UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

Uma abordagem *Forward-Looking* para estimar a PD segundo IFRS9

Luiz Henrique Outi Kauffmann

Dissertação de Mestrado do Programa de Mestrado Profissional em Matemática, Estatística e Computação Aplicadas à Indústria (MECAI)

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Luiz Henrique Outi Kauffmann

Uma abordagem *Forward-Looking* para estimar a PD segundo IFRS9

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Mestre em Ciências – Mestrado Profissional em Matemática, Estatística e Computação Aplicadas à Indústria. *EXEMPLAR DE DEFESA*

Área de Concentração: Matemática, Estatística e Computação

Orientador: Prof. Dr. Dorival Leão Pinto Jr.

USP – São Carlos
Julho de 2017

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados fornecidos pelo(a) autor(a)

- O21a Outi Kauffmann, Luiz Henrique
 Uma abordagem Foward Looking para estimar a PD
segundo IFRS9. / Luiz Henrique Outi Kauffmann;
orientador Dorival Leão Pinto Júnior. -- São
Carlos, 2017.
 83 p.
- Dissertação (Mestrado - Programa de Pós-Graduação
em Mestrado Profissional em Matemática, Estatística
e Computação Aplicadas à Indústria) -- Instituto de
Ciências Matemáticas e de Computação, Universidade
de São Paulo, 2017.
1. Risco de Crédito. 2. IFRS9. 3. PD. 4. Modelo
de Merton. 5. Regressão Logística. I. Pinto Júnior,
Dorival Leão , orient. II. Título.

Luiz Henrique Outi Kauffmann

**A Forward Looking Approach to estimate PD according to
IFRS9**

Master dissertation submitted to the Institute of Mathematics and Computer Sciences – ICMC-USP, in partial fulfillment of the requirements for the degree of the Master – Professional Masters in Mathematics, Statistics and Computing Applied to Industry. *EXAMINATION BOARD PRESENTATION COPY*

Concentration Area: Mathematics, Statistics and Computing

Advisor: Prof. Dr. Dorival Leão Pinto Jr.

USP – São Carlos
July 2017

Dedico este trabalho aos meus pais que sempre sonharam e lutaram pela educação dos seus filhos superando inúmeros problemas financeiros e pessoais.

AGRADECIMENTOS

Os agradecimentos principais são direcionados aos professores Dorival Leão Pinto Jr. e Antônio Castelo Filho que demonstraram energia e dedicação nesta turma do programa MECAl e sua disposição em ensinar, escutar e ajudar com esta iniciativa valiosa que uniu o mercado de trabalho e o mundo acadêmico. Tenho certeza que este é o primeiro grande passo para a geração e criação de novos conhecimentos e descobertas.

*“Nossa obrigação de sobreviver e prosperar é devida não apenas a nós mesmos,
mas também ao cosmos, antigo e vasto, do qual surgimos.”*
(Carl Sagan)

RESUMO

KAUFFMANN, L. H. O. **Uma abordagem *Forward-Looking* para estimar a PD segundo IFRS9**. 2017. 83 p. Dissertação (Mestrado em Ciências – Mestrado Profissional em Matemática, Estatística e Computação Aplicadas à Indústria) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2017.

Este trabalho tem por objetivo discutir as metodologias de estimação da PD utilizadas na indústria financeira. Além disso, contextualizar a aplicação do trabalho ao IFRS9 e seu direcionamento para o tema de Risco de Crédito. Historicamente os grandes bancos múltiplos utilizam variadas metodologias econométricas para modelar a Probabilidade de Descumprimento (PD), um dos métodos mais tradicionais é a regressão logística, entretanto com a necessidade do cálculo da Perda Esperada de Crédito através do IFRS9, se torna necessário mudar o paradigma de estimação para uma abordagem *forward-looking*, isto está sendo interpretado por muitas instituições e consultorias como a inclusão de fatores e variáveis projetadas dentro do processo de estimação, ou seja, não serão utilizados apenas os dados históricos para prever o descumprimento ou inadimplência. Dentro deste contexto será proposto uma abordagem que une a estimação da Probabilidade de Descumprimento com a inclusão de um fator *forward-looking*.

Palavras-chave: Probabilidade de Default, Probabilidade de Descumprimento, Impairment, Processo Estocástico, Risco de Crédito, IFRS9, Regressão Logística, *forward looking*..

ABSTRACT

KAUFFMANN, L. H. O. **A Forward Looking Approach to estimate PD according to IFRS9.** 2017. 83 p. Dissertação (Mestrado em Ciências – Mestrado Profissional em Matemática, Estatística e Computação Aplicadas à Indústria) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2017.

This paper aims to discuss the methodologies used to estimate the Probability Of Default used in the financial industry. In addition, contextualize the application of the work to IFRS9 requirements and its targeting to the Credit Risk theme. Historically large multi-banks use a variety of econometric methodologies to model the Probability of Default, one of the more traditional methods is logistic regression. However, with the need to calculate the expected credit loss through IFRS9, it becomes necessary to change the estimation paradigm to a *forward-looking approach*, this is being interpreted by many institutions and consultancies companies as the inclusion of factors and variables projected within the estimation process, that is, not only historical data are used to predict the default. Within this context will be proposed an approach that joins the estimation of Probability of Default with the inclusion of a *forward-looking* factor.

Keywords: Probability of Default, Impairment, Stochastic Process, Credit Risk, IFRS9, Logistic Regression, Forward Looking..

LISTA DE ILUSTRAÇÕES

Figura 1 – Information Value	43
Figura 2 – Categorização	44
Figura 3 – Parâmetros Estimados pela Regressão Logística	45
Figura 4 – Estatísticas de Ajuste	46
Figura 5 – Curva ROC	46
Figura 6 – Índices de KS e Gini entre as amostras	47
Figura 7 – Decomposição e Avaliação da Série	48
Figura 8 – Análise de Correlação Linear - Pearson	49
Figura 9 – Parâmetros do modelo ARIMAX	50
Figura 10 – Estatísticas	50
Figura 11 – Ajuste Final Modelo Foward Looking	51

SUMÁRIO

1	INTRODUÇÃO	19
2	O CONCEITO DE RISCO DE CRÉDITO	21
2.1	Risco Financeiro	21
2.2	Risco de Crédito e o Acordo de Basileia	22
2.3	IFRS9	23
2.3.1	<i>Diferenças entre Perda Esperada do acordo de Basileia e IFRS9 . .</i>	<i>24</i>
2.4	O Conceito de Probabilidade de Descumprimento	25
3	MODELAGEM QUANTITATIVA DA PD	27
3.1	Modelagem Quantitativa da PD	27
3.2	Modelos Estruturais	28
3.2.1	<i>Modelo de Merton</i>	<i>28</i>
3.2.2	<i>Aplicação na Indústria do Modelo de Merton e os modelos estruturais</i>	<i>31</i>
3.3	Modelos de Forma Reduzida	31
3.3.1	<i>Aplicação na Indústria dos Modelos da Forma Reduzida</i>	<i>32</i>
3.4	Modelos Econométricos	33
3.4.1	<i>Regressão Logística e Discrete-Choice-Models</i>	<i>34</i>
3.4.1.1	<i>Estrutura do Modelo Logístico aplicado à modelagem da PD</i>	<i>35</i>
3.4.2	<i>Análise de Sobrevivência e Duration Models</i>	<i>35</i>
3.4.2.1	<i>Estrutura do Modelo de Análise de Sobrevivência</i>	<i>36</i>
4	METODOLOGIA DE TRABALHO - ESTIMAÇÃO DA PD FORWARD LOOKING	39
4.1	Motivação do Problema	39
4.2	Base de Dados utilizada	40
4.3	Metodologia Proposta	40
4.4	Modelo de PD	41
4.5	Ajuste Temporal com modelo ARIMAX	41
4.6	Resultados	42
4.6.1	<i>Modelo de PD</i>	<i>42</i>
4.6.1.1	<i>Tratamento de Variáveis</i>	<i>43</i>
4.6.1.2	<i>Ajuste do Modelo</i>	<i>44</i>
4.6.2	<i>Modelo ARIMAX</i>	<i>47</i>

4.6.2.1	<i>Decomposição da Série de Taxa de Descumprimento</i>	47
5	CONCLUSÕES	51
5.1	Conclusões Gerais	51
	REFERÊNCIAS	55
APÊNDICE A	SCRIPTS SAS UTILIZADOS PARA MANIPULAÇÃO DE DADOS	57

INTRODUÇÃO

Este trabalho tem como objetivo contribuir com uma abordagem para a Modelagem da *Probabilidade de Descumprimento* considerando as recomendações do *IFRS9*. O *IFRS9* é um novo sistema de reportes financeiros, que recomenda as empresas a poupar (provisionar) um montante financeiro proporcional ao risco de descumprimento (inadimplência de crédito) de seus ativos de crédito. Entretanto, a maneira como se calcula este risco não deve utilizar apenas dados históricos das perdas ocorridas, e sim considerar sua projeção futura. Este novo método questiona os modelos econométricos tradicionais que sempre foram desenvolvidos baseados em um grande volume de dados históricos. Além disso, o *IFRS9* recomenda que para ativos com qualidade de crédito deteriorada seja calculado a estimativa até o vencimento dos contratos. Neste último caso a indústria tem utilizado os modelos conhecidos como modelos de análise de sobrevivência e colocado as variáveis macroeconômicas projetadas como inputs.

No caso das instituições financeiramente saudáveis é possível concluir que a maioria dos ativos de crédito não estejam qualificados com baixa qualidade, portanto para a maior parte dos ativos será necessário aplicar apenas a visao de projeção denominada como estimativa *forward-looking*. Neste contexto, muitas destas empresas podem aumentar seus custos de maneira considerável no processo de revisão e adequação dos seus modelos e sistemas financeiros. Considerando este possível custo será apresentado ao final deste documento uma abordagem que aproveita o sistema de modelagem de risco de crédito mais comum na indústria financeira, que é a modelagem da Probabilidade de Descumprimento através da regressão logística e uma combinação com um modelo de projeção futura. Esta abordagem foi proposta (KING; ZENG, 2001) para a modelagem de risco de crédito em dados com raros eventos de descumprimento e também foi proposta pela consultoria internacional *Mckinsey* para criar uma medida de risco de crédito agregada conhecida como capital econômico. Antes da proposição da metodologia será efetuada uma revisão da literatura dos principais métodos quantitativos existentes para a mensuração do risco de crédito e a Probabilidade de Descumprimento.

O CONCEITO DE RISCO DE CRÉDITO

O Objetivo deste capítulo é apresentar os conceitos de Riscos Financeiros e sua derivação para o Risco de Crédito com foco especificamente na modelagem da Probabilidade de Descumprimento (PD).

2.1 Risco Financeiro

Segundo (BESSIS, 2015) o conceito de risco apresenta a seguinte definição:

Definição 1. Risco é a existência da probabilidade de um evento assumir realizações positivas ou negativas.

Ao longo deste trabalho a palavra risco será utilizada de maneira recorrente, então é importante direcionar a definição anterior de risco para o Risco Financeiro. Será utilizada a definição de Risco Financeiro também dada por (BESSIS, 2015):

Definição 2. Risco Financeiro é a existência de um fator aleatório no retorno dos ativos e investimentos, tanto negativos quanto positivos.

Os Riscos Financeiros podem ser divididos em:

- Risco De Mercado
- Risco De Crédito
- Risco De Liquidez

O foco de estudo deste trabalho será o risco de crédito, este assim como cada um dos tipos de risco mencionados acima pode ser mensurado com parâmetros que são tratados como variáveis aleatórias. Atribuindo uma variável aleatória ou processo estocástico para cada parâmetro do

risco de crédito será possível mensurar e estimar o risco, gerando assim um conjunto de métricas de risco que ajudam as instituições a tomar decisões de investimento e alocação de capital.

2.2 Risco de Crédito e o Acordo de Basiléia

Dentro da Literatura de Gerenciamento de Riscos, o Risco de Crédito é considerado um dos tipos de riscos mais complexos e difíceis de mensurar. No Brasil ele ainda é o tipo de risco mais relevante dentro do setor bancário, pois é onde se concentra a maior parte do risco financeiro e capital alocado.

Desde julho de 2004, em que foi publicado o novo acordo de Basileia II, o setor Bancário passou por um momento de padronização de processos qualitativos e quantitativos para a gestão de risco de crédito, em que sua estrutura foi definida à partir de uma equação de perda esperada e parâmetros específicos.

A equação de Perda Esperada para Risco de Crédito contempla alguns parâmetros já conhecidos pela indústria financeira como:

- *PD*: Probabilidade de Descumprimento (*Probability Default*).
- *LGD*: Perda Financeira ou Severidade, dado o evento de descumprimento (*Loss Given Default*).
- *EAD*: Exposição financeira no momento do Descumprimento (*Exposure At Default*)

Dentro do acordo de Basiléia a fórmula de Perda Esperada foi definida com a seguinte combinação de parâmetros:

$$ECL = PD \times LGD \times EAD$$

No acordo de Basiléia II (2009) a Perda Esperada é utilizada para o objetivo de alocação de Capital, fato que afetaria o Patrimônio Líquido das Instituições Financeiras. Além disso, é recomendado pelo BCBS (*Basel Committee on Banking Supervision*), órgão regulador do Acordo de Basiléia, a utilização dos parâmetros de Perda Esperada também para a provisão de créditos.

O conceito de provisão de crédito apresenta diversas interpretações regionais, definidas para cada regulador local. No Brasil, por exemplo a metodologia de provisionamento de créditos para instituições financeiras é definida pela Resolução 2682, em que as instituições devem definir uma régua de classificação de crédito, semelhante as régua utilizadas pelas Agências de Classificação de Risco. O montante financeiro de provisão é definido para cada categoria da classificação de crédito definida pela mesma Resolução.

Na literatura contábil a provisão de crédito é conhecida como PCLD, provisão para créditos de liquidação duvidosa e existe como um princípio contábil independente da existência das normas do Banco Central ou Basiléia, é válida também para empresas que não sejam instituições financeiras. Entretanto, a idéia do BACEN é justamente padronizar o processo de classificação de risco de crédito que afeta o Balanço das Instituições Financeiras, pois estas empresas são responsáveis pela liquidez do sistema financeiro. De qualquer forma a provisão de crédito recomendada por Basiléia ou pelo Banco Central é o montante que existe no reporte financeiro das instituições financeiras para cobrir, da maneira mais precisa, a quantificação do risco de crédito de cada ativo.

Conforme foi mencionado, para cada país vai existir um agente regulador como o Banco Central que deve definir suas próprias regras de Provisão, muitas vezes inspirados no conhecido acordo de Basiléia, mas sempre levando em conta a realidade de suas respectivas economias.

O risco de crédito existe para qualquer empresa que contenha ativos de crédito, mas o Banco Central e BCBS regulam somente o sistema bancário e apenas para este podem criar um processo rígido de padronização de métodos e métricas contábeis. Portanto, o tema de padronização de métricas contábeis e suas metodologias sempre foram um grande desafio em todo o tipo de indústria, para este desafio foi criado, em 1973, um comitê chamado de *International Accounting Standards Committee* (IASC) que foi, em 2001, modificado para o *International Accounting Standards Board* (IASB). Este conselho tem por objetivo padronizar o processo de demonstração contábil de todas as empresas que são de capital aberto (incluindo instituições financeiras), tornando assim a análise das empresas que reportam sobre os padrões do IASB comparáveis. Os padrões contábeis, referentes ao princípio de PCLD, que foram publicados pelo IASB hoje são denominados de IFRS9. O IFRS9 possui a missão de padronização do cálculo e gestão da PCLD portanto, padronizarão diretamente o processo de cálculo de risco de crédito. A implantação do IFRS9 se torna menos complexa para os bancos que se candidataram ao processo de gestão de capital sugerido por Basiléia II, ou mesmo que criaram seus processos de gestão inspirados pelo acordo, devido a semelhança entre Basiléia e o IFRS9.

2.3 IFRS9

O IFRS9 faz parte das práticas contábeis sugeridas pelo IASB e se faz necessário para todas as empresas dos países aderiram às normas contábeis do IASB e todas as empresas de capital aberto. O Brasil é um dos países ao qual o mercado de capitais aderiu ao IFRS9, portanto todas as empresas de capital aberto no mercado brasileiro devem aderir ao IFRS9. O principal objetivo do IFRS9 é justamente a padronização do processo de gestão de risco de crédito e cálculo do PCLD através da mensuração da Perda Esperada. A perda esperada é um cálculo

que já existia no IAS39 (recomendação anterior ao IFRS9 pelo IASB) entretanto este cálculo considerava apenas o histórico ocorrido das perdas, sendo, neste caso, representado pela simples aplicação de um percentual sobre os ativos de crédito da Instituição. A filosofia deste modelo (anterior ao IFRS9) é denominada *Incurred Loss*. Entretanto o novo cálculo sugerido pelo IFRS9 deveria ser considerado na filosofia *forward-looking* ou mesmo considerando como cálculo de ECL (Expected Credit Loss), este modelo compartilha algumas semelhanças ao modelo sugerido por Basileia II entretanto adiciona requisitos novos que exigem uma maior sofisticação quantitativa e qualitativa.

2.3.1 Diferenças entre Perda Esperada do acordo de Basileia e IFRS9

Apesar da semelhança de Basileia e o IFRS9, existem algumas diferenças relevantes que devem ser destacadas:

- Perda Esperada utilizada para cálculo de Provisão: O Acordo de Basileia sugeria a utilização da Perda Esperada para o cálculo de provisão, mas utilizava de maneira direta apenas dentro do cálculo de capital alocado. Para o IFRS9 o foco é a utilização dentro cálculo de provisão de crédito.
- Visão Forward Looking: Diferente dos parâmetros de Basileia que eram estimados considerando um grande histórico e calibrados com os momentos de *stress financeiro*, para o IFRS9 esta estimativa deve considerar uma visão projetada para o futuro devendo ser a mais acurada possível.
- PD Lifetime: A Probabilidade de Descumprimento deve ser tratada diferente para ativos com qualidade creditícia deteriorada e deve ser estimada até o vencimento do ativo. Diferente de Basileia em que a visão era uma PD de 12 meses.

O IFRS9 exige que o cálculo de provisão de crédito seja derivado diretamente do cálculo da perda esperada, diferente da Resolução 2682, em que este estava atrelado ao sistema de classificação de crédito, portanto o IFRS9 vai introduzir um robusto processo de auditoria e controle dentro das áreas de riscos e crédito. Além disso, o acordo de Basileia permitia algumas abordagens mais simplificadas em termos quantitativos, ou seja, era possível substituir a utilização de modelos quantitativos por métricas de estatística descritivas ou fornecidas diretamente pelo Banco Central. Entretanto para o IFRS9 não existem diferentes abordagens, os bancos podem buscar a sua simplicidade em um modelo quantitativo mais simples, mas nunca optar pela não existência de modelos quantitativos. Com isto a dependência da sofisticação quantitativa da gestão de risco de crédito fica evidente e se torna um fator crítico na divulgação de resultados de empresas que possuem muita exposição ao risco de crédito.

O Objetivo deste trabalho vai focar diretamente no cálculo da Probabilidade de Descumprimento, que é um dos parâmetros centrais da mensuração de risco de crédito em ambas as abordagens (IFRS9 e Basiléia II) e desde de as primeiras pesquisas sobre o Risco de Crédito.

2.4 O Conceito de Probabilidade de Descumprimento

O Conceito de probabilidade de descumprimento é o parâmetro central de toda a literatura de *Credit Risk Analytics*. De uma maneira simples este parâmetro indica a probabilidade de descumprimento de um acordo ou contrato de crédito através da contraparte do contrato.

O conceito de Descumprimento ou Default pode variar de acordo com a instituição cedente do contrato de crédito, embora quando o parâmetro é utilizado para uma exigência regulatória ou normativa, deve atender uma padrão mínimo ou recomendação pré estabelecida, tanto para Basileia II quanto para o IFRS9.

O Conceito de Descumprimento para Basiléia II é:

- Atraso no pagamento dos fluxos de caixa em uma quantidade igual ou maior que 90 dias.
- Existe algum indicador de qualidade comprovado, como falência da contraparte ou acionamento de garantia.
- Para a modelagem de PD Basiléia é utilizado uma janela de descumprimento dos itens acima de até 12 meses após a data de referência de cálculo da estimativa da *PD*.

O Conceito de Descumprimento para IFRS9:

- Definição selecionada pela Instituição com base no seu processo de crédito e cobrança, mas o critério por atraso não deve ser maior que 90 dias de atraso.
- A janela de cálculo da PD vai variar de acordo com a qualidade de crédito:
- Para contratos que tiveram deteriorado sua qualidade de crédito, sua janela de cálculo será até o vencimento do contrato considerando a data de referência.
- Pode considerar 12 meses após a data de referência para créditos com boa qualidade de crédito, semelhante a PD de Basiléia II mas com uma visão projetada.

MODELAGEM QUANTITATIVA DA *PD*

Conforme mencionado no capítulo anterior o foco do trabalho será na modelagem da *PD*. O estudo da probabilidade de descumprimento é o tema mais maduro dentro da literatura quantitativa de risco de crédito e sua forma de medição pode seguir diversas abordagens, portanto este capítulo tem como objetivo discutir os principais modelos e abordagens de cálculo da *PD*, estes conceitos serão fundamentais para o modelo que vai ser proposto e o devido entendimento dos resultados que serão demonstrados ao final do trabalho.

3.1 Modelagem Quantitativa da *PD*

A literatura para a modelagem da Probabilidade de Descumprimento é relativamente antiga e começou após o artigo de *Robert Merton* utilizando o trabalho desenvolvido por (BLACK; SCHOLES, 1973), então à partir desta época muitos estudos e metodologias foram desenvolvidos e aplicados. Cada metodologia desenvolvida ou adaptada possui focos e premissas distintas de acordo com seu objetivo de utilização, com isto é possível encontrar Modelos Estocásticos, Econométricos e até mesmo Machine Learning em seu desenvolvimento. Esta literatura da modelagem da Probabilidade de Default é classificada em grupos de metodologias, neste trabalho será utilizada a classificação proposta por (BOHN; STEIN, 2009) que classifica as abordagens de modelagem quantitativa em três categorias:

- Modelos Estruturais.
- Modelos de Forma Reduzida.
- Modelos Econométricos.

3.2 Modelos Estruturais

Os Modelos Estruturais são aqueles que partem de uma premissa qualitativa ou econômica, entretanto assumem a forma de alguma equação específica e possuem uma justificativa causal. Um exemplo destes modelos são os modelos de apreçamento de opções que assumem que o comportamento econômico de um ativo que segue uma distribuição de probabilidade conhecida e justificada segundo premissas teóricas. No caso da Modelagem de Default, existe um conjunto de modelos conhecidos pela literatura, que são listados abaixo:

- Modelo de Merton.
- Modelo KMV.
- Longstaff and Schwartz.

Na realidade a base teórica e estrutural de todos estes modelos é o Modelo de Merton, que será detalhado na próxima seção, juntamente com suas propriedades matemáticas e premissas econômicas.

3.2.1 Modelo de Merton

Um dos primeiros casos da necessidade de mensuração do risco de crédito foi para precificar de uma maneira justa os títulos de crédito emitidos por grandes empresas no mercado de capitais. Estes títulos, diferente dos títulos do governo apresentam um risco de crédito considerável e cada contraparte, ou emissora deles apresenta seu próprio risco de crédito. No Brasil títulos como estes são chamados de Debêntures e no mercado estrangeiro são conhecidos como *Bonds*. De acordo com (MERTON, 1974) para mensurar o risco de crédito de um título de dívida (Debêture) é necessário comparar o valor de mercado dos ativos da empresa emissora com o valor de mercado de suas dívidas, sendo que o ativo é tratado como um processo estocástico. Uma empresa entrará em Descumprimento ou *Default* no momento em que o valor de mercado de seus ativos ficar menor que o valor de mercado de suas dívidas. O Modelo de Merton apresenta as seguintes premissas econômicas:

1. Taxa do Juros livre de risco é constante ao longo do tempo.
2. O Valor da Empresa apresenta desvio padrão constante.
3. No caso de Descumprimento ou evento de Default a empresa ou contraparte emissora não possui custos adicionais, como custos para advogados e afins.
4. Existe uma taxa de recuperação da dívida que seria dado pela fórmula $(1 - LGD)$, mas no caso será definida pela prioridade dos acionistas em recolher seus ativos no caso de um evento de Descumprimento (*Default*).

5. Toda a dívida da empresa esta concentrada em um único título de dívida com apenas um fluxo de caixa com vencimento em T . Esta foi a premissa mais discutida e os modelos herdeiros de Merton trataram esta premissa sem mudanças conceituais no processo de estimação da PD . O modelo herdeiro mais conhecido é o modelo KMV .
6. O gatilho do evento de Descumprimento ocorre no vencimento da Debênture (Título de Dívida) quando no final da maturidade o valor dos ativos da empresa ficam menores que o valor de suas dívidas.

Merton foi um dos ganhadores prêmio Nobel com o seu artigo [Merton \(1997\)](#), que é um trabalho que explica e detalha as interpretações matemáticas do famoso teorema de Black e Scholes. O modelo estrutural de Merton também é uma aplicação do teorema e considera o título de dívida como um instrumento chamado de opção de venda européia, chamada pela indústria de *put option*. Aplicando as premissas do modelo de Merton temos que os ativos da empresa se comportam como um processo estocástico conhecido como o Movimento Geométrico Browniano:

$$dV_t = \mu V_t dt + \sigma dW_t$$

E aqui $W(t)$ é um movimento Browniano sobre uma medida de probabilidade P sendo que o valor inicial dos ativos da empresa é representado por V_0 . Portanto a equação que representa o valor de ativos da empresa é dada por:

$$V(t) = V(0) \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W_t\right)$$

Conforme descrito anteriormente, no modelo de Merton, uma contraparte entrará no estado de Descumprimento (*Default*), quando o valor de mercado dos seus ativos estiver menor que o valor de suas dívidas, que no caso são representadas por um único título de dívida D_t . O Objetivo aqui será calcular a Probabilidade de Descumprimento na ótica do credor, no caso é a entidade que vai receber o valor prometido pelo título de dívida da empresa D_t . Assumindo isto, podemos escrever o valor recebido como uma variável aleatória da seguinte maneira:

$$D_t = \begin{cases} D_t, & \text{se } D_t \leq V_t, \\ V_t, & \text{se } D_t > V_t. \end{cases}$$

No caso o valor D_t depende apenas de sua atualização financeira, que seria sua correção pela taxa de juros, mas no modelo de Merton esta é uma taxa constante. É possível ver em [Lando \(2004\)](#) uma versão do modelo de Merton considerando a taxa de juros como um processo estocástico, mas este não é o objetivo deste trabalho. Portanto o único componente aleatório dentro desta relação matemática é o V_t . É importante notar que no caso do descumprimento o credor pega a diferença entre o valor dos ativos, utilizado como uma garantia da dívida e o valor de sua dívida, gerando assim um saldo recuperado, conforme colocado anteriormente a taxa de recuperação foi assumida apenas de acordo com a prioridade de recebimento dos credores.

Escrevendo matematicamente é possível calcular o valor esperado de recebimento, na ótica dos credores, através da relação:

$$E[D_t] = D_t - \max\{D_t - V_t, 0\}.$$

O trabalho proposto por Merton basicamente foi uma revisão aplicada do trabalho de Fisher Black e Myron Scholes, este trabalho apresenta a solução analítica para a precificação justa das opções de compra e venda. Diante deste contexto Merton considerou o segundo termo da equação anterior como uma posição vendida em uma opção de venda (*put*), ou seja, uma posição em que o credor vendeu uma opção de venda. Neste caso a opção definida apresentaria o valor de *strike* equivalente ao valor da dívida, pois este é um valor constante, e o valor do ativo objeto da opção é o valor dos ativos da empresa ou contraparte do título de dívida. Além do valor dos ativos ser regido pelo Movimento Geométrico Browniano, Merton considerou o valor de ativos da empresa como uma distribuição *Lognormal*, sendo assim é possível calcular de maneira analítica o segundo termo da equação aplicando a solução proposta por Black e Scholes. Portanto o valor esperado de recebimento pode ser escrito matematicamente através do Modelo de Black e Scholes, da seguinte maneira:

Seja uma opção de venda de um Ativo A_t com distribuição *Lognormal* que apresenta trajetória representada pelo Movimento Geométrico Browniano:

$$A_t = A_0 \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W_t\right).$$

Então o preço economicamente justo¹ para a opção de venda com o Strike dado por K é

$$p = K \exp^{-rT} \phi(d_2) - A(t) \phi(d_1)$$

Em que

$$d_1 = \frac{\ln(A(t) \backslash K) + (r + \sigma^2/2)T}{\sigma \sqrt{T}}$$

e

$$d_2 = d_1 - \sigma \sqrt{T}$$

Aplicando o modelo acima temos então que o valor esperado de uma dívida com risco de crédito pode ser dado por:

$$E[D_t] = D_t - (D_t \exp^{-rT} \phi(d_2) - V(t) \phi(d_1))$$

É importante observar que esta relação encontrada pelo Modelo de Merton implica que o valor esperado de uma dívida cedida para uma contraparte que apresente risco de crédito se torna menor do que o valor de face da dívida D_t corrigido pela taxa de juros, intuitivamente isto deveria ser verificado e é por isto que o Modelo Estrutural de Merton e seus derivados também podem

¹ O valor econômico justo é um conceito econômico que é estudado e revisado na teoria de precificação de ativos e na própria demonstração do modelo de *Black and Scholes*, teoricamente seria o valor que torna o ativo ou portfólio precificado livre de risco

ser utilizados para calcular o custo do risco de crédito conhecido como **spread de risco de crédito**. Esta discussão pode ser vista em Paschoarelli (2007). A aplicação do modelo de *Black and Scholes* não deixa a Probabilidade de Descumprimento explícita, entretanto sua verificação é simples e vem da interpretação do modelo de Black and Scholes, em que a probabilidade do Valor do Ativo ser menor que o valor de K é dada por $1 - \Phi(d_2)$, portanto no modelo de Merton encontramos a probabilidade de Descumprimento dada por:

$$PD = 1 - \Phi(d_2) = \Phi(-d_2)$$

Com

$$d_2 = \frac{\ln(A(t)/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}} - \sigma\sqrt{T}$$

Esta é a principal razão pela literatura chamar a medida d_2 de Distância ao Default (DD).

3.2.2 Aplicação na Indústria do Modelo de Merton e os modelos estruturais

Os modelos de Merton e outros estruturais apresentam uma dependência das informações de mercado das empresas e uma premissa de liquidez do mercado de dívidas, que aqui no Brasil seria o mercado de Debêntures. Estas duas dependências não são realidade do Brasil e da América Latina, pois poucas empresas possuem informações abertas e além disso o mercado de Debêntures possui uma estrutura não muito líquida. Além de tudo o mercado da América Latina é muito concentrado em bancos múltiplos, em que a maior parte da exposição de risco de crédito se concentra no empréstimo para pessoas físicas ou pequenas e micro empresas. De qualquer maneira, para Bancos de Investimento ou especializados em crédito para grandes empresas ainda pode ser uma boa opção.

3.3 Modelos de Forma Reduzida

Os modelos de Forma Reduzida, diferente dos modelos estruturais, não se preocupam em explicar economicamente a causa do evento de descumprimento (*Default*), ao invés de encontrar um modelo econômico eles tratam o evento de descumprimento como um processo estocástico buscando calibrar processos estocásticos aos dados de mercado diretamente relacionados ao evento de default, por exemplo:

- Mercado de CDS: é um mercado onde são negociados instrumentos de *SWAP* que trocam a exposição de risco de descumprimento por um prêmio financeiro materializado em uma estrutura de juros ou spread financeiro. Este mercado também não é líquido na América Latina.

- Mercado de Debêntures: é o mercado em que são negociadas as dívidas das empresas, que estão abertas para compra de qualquer contraparte e possuem uma estrutura de juros determinada pela demanda e apetite de risco dos investidores do mercado de capitais.
- Sistemas de Rating: no mercado nacional e internacional existem as famosas agências de classificação de rating que utilizam métodos econométricos, financeiros e qualitativos para qualificar todos os instrumentos e empresas abertas em uma única régua (cada agência possui sua régua) de risco de crédito.

Estes modelos foram estudados por vários pesquisadores, dentre eles [Jarrow, Lando e Turnbull \(1997\)](#) se destacaram e propuseram o chamado *DSL framework*. O Objetivo aqui é precificar um dos instrumentos de dívidas, como as debêntures, através de um modelo de intensidade. Em que:

$$D(T) = E_0^Q[\exp^{-\int_0^T (r(t) + h^Q(t))dt}]$$

$$r(t) = a_r(t) + b_r(t)X_c(t)$$

$$h^Q(t) = a_{h^Q}(t) + b_{h^Q}(t)X_c(t)$$

- $D(T)$ é o valor da dívida ajustada pelo risco de crédito no vencimento T .
- r é a taxa de juros livre de risco.
- h^Q é a intensidade do Descumprimento (*default*).
- X_c é um fator comum conectando o processo de juros com a intensidade de default.

O fator comum X_c pode ser derivado de dois processos estocásticos conhecidos pela literatura:

- Vasicek: $dX_c = k(\mu(x_c - X_c)dt + \sigma_{x_c}dz)$.
- Cox, Ingersoll e Ross (CIR): $dX_c = k(\mu(x_c - X_c)dt + \sigma_{x_c}\sqrt{X_c}dz)$.

3.3.1 Aplicação na Indústria dos Modelos da Forma Reduzida

No mundo acadêmico os Modelos da Forma Reduzida fazem parte das abordagens mais sofisticadas existentes, inclusive eles apresentam a solução para precificação de muitos instrumentos de risco de crédito, entretanto devido a sua complexidade matemática e numérica não são muito aplicados, embora fazem parte das discussões mais modernas para a mensuração de risco de crédito. Hoje com o surgimento do acordo de *Basileia III* a pressão para modelos de precificação e quantificação do risco de crédito através da métrica de CVA (*Credit Value*

Adjustment) de maneira detalhada tem crescido muito, mas a indústria ainda não conseguiu padronizar um framework único. Este inclusive é um ótimo assunto para novas pesquisas na modelagem quantitativa de risco de crédito. No mundo dos bancos múltiplos estes modelos poderiam ser aplicados na precificação dos empréstimos parcelados para pessoa física, porém se tornaria computacionalmente inviável seu cálculo para cada instrumento. Entretanto, com o advendo do *Big Data* provavelmente este problema será superado.

3.4 Modelos Econométricos

Os modelos econométricos não são classificados como modelos estruturais, por razões diretas, pois eles são aplicados de maneira empírica para modelar um padrão de comportamento baseado nos dados históricos e até mesmo projeções e correlações. Entretanto, poderiam, ser classificados dentro dos modelos de forma reduzida, entretanto muitos pesquisadores tratam eles de maneira separada justamente pela maior parte da aplicação não recorrer ao evento de default como um processo estocástico e sim como um modelo de classificação. Um dos primeiros modelos econométricos conhecido na literatura é o Modelo *Z-Score* de Altman que pode ser visto em [Altman \(1968\)](#), essencialmente este é um modelo em que foi utilizado uma técnica de análise de discriminante, em que o objetivo foi selecionar as variáveis de balanço que mais separam as empresas com alto e baixo risco de descumprimento.

A utilização de modelos econométricos é uma das abordagens mais aplicadas para tratar o risco de crédito de carteiras de pessoa física e pequenas empresas. Na América Latina a maior parte do risco de crédito está concentrada no crédito direto para o varejo (pessoas físicas e pequenas empresas), por isto a maior parte dos trabalhos acadêmicos do Brasil, como o de [Visconti \(2011\)](#) e [Sicsu \(2010\)](#) recorrem aos estudos econométricos, uma vez que o crédito direto ao varejo apresenta um grande volume de dados e grande histórico de informação. Além disto, neste caso não teremos como aplicar os modelos estruturais, pois cada pessoa possui seu próprio balanço financeiro, e também se torna difícil calibrar um único processo estocástico para cada contrato ou contraparte, como nos modelos de forma reduzida.

No mercado norte americano as taxas de juros são relativamente muito baixas e o mercado de securitização² é muito mais líquido, portanto os modelos estruturais e de forma reduzida podem ser mais fácil de aplicar. Além disso o mercado mais movimentado apresenta uma visão de risco mais agregada, em nível portfólio, pois os bancos revendem os créditos diretos em forma de portfólios para outras contrapartes constituindo os fundos de securitização, gerando uma complexidade econômica que já é estudada nas premissas dos modelos estruturais e de forma reduzida.

A taxonomia utilizada para esta seção considera os modelos classificados em dois tipos:

² são mercados em que os bancos agregam um conjunto de crédito diretos e vendem sobre a forma de fundos de investimento

- Modelos de Escolha Discreta (*Discrete-Choice-Models*): A técnica mais comum e mais aceita pela indústria é a regressão logística, embora exista uma tendência atual para a utilização de outros modelos econométricos ou técnicas de aprendizado de máquina como : Árvores de Decisão, Support Vector Machine, Redes Neurais e outras.
- Modelos de Sobrevivência (*duration models*): A técnica mais comum neste caso é a modelagem de análise de sobrevivência com o modelo de razões proporcionais, esta técnica está se tornando um *framework* para os bancos e instituições que querem calcular a *PD lifetime* para os modelos de risco de crédito alinhados ao *framework do IFRS9*.

As próximas seções será apresentado um detalhamento maior de cada uma destas metodologias.

3.4.1 Regressão Logística e Discrete-Choice-Models

O modelo de regressão logística vem sendo uma das técnicas mais utilizadas na indústria, não somente para o caso da modelagem de Descumprimento, mas também para temas como modelagem de propensão de compra, modelos de Churn³. Sua aceitação pela indústria se deve à uma série de fatores como:

- Premissas Estatísticas: O modelo de regressão logística não assume distribuição normal dos erros de estimação e da distribuição de variáveis explicativas.
- Ajuste: O modelo de regressão logística não assume relação linear das variáveis explicativas e variável resposta, logo para estudar risco de crédito que possui um limite que não é linear para valores extremos, se torna economicamente mais aceitável.
- Modelagem de eventos discretos: Em muitos casos de modelagem qunantitativa o objetivo não é acurácia, mas sim o poder de classificação dos modelos, então o modelo logístico apresenta como principal objetivo a ordenação dos dados em probabilidades relacionadas ao evento discreto modelado. A maior parte dos problemas de negócio acabam sendo tratados como eventos discretos, tais como: evento de descumprimento, evento de compra de produto, evento de fraude, resposta positiva ao tratamento de remédios ou mesmo penetração da política de crédito e descontos para cliente estratégicos. Nestes, é melhor estimar a probabilidade de uma contraparte cumprir seus pagamentos do que tentar estimar qual o valor de financeiro será utilizado, pois este último pode ser tratado via políticas flexíveis e qualitativas, ou mesmo apresentar uma distribuição de probabilidade de difícil mensuração.

³ Churn é o evento em que o cliente decide por deixar o relacionamento com uma empresa, é um dos parâmetros mais utilizados no mundo do CRM - Customer Relationship Management

3.4.1.1 Estrutura do Modelo Logístico aplicado à modelagem da PD

Seja o evento de descumprimento uma variável aleatória:

$$Y = \begin{cases} 1, & \text{se } \text{Descumprimento}, \\ 0, & \text{se } cc \end{cases}$$

A probabilidade de Descumprimento dado um conjunto de variáveis explicativas pode ser escrita como $P(Y = 1|X)$ e sua forma matemática pode ser escrita como:

Seja g , tal que,

$$g(X) = g(x_1, x_2, \dots, x_n) = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n$$

Teremos a função do modelo logístico dada por:

$$P(Y = 1|X) = f(X) = f(g(X)) = \frac{1}{1 + \exp^{-g(X)}}$$

Neste caso a Variável Y apresenta distribuição de Bernoulli, ou seja, $Y \sim \text{Bernoulli}(PD)$, sendo o parâmetro PD dado pela probabilidade de Descumprimento, no caso do modelo logístico acima. Em um conjunto de n registros, em que cada registro é uma contraparte ou um contrato de crédito a estimação do modelo logístico é através da seguinte função de verossimilhança:

$$L(Y_i; PD(x_i, \beta_i)) = \prod_{i=1}^n P(Y_i = y_i | X)$$

Esta estimação é solucionada pelo método de máxima verossimilhança sendo o método de Newton-Raphson aplicado para sua solução numérica, que é a base de quase todos os pacotes estatísticos do mercado para esta estimação. Basicamente a única grande premissa deste modelo é o de que a relação entre as variáveis explicativas seja linear de acordo com o $\text{Log}(\frac{PD}{1-PD})$, métrica comumente chamada de "Log odds".

3.4.2 Análise de Sobrevida e Duration Models

Os modelos de análise de sobrevivência começaram a ser aplicados por duas grandes necessidades:

- Trabalhar com dados censurados: para basileia a estimativa da PD é realizada em uma janela de previsão para os próximos 12 meses de vida do contrato, portanto quando é criado um planejamento de experimento para a realização da modelagem quantitativa os contratos são observados apenas 12 meses depois de sua data de referência, com isto se um contrato entrar em descumprimento no décimo terceiro mês ele vai ser modelado na amostra de desenvolvimento como um cliente bom, ou seja, que não entrou em descumprimento nos próximos 12 meses, portanto sua amostra vai apresentar censura relacionada ao

evento de descumprimento. Nos casos em que se quer modelar o tempo esperado para o descumprimento, é recomendado utilizar a análise de sobrevivência (BAESENS; ROESCH; SCHEULE, 2016).

- Criar uma estrutura temporal de Default: muitas vezes para um credor é interessante saber não somente a probabilidade de descumprimento mas qual o tempo esperado até que o contrato entre em descumprimento, ou o tempo de sobrevivência deste contrato. Esta é inclusive uma das características necessárias para a *Lifetime PD* solicitada pelas normas do IFRS9. Além disso, a criação da estrutura temporal permite uma modelagem da curva de descumprimento utilizada para a definição do preço justo do crédito dado seu risco, para qualquer prazo de contratação.

3.4.2.1 Estrutura do Modelo de Análise de Sobrevivência

Conforme explicado anteriormente o modelo de sobrevivência tem por objetivo mensurar o tempo esperado até o evento de descumprimento, e esta estimativa vem da definição do próprio conceito de sobrevivência. Matematicamente podemos definir a probabilidade instantânea do evento de descumprimento no tempo T dada pela função f :

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t}$$

Com isto definimos a probabilidade de sobrevivência até o momento T como a função de distribuição acumulada:

$$F(t) = P(T \leq t) = \int_0^t f(u) du$$

Então a função de sobrevivência a partir do tempo t se torna

$$S(t) = 1 - F(t) = \int_t^\infty f(u) du$$

Definindo a função de sobrevivência para a devida explicação do modelo de sobrevivência é necessário a definição da função de *hazard*, que é dada pela relação:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$$

Esta função de *hazard* seria a probabilidade instantânea de descumprimento no tempo T dado que não ocorreu nenhum evento de descumprimento antes de t . Para a academia e indústria é conhecida como função intensidade ou risco (*hazard*) Geralmente para aplicar a modelagem de sobrevivência é necessário observar empiricamente a função de *Hazard* e escolher a função matemática que se adequa aos dados observados. Existem muitos tipos de função de *Hazard* na literatura, que são:

- Constante: onde o evento de descumprimento não muda com a idade do contrato ou acordo de crédito

- Crescente: quando o evento de descumprimento cresce de acordo com a idade do contrato ou acordo de crédito.
- Decrescente: quando o evento de descumprimento decresce de acordo com a idade do contrato ou acordo de crédito.
- Convexa: quando a partir de uma idade o contrato aumenta sua probabilidade de descumprimento, mas depois de certa idade esta probabilidade diminui. Este tipo de *Hazard* é aplicado quando o evento modelado é a taxa de sobrevivência humana.

A escolha do tipo de função de intensidade para os eventos de descumprimento é tipicamente a função decrescente, por questões lógicas, pois conforme o contrato de crédito vai sendo pago, ou seja, sobrevive ao longo do tempo sua probabilidade de descumprimento decresce até o final de sua vida. O modelo mais comum utilizado pela indústria, conforme *Baesens(2016)* é a função de intensidade com razões proporcionais conhecida como o modelo não paramétrico de cox. Com isto a estrutura do modelo econométrico de sobrevivência apresenta a seguinte estrutura:

$$h(t, x_i) = h_0(t) \exp(\beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_n x_{in})$$

Sendo que o vetor $X = (x_{i1}, x_{i2}, \dots, x_{in})$ é uma estrutura que contém as variáveis explicativas do contrato ou contraparte i dentro de uma base de dados. É importante citar que existem outros modelos de intensidade conhecidos na indústria, mas cada tipo é escolhido de acordo com a observação do evento que será modelado.

METODOLOGIA DE TRABALHO - ESTIMAÇÃO DA PD FORWARD LOOKING

Após a explicação das diferentes abordagens de modelagem da *PD*, será introduzida neste capítulo a metodologia de estimação proposta para empresas e instituições financeiras que desejam transformar um modelo econométrico de PD desenvolvido através da regressão logística em uma visão *forward-looking* apenas com a modelagem de um único componente da função logística, transformando este componente em um modelo de série de tempo.

4.1 Motivação do Problema

Conforme contextualizado no segundo capítulo, a regulação financeira é um dos grandes instrumentos de sofisticação quantitativa para as práticas de risco, incluindo o risco de crédito. A chegada do IFRS9 vai obrigar todas as empresas e instituições financeiras que são classificadas como sociedade anônima¹ a mudarem a maneira como se calcula a sua perda esperada. Para o objetivo deste trabalho que é a aplicação da modelagem de PD, as mudanças que mais afetam estes parâmetros são:

1. Abordagem Forward Looking: parâmetro de PD deve conter uma estimativa que não seja obtida apenas pelo histórico de perdas ocorridas ou *incurred loss* e sim de uma abordagem de projeção futura de perdas de crédito.
2. *Lifetime PD*: Para os contratos com baixa qualidade de crédito as probabilidades de descumprimento devem ser estimadas até o final da vida do contrato.

¹ Sociedades Anônimas são empresas que possuem uma constituição jurídica em que as participações da empresa são distribuídas no mercado de capitais através de ações que podem ser transacionadas por qualquer contraparte do mercado de capitais

O segundo item acima pode ser tratado com a aplicação dos modelos de forma reduzida, ou mesmo o modelo de sobrevivência com razões proporcionais, que são técnicas relativamente conhecidas no mercado e já estudadas em diversos trabalhos. Entretanto o primeiro item não pode ser subestimado, uma vez que a maior parte dos ativos de crédito de uma instituição financeira geralmente não está classificado com baixa qualidade de crédito e mesmo que estes não tenham que passar por uma estimativa *lifetime* eles devem ser mensurados com uma visão *forward-looking*. Portanto o objetivo deste trabalho será propor uma metodologia de estimação da *forward-looking*

4.2 Base de Dados utilizada

Para embasar a metodologia proposta, será utilizada uma base de dados na forma de painel que contém cinquenta mil contratos de crédito imobiliário dos estados unidos conhecidos como *U.S. residential mortgage-backed securities* (RMBS) extraídos do órgão International Financial Research. Dentro desta base temos a identificação por contrato de crédito e uma variável de data de referência que contém informações características do contrato de crédito, como seu saldo devedor e idade, além disso possuímos as variáveis macroeconômicas diretamente observadas pela respectiva data de referência. Esta base contém 50 mil contratos observados durante 60 meses, totalizando 622.489 registros. A base de dados pode ser obtida através do site <http://www.creditriskanalytics.net/> seu nome estará como *mortgage*.

4.3 Metodologia Proposta

King e Zheng em seu artigo *Logistic Regression in Rare Events Data* (King e Zeng (2001)) propuseram um fator de correção temporal dentro da regressão logística tradicional, que pode ser modelada por uma componente de série de tempo. Sendo assim teremos a seguinte estrutura do Modelo Logístico.

$$PD_{igt} = \frac{1}{1 + \exp(-f_{igt})}$$

Em que

$$f_{igt} = f_i + c_{gt}$$

E

$$c_{gt} = \ln\left[\frac{D_{gt}}{1 - D_{gt}} \times \frac{1 - P_{gt}}{P_{gt}}\right]$$

Neste caso a PD_{igt} é a Probabilidade de Descumprimento ajustada pelo modelo logístico do cliente i pertencente ao setor g no momento t . Os elementos f_i e c_{gt} são respectivamente, o preditor linear do cliente i do modelo de PD sem o fator de ajuste macroeconômico e o fator macroeconômico da classe de risco g no tempo t . D_{gt} é a componente de dependente do modelo de séries e P_{gt} é a taxa média de default no mês t para a classe de risco g . Um tratamento

semelhante pode ser observado no modelo desenvolvido pela empresa *Mckinsey* chamado de *Credit Portfolio View* que foi baseado em [Wilson \(1997a\)](#). Diante desta metodologia serão estimados através da base de dados descrita no tópico anterior dois modelos quantitativos:

1. Um Modelo de Probabilidade de Descumprimento
2. Um Modelo conhecido como ARIMAX(*Autoregressive Integrated Moving Average with eXtra*) que trata de um modelo ARIMA com a inclusão de covariáveis explicativas

4.4 Modelo de PD

O Modelo de PD ajustado terá a estrutura classica de um modelo logístico dada por

$$PD_i = \frac{1}{1 + \exp(-f)}$$

Em que

$$f(X) = \beta_1 X_1 + \beta_2 X_2 + \dots \beta_n X_n$$

A variável de Descumprimento será o evento de Descumprimento em uma janela de até 12 meses após a data de referência.

4.5 Ajuste Temporal com modelo ARIMAX

O Modelo ARIMAX apresentará a seguinte estrutura:

$$Y_t = \mu + \sum_i \Pi_i X_{it} + \frac{\Theta(B)}{\Phi(B)} \varepsilon_t$$

Em que:

- Y_t : taxa de descumprimento no tempo t
- μ : nível médio, intercepto
- X_{it} : variáveis preditoras exógenas, índices de mercado
- Π_i : parâmetro de elasticidade
- $\Theta(B)$: polinômio em B da parte AR – parâmetros da parte AR
- $\Phi(B)$: polinômio em B da parte MA – parâmetros da parte MA

Serão utilizadas no processo de estimação da Série as variáveis Macroeconômicas contidas na base de dados que são:

- Default-time: taxa de default observada na data t
- HPI: índice de preços médios das casas na data t
- GDP-time: taxa de crescimento do PIB americano na data t
- UER-time: taxa de desemprego nos Estados Unidos na data t
- Interest-rate-time Taxa de juros básica dos Estados Unidos na data t

Como o IFRS9 para os créditos que não estão com qualidade deteriorada pressupõe a estimativa de Probabilidade de Descumprimento para uma janela de 12 meses, o componente de séries de tempo estimado terá como objetivo explicar a taxa de default da carteira para os próximos 12 meses, utilizando as variáveis macroeconômicas como co-variáveis.

4.6 Resultados

4.6.1 Modelo de PD

Conforme definido no tópico anterior foi realizado um modelo de Regressão Logística, entretanto para o processo de estimação foram realizados alguns ajustes nos dados que seguem abaixo:

1. Tratamento da variável resposta: Os dados estão em formato painel, então cada contrato aparece repetidamente em inúmeros registros, foi selecionada para cada safra uma janela de 12 meses de observação e caso o contrato tenha entrado em descumprimento foi criada uma variável resposta com valor 1, caso contrário esta variável teria valor 0.
2. Tratamento de Repetição dos contratos: No modelo de regressão é assumido que cada elemento amostral não apresenta dependência, justamente pela função de máxima verossimilhança ser composta pela multiplicação das probabilidades de cada elemento apresentar descumprimento nos próximos 12 meses. Entretanto, esta tabela contém uma repetição de contratos pois cada contrato aparece mensalmente na base e queremos estimar uma probabilidade mensal de descumprimento, ou seja, queremos que para dado um mês de referência, saber a probabilidade (com variáveis explicativas deste do mês de referência) do contrato entrar em descumprimento nos próximos 12 meses. Portanto foi feito um sorteio de uma variável aleatória para cada repetição do contrato, e escolhido apenas um mês de referência apenas, sendo este contrato um contrato que apresentou descumprimento ou não.
3. Seleção de safras: foram selecionados apenas contratos que tinham no mínimo 12 safras de observação, ou seja, tinha a variável de data de referência menor ou igual a 48.

Ao final das etapas anteriores chegamos ao que é conhecido na indústria como *ABT - Analytical Base Table* que é uma tabela com os elementos amostrais para cada linha e em suas colunas as variáveis independentes e dependentes. Além disso, para todos os processo posteriores foi selecionado aleatoriamente uma amostra com 70% dos registros chamada de amostra de treinamento e 30% como amostra de validação.

4.6.1.1 Tratamento de Variáveis

A regressão logística é comumente modelada com variáveis explicativas categóricas devido a sua interpretação, pois como a variável dependente é binária, se torna mais fácil a interpretação das taxas da variável resposta para cada categoria da variável explicativa ou independente. Portanto é uma boa prática a categorização de cada variável independente. Com isto foi aplicado um procedimento de categorização das variáveis com o objetivo de que ao longo das categorias o $\ln\left(\frac{\%Descumprimento}{1-\%Descumprimento}\right)$ seja linear, esta métrica que é conhecida como "*Log Odds*", além disso foi observado as métricas de poder explicativo de cada variável e a lógica baseada no conhecimento do assunto. Por exemplo: Quanto maior o Escore da FICO menor a taxa de contratos em descumprimento. Segue abaixo os resultados destas análises:

Figura 1 – Information Value

Fonte – Elaborada pelo autor.

O Information Value mostrado acima, é uma das medidas mais importantes para avaliar o poder preditivo de uma variável explicativa, de acordo com [Siddiqi \(2006\)](#) temos a seguinte referencia:

- Menor que 0.02: Sem valor preditivo
- Entre 0.02 e 0.1: Baixo poder preditivo

- Entre 0.1 e 0.3: Médio poder preditivo
- Entre 0.3 e 0.5: Forte poder preditivo
- Acima de 0.5: Outlier o valor suspeito de poder preditivo, pode ser uma variável diretamente derivada da variável dependente e deveria ser descartada do processo de modelagem.

A fórmula do *Information Value* - IV é dada por:

$$IV = \sum_{i=1}^{k-\text{categorias}} \%Descumprimento - \%Normalidade \times \ln \frac{\%Descumprimento}{\%Normalidade}$$

Figura 2 – Categorização

Fonte – Elaborada pelo autor.

Portanto através da categorização das variáveis foram excluídas as variáveis que possuem menos um *information value* menor que 0.1. Além delas foram removidas as variáveis macroeconômicas independente do seu *information value*, pois elas variam dependendo do mês de referência, e não apresentam variabilidade para diferentes contratos observados no mesmo mês de referência. A ideia é utiliza-las no modelo de série de tempo.

4.6.1.2 Ajuste do Modelo

Após todos os procedimentos descritos anteriormente foi executado um procedimento de estimação dentro do software estatístico SAS Enterprise Miner utilizando um método de seleção de variáveis conhecido por *stepwise*. Com isto foi obtido a estimação dos seguintes parâmetros.

Figura 3 – Parâmetros Estimados pela Regressão Logística

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Exp(Est)
Intercept	1	-1.1283	0.1345	70.34	<.0001	0.324
GRP_FICO_orig_time 1	1	0.5130	0.0231	492.39	<.0001	1.670
GRP_FICO_orig_time 2	1	0.2799	0.0280	99.79	<.0001	1.323
GRP_FICO_orig_time 3	1	0.0611	0.0316	3.73	0.0536	1.063
GRP_FICO_orig_time 4	1	-0.1332	0.0391	11.60	0.0007	0.875
GRP_LTV_orig_time 1	1	-0.3319	0.0484	46.94	<.0001	0.718
GRP_LTV_orig_time 2	1	-0.0230	0.0399	0.33	0.5651	0.977
GRP_LTV_orig_time 3	1	0.0599	0.0372	2.60	0.1072	1.062
GRP_LTV_orig_time 4	1	0.3003	0.0242	154.03	<.0001	1.350
GRP_LTV_time 1	1	-0.3716	0.1359	7.48	0.0062	0.690
GRP_LTV_time 2	1	-0.1856	0.1380	1.81	0.1785	0.831
GRP_LTV_time 3	1	-0.00296	0.1376	0.00	0.9829	0.997
GRP_LTV_time 4	1	0.2245	0.1358	2.73	0.0982	1.252
GRP_mat_time 1	1	-0.9878	0.0322	939.06	<.0001	0.372
GRP_mat_time 2	1	-0.4444	0.0366	147.58	<.0001	0.641
GRP_mat_time 3	1	0.1663	0.0272	37.52	<.0001	1.181
GRP_mat_time 4	1	0.5293	0.0304	302.93	<.0001	1.698

Fonte – Elaborada pelo autor.

A figura acima representa a saída do processo de estimação da máxima verossimilhança e os β 's estimados que maximizam a verossimilhança. Além disso foi aplicado o teste de hipótese de Wald, que a estatística é apresentada na coluna *Wald-Chi-Square* e seu p-valor na coluna *PR > ChiSq*. Segue abaixo a explicação variáveis finalistas do modelo:

- **GRP_FICO_orig_time**: é o escore de crédito, fornecido por uma agência de *bureau* de crédito conhecida como FICO, analisando o sinal e valores dos β 's de cada categoria é possível observar que quanto maior o escore menor é a probabilidade de descumprimento, que é um comportamento esperado para um escore de crédito
- **GRP_LTV_orig_time**: LTV é uma métrica que analise o valor do empréstimo, ou saldo devedor, dividido pelo valor total do imóvel. Esta métrica representa o percentual do imóvel que o cliente esta tomando emprestado, entretanto esta variável é constante desde a origem do empréstimo. Analisando o sinal e valores dos β 's de cada categoria é possível observar que quanto maior o LTV maior é a probabilidade de descumprimento, que também é o comportamento esperado, pois implica que o cliente esta muito alavancado e quase toda o valor do imóvel teve que ser financiado.
- **GRP_LTV_time**: é a mesma métrica de LTV explicada no item anterior, entretanto esta não é constante e é o valor do LTV atualizado para a data de referência indicando a alavancagem atual do contrato. Assim como a métrica anterior apresenta correlação positiva com a PD, portanto um comportamento esperado, ou seja, clientes alavancados tem maior probabilidade de descumprimento.
- **GRP_mat_time**: é o tempo de vencimento do contrato atualizado na data de referência e também apresenta correlação positiva com a probabilidade de descumprimento, indicando que contratos com maior prazo apresentam maior risco de descumprimento.

De maneira geral as variáveis finalistas apresentam um comportamento alinhado com suas definições, agora será importante avaliar o poder preditivo do modelo.

O poder de classificação de uma Regressão Logística pode ser medido por diversas variáveis, como: Gini, KS e ROC. Conforme sugerido por Siddiqi (2006). O KS obtido pela amostra de validação e treinamento foi no valor de 40.5 pontos o que segundo Picinini, Oliveira e Monteiro (2003) é considerado um bom KS.

Figura 4 – Estatísticas de Ajuste

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation
TARGET		_AIC_	Akaike's Information Criterion	31863.82	
TARGET		_ASE_	Average Squared Error	0.171382	0.172689
TARGET		_AVERR_	Average Error Function	0.512971	0.516936
TARGET		_DFE_	Degrees of Freedom for Error	31008	
TARGET		_DFM_	Model Degrees of Freedom	17	
TARGET		_DFT_	Total Degrees of Freedom	31025	
TARGET		_DIV_	Divisor for ASE	62050	26592
TARGET		_ERR_	Error Function	31829.82	13746.36
TARGET		_FPE_	Final Prediction Error	0.17157	
TARGET		_MAX_	Maximum Absolute Error	0.971815	0.971815
TARGET		_MSE_	Mean Square Error	0.171476	0.172689
TARGET		_NOBS_	Sum of Frequencies	31025	13296
TARGET		_NW_	Number of Estimate Weights	17	
TARGET		_RASE_	Root Average Sum of Squares	0.413983	0.415559
TARGET		_RFPE_	Root Final Prediction Error	0.41421	
TARGET		_RMSE_	Root Mean Squared Error	0.414096	0.415559
TARGET		_SBC_	Schwarz's Bayesian Criterion	32005.65	
TARGET		_SSE_	Sum of Squared Errors	10634.24	4592.144
TARGET		_SUMW_	Sum of Case Weights Times Freq	62050	26592
TARGET		_MISC_	Misclassification Rate	0.265721	0.266772

Fonte – Elaborada pelo autor.

As métricas acima também foram geradas pela ferramenta estatística utilizada, e dentre elas temos o destaque para o *Misclassification Rate* que seria a taxa de erro que está em torno de 26% o que implica em uma taxa de acerto de mais de 74% que também segundo Picinini, Oliveira e Monteiro (2003) e Selau e Ribeiro (2009) é considerada uma taxa de acerto de um bom modelo de *credit scoring*.

Figura 5 – Curva ROC

Fonte – Elaborada pelo autor.

Acima temos a curva de ROC que é conhecida como curva de *Lorenz* e a partir dela obtemos o índice Gini e o índice ROC que também podem ser utilizados para avaliar a qualidade de ajuste. Entretanto, o objetivo desta figura foi comparar a curva de *Lorenz* para diferentes amostras e observar que elas se comportam de maneira semelhante, o que garante que o modelo não teve um super ajuste apenas para base de treinamento, o que evita o *overfitting*, garantindo

que seu modelo não tem viés de seleção e se comporta bem apenas para a base utilizada para estimação da máxima verossimilhança.

Figura 6 – Índices de KS e Gini entre as amostras

Train: Gini Coefficient	Valid: Gini Coefficient	Train: Kolmogorov-Smirnov Statistic	Valid: Kolmogorov-Smirnov Statistic
0.516	0.506	0.405	0.405

Fonte – Elaborada pelo autor.

A figura acima é apenas o cálculo de KS aberto para amostra de treinamento e validação, assim conseguimos observar que o modelo apresentou bons resultados para ambas amostras.

4.6.2 Modelo ARIMAX

O grande objetivo deste trabalho é apresentar uma alternativa para inclusão de um fator *forward looking* dentro do tradicional Modelo Logístico, então com a taxa agregada de descumprimento foi estimado um modelo *ARIMAX* pois a ideia é justamente utilizar variáveis explicativas que sejam macroeconômicas para explicar a taxa de descumprimento do portfólio combinadas ao risco idiossincrático modelado pelo Modelo Logístico, assim dentro da estimação da Probabilidade de Descumprimento de cada contrato, pode ser colocado um fator de forecast (modelo de série de tempo) baseado em um ou mais cenários econômicos projetados, com isto teremos um modelo de PD *forward looking* que transformará o cálculo de Perda Esperada em um modelo de provisionamento projetado, mesmo com este fator projetado o seu ajuste será tão bom quanto a qualidade do modelo de série temporal desenvolvido, além disso este modelo ainda irá refletir a sensibilidade do portfólio estudado aos fatores macroeconômicos analisados e disponíveis no histórico garantindo uma melhor projeção de perda esperada baseada nos cenários econômicos capacidade de proteção e imunidade do portfólio analisado, assim será atendido o objetivo do IFRS9 que é um modelo de provisionamento bem acurado que conterá perdas ocorridas e projetadas.

A estimação de um modelo de série de tempo pode ser sofisticada e feitas em vários passos, dentro deste trabalho este processo foi simplificado pois sofisticar o modelo de série de tempo não era o objetivo deste trabalho.

4.6.2.1 Decomposição da Série de Taxa de Descumprimento

Dentro da mesma base de dados utilizada para a modelagem de PD foi calculada a taxa de descumprimento agregada do portfólio para cada mês de referência, esta foi transformada e decomposta para a observação de seus fatores *autoregressivos* e *choques ou média móvel*. Este trabalho é fundamental para avaliar quais as possíveis ordens do modelo ARIMA vai ser utilizada. Com isto temos o seguinte resultado: Dentro da figura acima observamos o primeiro

Figura 7 – Decomposição e Avaliação da Série

Fonte – Elaborada pelo autor.

gráfico, superior esquerdo, é a própria taxa de descumprimento observada ao longo dos 60 meses de observação. Depois disso o próprio pacote estatístico (*SAS ETS*) decompõe a série em taxas de retorno gerando três gráficos importantes para a modelagem de séries que são:

- ACF: função de autocorrelação
- PACF: função de autocorrelação parcial
- IACF: função inversa de autocorrelação

A ideia de observar estes gráficos é identificar a ordem do fator autoregressivo através da ACF e para descobrirmos a ordem do componente de média móvel é necessário a PACF. A ideia é que com esta decomposição visualizamos o intervalo de confiança de cada ordem das funções e para cada nível ou *lag* analisamos quais ficam fora do intervalo de confiança, com isto sugerimos para a estimação do modelo arimax, em seu componente ARMA a ordem 1 para o componente MA, e ordem 4 para o componente AR.

O modelo ARIMAX também contém outras séries como variáveis explicativas, como já foi mencionado a ideia é ajustar um modelo mais simples. Então foi será analisado a correlação

entre a série de taxa de descumprimento e as outras séries das variáveis macroeconômicas dentro da base de dados.

Figura 8 – Análise de Correlação Linear - Pearson

Pearson Correlation Coefficients, N = 60								
Prob > r under H0: Rho=0								
	time	max_hpi	max_gdp	max_ir	max_uer	Default	QTD	Y_t
time	1.00000	0.32053	-0.20482	0.17404	0.62239	0.26400	0.41856	-0.00372
		0.0125	0.1164	0.1835	<.0001	0.0415	0.0009	0.9775
max_hpi	0.32053	1.00000	0.15556	0.44436	-0.20039	0.23017	0.68028	-0.30123
	0.0125		0.2353	0.0004	0.1247	0.0769	<.0001	0.0193
max_gdp	-0.20482	0.15556	1.00000	-0.14008	-0.35917	-0.63546	-0.35524	-0.75942
	0.1164	0.2353		0.2857	0.0048	<.0001	0.0053	<.0001
max_ir	0.17404	0.44436	-0.14008	1.00000	0.14098	0.39865	0.57143	0.15475
	0.1835	0.0004	0.2857		0.2826	0.0016	<.0001	0.2378
max_uer	0.62239	-0.20039	-0.35917	0.14098	1.00000	0.22846	0.09845	0.36101
	<.0001	0.1247	0.0048	0.2826		0.0791	0.4542	0.0046
Default	0.26400	0.23017	-0.63546	0.39865	0.22846	1.00000	0.82009	0.78562
	0.0415	0.0769	<.0001	0.0016	0.0791		<.0001	<.0001
QTD	0.41856	0.68028	-0.35524	0.57143	0.09845	0.82009	1.00000	0.37333
	0.0009	<.0001	0.0053	<.0001	0.4542	<.0001		0.0033
Y_t	-0.00372	-0.30123	-0.75942	0.15475	0.36101	0.78562	0.37333	1.00000
	0.9775	0.0193	<.0001	0.2378	0.0046	<.0001	0.0033	

Fonte – Elaborada pelo autor.

Neste caso foi executado o cálculo da matriz de correlação entre a taxa de descumprimento e as outras variáveis macroeconômicas, foi utilizado a métrica de correlação de Pearson que considera apenas a correlação linear. Nesta análise se destacou a correlação entre a taxa de default e a variação do PIB representada pela variável max_gdp, que apresentam uma correlação negativa de 75%. Fato teoricamente sustentado, pois quando a taxa de default sobe o crescimento do PIB tende a ser menor, economicamente é esperado que nos momentos de crescimento econômico a taxa de descumprimento decresça.

O ajuste do modelo ARIMAX estará bem sucedido se for possível encontrar no mínimo uma outra variável explicativa, portanto foi executado uma rotina de estimação dos parâmetros considerando o modelo ARMA(4,1) e a inclusão da co-variável de crescimento de PIB. Esta estimação foi viabilizada através da ferramenta SAS Enterprise Guide e utilização de um procedimento de máxima verossimilhança, obtendo os seguintes resultados:

Figura 9 – Parâmetros do modelo ARIMAX

Model for variable Y_t	
Estimated Intercept	0.025831
Autoregressive Factors	
Factor 1:	1 - 1.36723 B**(1) + 0.01354 B**(2) + 0.1838 B**(3) + 0.20988 B**(4)
Moving Average Factors	
Factor 1:	1 - 1 B**(1)
Input Number 1	
Input Variable	max_gdp
Overall Regression Factor	-0.00238

Fonte – Elaborada pelo autor.

Na próxima figura vemos as estatísticas de ajuste obtidas, em que destacamos as estatísticas AIC, conhecido por *Akaike's Information Criteria* e o SBC conhecido como *Schwarz's Bayesian Criteria*, que são utilizados para comparar modelos estimados e quanto menor seus valores melhor é o modelo, entretanto foi estimado apenas um modelo, portanto não temos como comparar este ajuste e não é a ideia do trabalho. Também não existem referências para estes números como existem para a regressão logística relacionado aos indicadores de acerto e KS, além desta destacamos a estimativa do desvio padrão do erro, representada por *Std Error Estimate* que está bem baixa, indicando pouca variabilidade do erro o que é um objetivo da modelagem de série de tempo, pois a premissa é que depois dos devidos ajustes não exista variabilidade dos erros e ausência de autocorrelação.

Figura 10 – Estatísticas

Constant Estimate	0.001033
Variance Estimate	0.000021
Std Error Estimate	0.004591
AIC	-469.209
SBC	-454.548
Number of Residuals	60

Fonte – Elaborada pelo autor.

CONCLUSÕES

Este capítulo apresenta a conclusão da estimação dos modelos, analisando a série de tempo estimada pelo modelo ARIMAX, a taxa média de PD estimada pela regressão logística aplicada em cada contrato e por final a combinação entre a regressão e o modelo ARIMAX proposta pelo capítulo anterior.

5.1 Conclusões Gerais

Depois dos modelos estimados e explicados no capítulo anterior temos os seguintes resultados:

Figura 11 – Ajuste Final Modelo Foward Looking

Fonte – Elaborada pelo autor.

Analisando os resultados apresentados no gráfico acima é possível verificar um conjunto de conclusões:

1. O Modelo de PD: O modelo apresentou métricas de ordenamento e classificações boas, e a comparação entre as curvas ROC da amostra de treinamento e validação garantem a estabilidade de seu poder de classificação, no gráfico de séries está representado pela curva chamada de "PD Média" que é basicamente a média simples da escoragem com o modelo logístico de cada contrato da base de dados, portanto seria a previsão do Evento de Descumprimento para os próximos 12 meses. Este modelo justamente por levar em conta as características idiossincráticas subestima o evento de estresse observado, onde a curva de descumprimento observada atinge seu valor máximo. E este ponto é uma materialização da principal crítica da utilização de modelos econométricos sem fatores de projeção. Basiléia torna compulsório a estimação do parâmetro de PD considerando um ciclo econômico completo e uma visão já estressada deste parâmetro, entretanto mesmo se isto tivesse sido realizado o modelo ainda sim apresentaria um valor máximo menor que a taxa de descumprimento observada no momento de estresse, comprometendo sua acurácia na visão do portfólio.
2. O Modelo ARIMAX: Como o objetivo deste trabalho é apresentar uma forma de integrar um modelo de série de tempo dentro do processo tradicional de um modelo de classificação (regressão logística) considerando a limitação obter um conjunto maior de base de dados reais e com mais variáveis macroeconômicas, não foi possível fazer um bom modelo de série de tempo, embora este apresente uma boa correlação com a taxa observada que é visível no gráfico, mas também é possível notar uma aparente evidência de sazonalidade da taxa observada, que se tivesse sido capturada pelo modelo teria gerado uma melhor estimativa. De qualquer maneira, considerando o nível agregado de análise de séries, este modelo apresenta uma acurácia melhor no processo de risco de crédito. Esta é uma boa evidência ao fato de que muitos bancos pequenos e com poucos contratos preferem realizar estimativas alto nível (em nível agregado) para fins regulatórios. A desvantagem desta escolha é que estes modelos não ajudam a separar e ordenar o risco individual (idiossincrático) das contrapartes, gerando políticas de crédito com uma abordagem menos quantitativa e mais qualitativa. No gráfico este modelo é representado pela curva "PREVISTO 12 meses".
3. O Modelo de PD Forward Looking: O Modelo apresentou um melhor resultado do que o modelo tradicional, e claramente corrigiu a subestimação da taxa de descumprimento no momento de estresse. Além disso, se adequou melhor ao comportamento da taxa de descumprimento observada. Foi possível verificar que este modelo aproveitou as vantagens dos dois tipos de modelos individuais, ou seja, apresentou uma boa adequação ao histórico observado de taxa de descumprimento e ainda possui as vantagens do bom modelo de classificação, até porque nesta metodologia não são alterados nenhum dos parâmetros da

regressão logística. Portanto este modelo pode servir como um bom forecast de risco e ainda uma ferramenta de classificação individual de risco de crédito. De qualquer maneira, poderia ter apresentado um resultado maior se tivesse sido utilizada uma base com mais fatores macroeconômicos e maiores iterações do processo de estimação, o que não era o objetivo deste trabalho.

Em termos gerais, o objetivo do trabalho foi atingido e apresentou uma boa abordagem para a combinação de modelos, utilizando uma estimativa de *forecast* e outro modelo de classificação, esta abordagem poder servir como alternativa para as instituições financeiras atenderem as exigências do IFRS9 reaproveitando os modelos de regressão logística existentes, ou mantendo o processo tradicional de estimação. Além disso este modelo pode servir inclusive para a modelagem do estresse de uma carteira através dos fatores macroeconômicos que estão diretamente isolados neste modelo. Entretanto, quando este for o objetivo será necessário avaliar a possibilidade de segmentação do portfólio, pois provavelmente cada *cluster* de um portfólio pode ser afetado de maneira distinta pelo estresse do fator macroeconomico. Além disso, a escolha do modelo ARIMAX foi uma opção adequada ao objetivo deste trabalho, a recomendação é que ao ser realizada uma abordagem com o objetivo de encontrar a melhor acurácia será importante discutir qual a melhor técnica para o modelo de forecast.

REFERÊNCIAS

ALTMAN, E. I. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. **The Journal of Finance**, John Wiley and Sons Ltd, p. 589–609, 1968. Disponível em: <<http://dx.doi.org/10.1111/j.1540-6261.1968.tb00843.x>>. Citado na página 33.

BAESENS, B.; ROESCH, D.; SCHEULE, H. **Credit Risk Analytics: Measurement, Techniques, Applications, and Examples in SAS**. 1. ed. Wiley, 2016. ISBN 9781119143987. Disponível em: <<http://www.creditriskanalytics.net/>>. Citado na página 36.

BESSIS, J. **Risk Managament in Banking**. 4. ed. Wiley, 2015. ISBN 9781118660218. Disponível em: <<http://www.wiley.com/WileyCDA/WileyTitle/productCd-1118660218.html>>. Citado na página 21.

BLACK, F.; SCHOLES, M. The pricing of options and corporate liabilities. **The Journal of Political Economy**, p. 637–654, 1973. Disponível em: <<http://www.jstor.org/stable/1831029>>. Citado na página 27.

BOHN, J. R.; STEIN, R. M. **Active Credit Portfolio Management in Practice**. 1. ed. Wiley, 2009. ISBN 9781118660218. Disponível em: <<https://www.amazon.com/Active-Credit-Portfolio-Management-Practice/dp/0470080183>>. Citado na página 27.

JARROW, R. A.; LANDO, D.; TURNBULL, S. M. A markov model for the term structure of credit risk spreads. **The Review of Financial Studies Summer**, The Review of Financial Studies, p. 481–523, 1997. Citado na página 32.

KING, G.; ZENG, L. Logistic regression in rare events data. **Society for Political Methodology**, n. 9, p. 137–163, 2001. Citado nas páginas 19 e 40.

LANDO, D. **Credit Risk Modeling : Theory and Applications**. 1. ed. Princeton University Press, 2004. ISBN 9780691089294. Disponível em: <<http://press.princeton.edu/titles/7726.html>>. Citado na página 29.

MERTON, R. C. On the pricing of corporate debt: The risk structure of interest rates. **The Journal of Finance**, Blackwell Publishing Ltd, p. 449–470, 1974. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.1111/j.1540-6261.1974.tb03058.x/abstract>>. Citado na página 28.

_____. Applications of option-pricing theory: Twenty-five years later. **Nobel Lectures**, Torsten Persson, World Scientific Publishing Co., p. 449–470, 1997. Disponível em: <http://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/1997/merton-lecture.html>. Citado na página 29.

PASCHOARELLI, R. **Probabilidade de Default**. 1. ed. Saint Paul, 2007. ISBN 9788598838380. Disponível em: <<http://www.saintpaul.com.br/livro-probabilidade-default-arvores-binomiais-1155.aspx>>. Citado na página 31.

PICININI, R.; OLIVEIRA, G. M. B.; MONTEIRO, L. H. A. Mineração de critério de credit scoring utilizando algoritmos genéticos. In: **VI Simpósio Brasileiro de Automação Inteligente**. [S.l.]: Anais do VI Simpósio Brasileiro de Automação Inteligente, 2003. (SABI '03), p. 463–466. ISBN 1-58113-534-3. Citado na página 46.

SICSU, A. L. **Credit Scoring**. 1. ed. Blucher, 2010. ISBN 9788521205333. Disponível em: <<https://www.blucher.com.br/livro/detalhes/credit-scoring-842>>. Citado na página 33.

SIDDIQI, N. **Credit Risk Scorecards**. 1. ed. Wiley, 2006. ISBN 9780471754510. Disponível em: <<http://onlinelibrary.wiley.com/book/10.1002/9781119201731>>. Citado nas páginas 43 e 46.

VISCONTI, M. F. **Modelagem de Capital Econômico no varejo conforme as diretrizes de Basiléia II**. 2011. Citado na página 33.

WILSON, T. C. Credit portfolio risk(i). **Risk Magazine**, Risk Magazine, 1997a. Citado na página 41.

SCRIPTS SAS UTILIZADOS PARA MANIPULAÇÃO DE DADOS

Código-fonte 1 – Código SAS para preparar a base de Mortgage para a modelagem de PD

```
1:
2: libname mecai 'C:\Arquivos Luiz\MECAI\Projeto\Projeto_vfinal\
   Modelagem Estatística\datasets';
3: data mecai.ABT_0;
4: set mecai.mortgage;
5: run;
6:
7: /*Separa Base de Contratos*/
8: data contratos;
9: set mecai.ABT_0(keep=id);
10: run;
11:
12: proc sort data=contratos out=contratos_nd nodupkey;
13: by id;
14: run;
15:
16: %let meses =
17: (
18: 1 2 3 4 5 6 7 8 9 10 11 12
19: 13 14 15 16 17 18 19 20 21
20: 22 23 24 25 26 27 28 29 30
21: 31 32 33 34 35 36 37 38 39
22: 40 41 42 43 44 45 46 47 48
23: 49 50 51 52 53 54 55 56 57
```

```
24: 58 59 60
25: );
26:
27: %macro separa_safra;
28: %do i=1 %to 60;
29: data safra_%scan(&meses,&i)
30: ( KEEP= id default_time rename=default_time=default_%scan(&
      meses,&i));
31: set mecai.ABT_0;
32: if time eq %scan(&meses,&i);
33: ;
34: run;
35: proc sort data= safra_%scan(&meses,&i);
36: by id;
37: run;
38:
39: %end;
40: %mend;
41: %separa_safra;
42:
43: %macro painel;
44: data default_horizonte;
45: merge contratos_nd (in=a)
46: %do i=1 %to 60;
47: safra_%scan(&meses,&i)
48: %end;;
49: by id;
50: if a=1;
51: run;
52: %mend;
53: %painel;
54:
55:
56: data auxiliar;
57: set mecai.ABT_0(keep= id time);
58: run;
59:
60: proc sort data=auxiliar;
61: by id;
62: run;
63:
64: data default_desl_0;
```

```

65: merge auxiliar(in=a) default_horizonte(in=b);
66: by id;
67: run;
68:
69: %macro Deslocamento;
70: data default_desl_1(keep= id time D_default_1-D_default_12
      where=(time <=48));
71: set default_desl_0;
72: %do i=1 %to 48;
73: if time EQ %scan(&meses,&i) then
74: do;
75: %do j=%eval(&i ) %to %eval(&i + 12);
76: D_default_%eval(&j-&i) = default_%scan(&meses,&j) ;
77:
78: %end;
79: end;
80: %end;
81: run;
82: %mend;
83: %Deslocamento;
84:
85: /*Marcação de Default em 12 meses*/
86: data mecai.TARGET_12(keep= id time TARGET);
87: set default_desl_1(where=(missing(D_default_1) eq 0));
88: TARGET=
89: MAX(.,
90: D_default_1,
91: D_default_2,
92: D_default_3,
93: D_default_4,
94: D_default_5,
95: D_default_6,
96: D_default_7,
97: D_default_8,
98: D_default_9,
99: D_default_10,
100: D_default_11,
101: D_default_12
102: );
103: run;
104:
105: /*Vou seleccionar um contrato apenas uma única vez,*/

```

```
106: /*evitando o contrato aparecer como default e não default*/
107: data mecai.sample;
108: set mecai.TARGET_12;
109: selecao=ranuni(123);
110: run;
111:
112: proc sort data=mecai.sample;
113: by id descending selecao ;
114: run;
115:
116: proc sort data=mecai.sample out=mecai.sample_final nodupkey;
117: by id ;
118: run;
119:
120: /*Buscando as variáveis explicativas*/
121: proc sort data= mecai.sample_final;
122: by id time;
123: run;
124:
125: proc sort data= mecai.ABT_0;
126: by id time;
127: run;
128:
129: data mecai.ABT_final;
130: merge mecai.sample_final(in=a) mecai.ABT_0(in=b);
131: by id time;
132: if a=1;
133: run;
```

Código-fonte 2 – Código SAS analisar Série de Tempo

```
1: libname mecai 'C:\Arquivos Luiz\MECAI\Projeto\Projeto_vfinal\
   Modelagem Estatística\datasets';
2: data mecai.mortgage;
3: set mortgage;
4: run;
5:
6: proc sql;
7: create table mecai.TIMESERIE_0
8: as select
9: time,
10: max(hpi_time) as max_hpi,
```

```
11: max(gdp_time) as max_gdp,
12: max(interest_rate_time) as max_ir,
13: max(uer_time) as max_uer,
14: sum(default_time) as Default,
15: count(*) as QTD
16: from mecai.mortgage
17: group by 1;
18: quit;
19:
20: data mecai.TIMESERIE_1;
21: set mecai.TIMESERIE_0;
22: Y_t=Default/QTD; /*Taxa de Default em t*/
23: run;
24:
25: proc sgplot data=mecai.TIMESERIE_1;
26: scatter y=Y_t x=time;
27: run;
28:
29: proc arima data=mecai.TIMESERIE_1 ;
30: identify var=Y_t nlag=48; /*temos 60 meses então para projetar
    12 meses podemos ter no máximo 48 lags*/
31: run;
32:
33:
34: /*AR(4) e MA(1) agora a ideia sera estimar o ARIMAX*/
35: proc arima data=mecai.TIMESERIE_1;
36: identify var=Y_t crosscorr=(max_hpi max_gdp max_ir max_uer)
    nlag=48;
37: estimate p=4 q=1 input=(max_hpi max_gdp max_ir max_uer) plot;
38: run;
39:
40: /*AR(4) e MA(1) agora a ideia sera estimar o ARIMAX*/
41: proc arima data=mecai.TIMESERIE_1;
42: identify var=Y_t crosscorr=( max_gdp ) nlag=48;
43: estimate p=4 q=1 input=( max_gdp ) plot;
44: run;
45:
46:
47: proc arima data=mecai.TIMESERIE_1;
48: identify var=Y_t nlag=48;
49: estimate p=4 q=1 plot;
50: run;
```

```

51:
52: proc corr data=mecai.TIMESERIE_1 outs=spearman1 outp=pearson1;
53: run;

```

Código-fonte 3 – Código SAS que contém o processo de escoragem através da regressão logística estimada para a PD

```

1:  data mecai.ABT_tm_pd;
2:  set mecai.ABT_0;
3:  *-----
   *;
4:  * EM SCORE CODE;
5:  *-----
   *;
6:  *-----
   *;
7:  * TOOL: Input Data Source;
8:  * TYPE: SAMPLE;
9:  * NODE: Ids;
10: *-----
   *;
11: *-----
   *;
12: * TOOL: Partition Class;
13: * TYPE: SAMPLE;
14: * NODE: Part;
15: *-----
   *;
16: *-----
   *;
17: * TOOL: Extension Class;
18: * TYPE: CREDSCORE;
19: * NODE: IGN;
20: *-----
   *;
21: length _UFormat $200;
22: drop _UFormat;
23: _UFormat='';
24:
25: *-----
   *;

```



```
26:  * Variable: FICO_orig_time;
27:  *-----
   *;
28:  LABEL GRP_FICO_orig_time = "Grouped: FICO_orig_time";
29:  LABEL WOE_FICO_orig_time = "Weight of Evidence: FICO_orig_
   time";
30:
31:  if MISSING(FICO_orig_time) then do;
32:  GRP_FICO_orig_time = 1;
33:  WOE_FICO_orig_time = -0.250393933;
34:  end;
35:  else if NOT MISSING(FICO_orig_time) then do;
36:  if FICO_orig_time < 647 then do;
37:  GRP_FICO_orig_time = 1;
38:  WOE_FICO_orig_time = -0.250393933;
39:  end;
40:  else
41:  if 647 <= FICO_orig_time AND FICO_orig_time < 685 then do;
42:  GRP_FICO_orig_time = 2;
43:  WOE_FICO_orig_time = -0.166099398;
44:  end;
45:  else
46:  if 685 <= FICO_orig_time AND FICO_orig_time < 718 then do;
47:  GRP_FICO_orig_time = 3;
48:  WOE_FICO_orig_time = 0.064655925;
49:  end;
50:  else
51:  if 718 <= FICO_orig_time AND FICO_orig_time < 744 then do;
52:  GRP_FICO_orig_time = 4;
53:  WOE_FICO_orig_time = 0.2779870221;
54:  end;
55:  else
56:  if 744 <= FICO_orig_time then do;
57:  GRP_FICO_orig_time = 5;
58:  WOE_FICO_orig_time = 0.8404936406;
59:  end;
60:  end;
61:
62:  *-----
   *;
63:  * Variable: Interest_Rate_orig_time;
64:  *-----
```

```

        *;
65:  LABEL GRP_Interest_Rate_orig_time = "Grouped: Interest_Rate_
      orig_time";
66:  LABEL WOE_Interest_Rate_orig_time = "Weight of Evidence:
      Interest_Rate_orig_time";
67:
68:  if MISSING(Interest_Rate_orig_time) then do;
69:  GRP_Interest_Rate_orig_time = 5;
70:  WOE_Interest_Rate_orig_time = -0.390108296;
71:  end;
72:  else if NOT MISSING(Interest_Rate_orig_time) then do;
73:  if Interest_Rate_orig_time < 3 then do;
74:  GRP_Interest_Rate_orig_time = 1;
75:  WOE_Interest_Rate_orig_time = 0.0085605842;
76:  end;
77:  else
78:  if 3 <= Interest_Rate_orig_time AND Interest_Rate_orig_time <
      5.63 then do;
79:  GRP_Interest_Rate_orig_time = 2;
80:  WOE_Interest_Rate_orig_time = 1.1923864736;
81:  end;
82:  else
83:  if 5.63 <= Interest_Rate_orig_time AND Interest_Rate_orig_
      time < 6.13 then do;
84:  GRP_Interest_Rate_orig_time = 3;
85:  WOE_Interest_Rate_orig_time = 0.71293339;
86:  end;
87:  else
88:  if 6.13 <= Interest_Rate_orig_time AND Interest_Rate_orig_
      time < 6.63 then do;
89:  GRP_Interest_Rate_orig_time = 4;
90:  WOE_Interest_Rate_orig_time = 0.3059787096;
91:  end;
92:  else
93:  if 6.63 <= Interest_Rate_orig_time then do;
94:  GRP_Interest_Rate_orig_time = 5;
95:  WOE_Interest_Rate_orig_time = -0.390108296;
96:  end;
97:  end;
98:
99:  *-----
      *;

```

```

100:  * Variable: LTV_orig_time;
101:  *-----
    *;
102:  LABEL GRP_LTV_orig_time = "Grouped: LTV_orig_time";
103:  LABEL WOE_LTV_orig_time = "Weight of Evidence: LTV_orig_time
    ";
104:
105:  if MISSING(LTV_orig_time) then do;
106:  GRP_LTV_orig_time = 4;
107:  WOE_LTV_orig_time = -0.217846514;
108:  end;
109:  else if NOT MISSING(LTV_orig_time) then do;
110:  if LTV_orig_time < 66.2 then do;
111:  GRP_LTV_orig_time = 1;
112:  WOE_LTV_orig_time = 0.8233118654;
113:  end;
114:  else
115:  if 66.2 <= LTV_orig_time AND LTV_orig_time < 75 then do;
116:  GRP_LTV_orig_time = 2;
117:  WOE_LTV_orig_time = 0.3749861929;
118:  end;
119:  else
120:  if 75 <= LTV_orig_time AND LTV_orig_time < 80 then do;
121:  GRP_LTV_orig_time = 3;
122:  WOE_LTV_orig_time = 0.1842906019;
123:  end;
124:  else
125:  if 80 <= LTV_orig_time AND LTV_orig_time < 93.8 then do;
126:  GRP_LTV_orig_time = 4;
127:  WOE_LTV_orig_time = -0.217846514;
128:  end;
129:  else
130:  if 93.8 <= LTV_orig_time then do;
131:  GRP_LTV_orig_time = 5;
132:  WOE_LTV_orig_time = 0.0379449662;
133:  end;
134:  end;
135:
136:  *-----
    *;
137:  * Variable: LTV_time;
138:  *-----

```

```

    *;
139: LABEL GRP_LTV_time = "Grouped: LTV_time";
140: LABEL WOE_LTV_time = "Weight of Evidence: LTV_time";
141:
142: if MISSING(LTV_time) then do;
143: GRP_LTV_time = 5;
144: WOE_LTV_time = 0.3087061611;
145: end;
146: else if NOT MISSING(LTV_time) then do;
147: if LTV_time < 72.29 then do;
148: GRP_LTV_time = 1;
149: WOE_LTV_time = 1.0454945427;
150: end;
151: else
152: if 72.29 <= LTV_time AND LTV_time < 77.04 then do;
153: GRP_LTV_time = 2;
154: WOE_LTV_time = 0.4029253822;
155: end;
156: else
157: if 77.04 <= LTV_time AND LTV_time < 80.91 then do;
158: GRP_LTV_time = 3;
159: WOE_LTV_time = -0.006004584;
160: end;
161: else
162: if 80.91 <= LTV_time then do;
163: GRP_LTV_time = 4;
164: WOE_LTV_time = -0.631951597;
165: end;
166: end;
167:
168: *-----
    *;
169: * Variable: first_time;
170: *-----
    *;
171: LABEL GRP_first_time = "Grouped: first_time";
172: LABEL WOE_first_time = "Weight of Evidence: first_time";
173:
174: if MISSING(first_time) then do;
175: GRP_first_time = 1;
176: WOE_first_time = 1.2032581308;
177: end;

```

```

178:   else if NOT MISSING(first_time) then do;
179:     if first_time < 22 then do;
180:       GRP_first_time = 1;
181:       WOE_first_time = 1.2032581308;
182:     end;
183:   else
184:     if 22 <= first_time AND first_time < 24 then do;
185:       GRP_first_time = 2;
186:       WOE_first_time = 0.5050983018;
187:     end;
188:   else
189:     if 24 <= first_time AND first_time < 26 then do;
190:       GRP_first_time = 3;
191:       WOE_first_time = 0.0272462108;
192:     end;
193:   else
194:     if 26 <= first_time AND first_time < 28 then do;
195:       GRP_first_time = 4;
196:       WOE_first_time = -0.591294618;
197:     end;
198:   else
199:     if 28 <= first_time then do;
200:       GRP_first_time = 5;
201:       WOE_first_time = -0.837629652;
202:     end;
203:   end;
204:
205:   *-----
206:   *;
207:   * Variable: gdp_time;
208:   *-----
209:   *;
210:
211:   LABEL GRP_gdp_time = "Grouped: gdp_time";
212:   LABEL WOE_gdp_time = "Weight of Evidence: gdp_time";
213:
214:   if MISSING(gdp_time) then do;
215:     GRP_gdp_time = 1;
216:     WOE_gdp_time = -0.557900977;
217:   end;
218:   else if NOT MISSING(gdp_time) then do;
219:     if gdp_time < 1.23 then do;
220:       GRP_gdp_time = 1;

```

```
218:   WOE_gdp_time = -0.557900977;
219:   end;
220:   else
221:     if 1.23 <= gdp_time AND gdp_time < 2.36 then do;
222:       GRP_gdp_time = 2;
223:       WOE_gdp_time = -0.358699251;
224:     end;
225:     else
226:       if 2.36 <= gdp_time AND gdp_time < 2.85 then do;
227:         GRP_gdp_time = 3;
228:         WOE_gdp_time = 0.030006185;
229:       end;
230:       else
231:         if 2.85 <= gdp_time AND gdp_time < 3.28 then do;
232:           GRP_gdp_time = 4;
233:           WOE_gdp_time = 0.8279076463;
234:         end;
235:         else
236:           if 3.28 <= gdp_time then do;
237:             GRP_gdp_time = 5;
238:             WOE_gdp_time = 1.5391311139;
239:           end;
240:         end;
241:
242:         *-----
243:         *;
244:         * Variable: hpi_orig_time;
245:         *-----
246:         *;
247:         LABEL GRP_hpi_orig_time = "Grouped: hpi_orig_time";
248:         LABEL WOE_hpi_orig_time = "Weight of Evidence: hpi_orig_time";
249:
250:         if MISSING(hpi_orig_time) then do;
251:           GRP_hpi_orig_time = 5;
252:           WOE_hpi_orig_time = -0.682807322;
253:         end;
254:         else if NOT MISSING(hpi_orig_time) then do;
255:           if hpi_orig_time < 144.84 then do;
256:             GRP_hpi_orig_time = 1;
257:             WOE_hpi_orig_time = 0.8034826308;
258:           end;
```

```

257:  else
258:  if 144.84 <= hpi_orig_time AND hpi_orig_time < 191.42 then do
    ;
259:  GRP_hpi_orig_time = 2;
260:  WOE_hpi_orig_time = 1.2645204071;
261:  end;
262:  else
263:  if 191.42 <= hpi_orig_time AND hpi_orig_time < 208.86 then do
    ;
264:  GRP_hpi_orig_time = 3;
265:  WOE_hpi_orig_time = 0.6266249899;
266:  end;
267:  else
268:  if 208.86 <= hpi_orig_time AND hpi_orig_time < 219.67 then do
    ;
269:  GRP_hpi_orig_time = 4;
270:  WOE_hpi_orig_time = -0.063474561;
271:  end;
272:  else
273:  if 219.67 <= hpi_orig_time then do;
274:  GRP_hpi_orig_time = 5;
275:  WOE_hpi_orig_time = -0.682807322;
276:  end;
277:  end;
278:
279:  *-----
    *;
280:  * Variable: hpi_time;
281:  *-----
    *;
282:  LABEL GRP_hpi_time = "Grouped: hpi_time";
283:  LABEL WOE_hpi_time = "Weight of Evidence: hpi_time";
284:
285:  if MISSING(hpi_time) then do;
286:  GRP_hpi_time = 4;
287:  WOE_hpi_time = -0.252358704;
288:  end;
289:  else if NOT MISSING(hpi_time) then do;
290:  if hpi_time < 167.91 then do;
291:  GRP_hpi_time = 1;
292:  WOE_hpi_time = 0.0563660439;
293:  end;

```

```
294:   else
295:     if 167.91 <= hpi_time AND hpi_time < 186.91 then do;
296:       GRP_hpi_time = 2;
297:       WOE_hpi_time = -0.507805295;
298:     end;
299:   else
300:     if 186.91 <= hpi_time AND hpi_time < 212.73 then do;
301:       GRP_hpi_time = 3;
302:       WOE_hpi_time = 0.481283137;
303:     end;
304:   else
305:     if 212.73 <= hpi_time AND hpi_time < 223.75 then do;
306:       GRP_hpi_time = 4;
307:       WOE_hpi_time = -0.252358704;
308:     end;
309:   else
310:     if 223.75 <= hpi_time then do;
311:       GRP_hpi_time = 5;
312:       WOE_hpi_time = 0.4797311312;
313:     end;
314:   end;
315:
316:   *-----
317:   *;
318:   * Variable: interest_rate_time;
319:   *-----
320:   *;
321:
322:   LABEL GRP_interest_rate_time = "Grouped: interest_rate_time";
323:   LABEL WOE_interest_rate_time = "Weight of Evidence: interest_
324:     rate_time";
325:
326:   if MISSING(interest_rate_time) then do;
327:     GRP_interest_rate_time = 5;
328:     WOE_interest_rate_time = -0.419120244;
329:   end;
330:   else if NOT MISSING(interest_rate_time) then do;
331:     if interest_rate_time < 5.63 then do;
332:       GRP_interest_rate_time = 1;
333:       WOE_interest_rate_time = 1.1494237359;
334:     end;
335:   end;
336:   else
337:     if 5.63 <= interest_rate_time AND interest_rate_time < 6.13
```



```

        then do;
333:   GRP_interest_rate_time = 2;
334:   WOE_interest_rate_time = 0.7103311528;
335:   end;
336:   else
337:   if 6.13 <= interest_rate_time AND interest_rate_time < 6.75
        then do;
338:   GRP_interest_rate_time = 3;
339:   WOE_interest_rate_time = 0.3320712269;
340:   end;
341:   else
342:   if 6.75 <= interest_rate_time AND interest_rate_time < 7.02
        then do;
343:   GRP_interest_rate_time = 4;
344:   WOE_interest_rate_time = -0.101842842;
345:   end;
346:   else
347:   if 7.02 <= interest_rate_time then do;
348:   GRP_interest_rate_time = 5;
349:   WOE_interest_rate_time = -0.419120244;
350:   end;
351:   end;
352:
353:   *-----
        *;
354:   * Variable: mat_time;
355:   *-----
        *;
356:   LABEL GRP_mat_time = "Grouped: mat_time";
357:   LABEL WOE_mat_time = "Weight of Evidence: mat_time";
358:
359:   if MISSING(mat_time) then do;
360:   GRP_mat_time = 1;
361:   WOE_mat_time = 1.0753588915;
362:   end;
363:   else if NOT MISSING(mat_time) then do;
364:   if mat_time < 140 then do;
365:   GRP_mat_time = 1;
366:   WOE_mat_time = 1.0753588915;
367:   end;
368:   else
369:   if 140 <= mat_time AND mat_time < 142 then do;

```

```
370:   GRP_mat_time = 2;
371:   WOE_mat_time = 0.4140399742;
372:   end;
373:   else
374:   if 142 <= mat_time AND mat_time < 145 then do;
375:   GRP_mat_time = 3;
376:   WOE_mat_time = -0.308469587;
377:   end;
378:   else
379:   if 145 <= mat_time AND mat_time < 147 then do;
380:   GRP_mat_time = 4;
381:   WOE_mat_time = -0.825487454;
382:   end;
383:   else
384:   if 147 <= mat_time then do;
385:   GRP_mat_time = 5;
386:   WOE_mat_time = -0.975371939;
387:   end;
388:   end;
389:
390:   *-----
391:   *;
392:   * Variable: orig_time;
393:   *-----
394:   *;
395:   LABEL GRP_orig_time = "Grouped: orig_time";
396:   LABEL WOE_orig_time = "Weight of Evidence: orig_time";
397:
398:   if MISSING(orig_time) then do;
399:   GRP_orig_time = 1;
400:   WOE_orig_time = 1.0913370812;
401:   end;
402:   else if NOT MISSING(orig_time) then do;
403:   if orig_time < 20 then do;
404:   GRP_orig_time = 1;
405:   WOE_orig_time = 1.0913370812;
406:   end;
407:   else
408:   if 20 <= orig_time AND orig_time < 22 then do;
409:   GRP_orig_time = 2;
410:   WOE_orig_time = 0.3490811284;
411:   end;
```

```

410:  else
411:    if 22 <= orig_time AND orig_time < 24 then do;
412:      GRP_orig_time = 3;
413:      WOE_orig_time = -0.127022273;
414:    end;
415:  else
416:    if 24 <= orig_time AND orig_time < 26 then do;
417:      GRP_orig_time = 4;
418:      WOE_orig_time = -0.488913041;
419:    end;
420:  else
421:    if 26 <= orig_time then do;
422:      GRP_orig_time = 5;
423:      WOE_orig_time = -0.921963952;
424:    end;
425:  end;
426:
427:  *-----
      *;
428:  * Variable: uer_time;
429:  *-----
      *;
430:  LABEL GRP_uer_time = "Grouped: uer_time";
431:  LABEL WOE_uer_time = "Weight of Evidence: uer_time";
432:
433:  if MISSING(uer_time) then do;
434:    GRP_uer_time = 2;
435:    WOE_uer_time = -0.119681092;
436:  end;
437:  else if NOT MISSING(uer_time) then do;
438:    if uer_time < 4.7 then do;
439:      GRP_uer_time = 1;
440:      WOE_uer_time = -0.407889771;
441:    end;
442:  else
443:    if 4.7 <= uer_time AND uer_time < 5.2 then do;
444:      GRP_uer_time = 2;
445:      WOE_uer_time = -0.119681092;
446:    end;
447:  else
448:    if 5.2 <= uer_time AND uer_time < 5.8 then do;
449:      GRP_uer_time = 3;

```

```

450:   WOE_uer_time = 1.4947854522;
451:   end;
452:   else
453:   if 5.8 <= uer_time AND uer_time < 8.3 then do;
454:   GRP_uer_time = 4;
455:   WOE_uer_time = -0.402271082;
456:   end;
457:   else
458:   if 8.3 <= uer_time then do;
459:   GRP_uer_time = 5;
460:   WOE_uer_time = 0.1625346129;
461:   end;
462:   end;
463:
464:   *-----
      *;
465:   * Special Code Values
466:   *-----
      *;
467:   *-----
      *;
468:   * TOOL: Regression;
469:   * TYPE: MODEL;
470:   * NODE: Reg;
471:   *-----
      *;
472:   *****;
473:   *** begin scoring code for regression;
474:   *****;
475:
476:   length _WARN_ $4;
477:   label _WARN_ = 'Warnings' ;
478:
479:   length I_TARGET $ 12;
480:   label I_TARGET = 'Into: TARGET' ;
481:   *** Target Values;
482:   array REGDRF [2] $12 _temporary_ ('1' '0' );
483:   label U_TARGET = 'Unnormalized Into: TARGET' ;
484:   *** Unnormalized target values;
485:   ARRAY REGDRU[2] _TEMPORARY_ (1 0);
486:
487:   drop _DM_BAD;

```

```

488:  _DM_BAD=0;
489:
490:  *** Generate dummy variables for GRP_FICO_orig_time ;
491:  drop _1_0 _1_1 _1_2 _1_3 ;
492:  *** encoding is sparse, initialize to zero;
493:  _1_0 = 0;
494:  _1_1 = 0;
495:  _1_2 = 0;
496:  _1_3 = 0;
497:  if missing( GRP_FICO_orig_time ) then do;
498:  _1_0 = .;
499:  _1_1 = .;
500:  _1_2 = .;
501:  _1_3 = .;
502:  substr(_warn_,1,1) = 'M';
503:  _DM_BAD = 1;
504:  end;
505:  else do;
506:  length _dm12 $ 12; drop _dm12 ;
507:  _dm12 = put( GRP_FICO_orig_time , BEST12. );
508:  %DMNORMIP( _dm12 )
509:  _dm_find = 0; drop _dm_find;
510:  if _dm12 <= '3' then do;
511:  if _dm12 <= '2' then do;
512:  if _dm12 = '1' then do;
513:  _1_0 = 1;
514:  _dm_find = 1;
515:  end;
516:  else do;
517:  if _dm12 = '2' then do;
518:  _1_1 = 1;
519:  _dm_find = 1;
520:  end;
521:  end;
522:  end;
523:  else do;
524:  if _dm12 = '3' then do;
525:  _1_2 = 1;
526:  _dm_find = 1;
527:  end;
528:  end;
529:  end;

```

```
530:   else do;
531:     if _dm12 = '4' then do;
532:       _1_3 = 1;
533:       _dm_find = 1;
534:     end;
535:   else do;
536:     if _dm12 = '5' then do;
537:       _1_0 = -1;
538:       _1_1 = -1;
539:       _1_2 = -1;
540:       _1_3 = -1;
541:       _dm_find = 1;
542:     end;
543:   end;
544: end;
545: if not _dm_find then do;
546:   _1_0 = .;
547:   _1_1 = .;
548:   _1_2 = .;
549:   _1_3 = .;
550:   substr(_warn_,2,1) = 'U';
551:   _DM_BAD = 1;
552: end;
553: end;
554:
555: *** Generate dummy variables for GRP_LTV_orig_time ;
556: drop _2_0 _2_1 _2_2 _2_3 ;
557: *** encoding is sparse, initialize to zero;
558: _2_0 = 0;
559: _2_1 = 0;
560: _2_2 = 0;
561: _2_3 = 0;
562: if missing( GRP_LTV_orig_time ) then do;
563:   _2_0 = .;
564:   _2_1 = .;
565:   _2_2 = .;
566:   _2_3 = .;
567:   substr(_warn_,1,1) = 'M';
568:   _DM_BAD = 1;
569: end;
570: else do;
571:   length _dm12 $ 12; drop _dm12 ;
```

```

572:  _dm12 = put( GRP_LTV_orig_time , BEST12. );
573:  %DMNORMIP( _dm12 )
574:  if _dm12 = '4' then do;
575:    _2_3 = 1;
576:  end;
577:  else if _dm12 = '3' then do;
578:    _2_2 = 1;
579:  end;
580:  else if _dm12 = '2' then do;
581:    _2_1 = 1;
582:  end;
583:  else if _dm12 = '5' then do;
584:    _2_0 = -1;
585:    _2_1 = -1;
586:    _2_2 = -1;
587:    _2_3 = -1;
588:  end;
589:  else if _dm12 = '1' then do;
590:    _2_0 = 1;
591:  end;
592:  else do;
593:    _2_0 = .;
594:    _2_1 = .;
595:    _2_2 = .;
596:    _2_3 = .;
597:    substr(_warn_,2,1) = 'U';
598:    _DM_BAD = 1;
599:  end;
600: end;
601:
602: *** Generate dummy variables for GRP_LTV_time ;
603: drop _3_0 _3_1 _3_2 _3_3 ;
604: *** encoding is sparse, initialize to zero;
605: _3_0 = 0;
606: _3_1 = 0;
607: _3_2 = 0;
608: _3_3 = 0;
609: if missing( GRP_LTV_time ) then do;
610:   _3_0 = .;
611:   _3_1 = .;
612:   _3_2 = .;
613:   _3_3 = .;

```

```
614: substr(_warn_,1,1) = 'M';
615: _DM_BAD = 1;
616: end;
617: else do;
618: length _dm12 $ 12; drop _dm12 ;
619: _dm12 = put( GRP_LTV_time , BEST12. );
620: %DMNORMIP( _dm12 )
621: _dm_find = 0; drop _dm_find;
622: if _dm12 <= '3' then do;
623: if _dm12 <= '2' then do;
624: if _dm12 = '1' then do;
625: _3_0 = 1;
626: _dm_find = 1;
627: end;
628: else do;
629: if _dm12 = '2' then do;
630: _3_1 = 1;
631: _dm_find = 1;
632: end;
633: end;
634: end;
635: else do;
636: if _dm12 = '3' then do;
637: _3_2 = 1;
638: _dm_find = 1;
639: end;
640: end;
641: end;
642: else do;
643: if _dm12 = '4' then do;
644: _3_3 = 1;
645: _dm_find = 1;
646: end;
647: else do;
648: if _dm12 = '5' then do;
649: _3_0 = -1;
650: _3_1 = -1;
651: _3_2 = -1;
652: _3_3 = -1;
653: _dm_find = 1;
654: end;
655: end;
```



```
656:   end;
657:   if not _dm_find then do;
658:     _3_0 = .;
659:     _3_1 = .;
660:     _3_2 = .;
661:     _3_3 = .;
662:     substr(_warn_,2,1) = 'U';
663:     _DM_BAD = 1;
664:   end;
665: end;
666:
667: *** Generate dummy variables for GRP_mat_time ;
668: drop _4_0 _4_1 _4_2 _4_3 ;
669: *** encoding is sparse, initialize to zero;
670: _4_0 = 0;
671: _4_1 = 0;
672: _4_2 = 0;
673: _4_3 = 0;
674: if missing( GRP_mat_time ) then do;
675:   _4_0 = .;
676:   _4_1 = .;
677:   _4_2 = .;
678:   _4_3 = .;
679:   substr(_warn_,1,1) = 'M';
680:   _DM_BAD = 1;
681: end;
682: else do;
683:   length _dm12 $ 12; drop _dm12 ;
684:   _dm12 = put( GRP_mat_time , BEST12. );
685:   %DMNORMIP( _dm12 )
686:   _dm_find = 0; drop _dm_find;
687:   if _dm12 <= '3' then do;
688:     if _dm12 <= '2' then do;
689:       if _dm12 = '1' then do;
690:         _4_0 = 1;
691:         _dm_find = 1;
692:       end;
693:     else do;
694:       if _dm12 = '2' then do;
695:         _4_1 = 1;
696:         _dm_find = 1;
697:       end;
```

```
698:   end;
699:   end;
700:   else do;
701:     if _dm12 = '3' then do;
702:       _4_2 = 1;
703:       _dm_find = 1;
704:     end;
705:   end;
706:   end;
707:   else do;
708:     if _dm12 = '4' then do;
709:       _4_3 = 1;
710:       _dm_find = 1;
711:     end;
712:   end;
713:   if _dm12 = '5' then do;
714:     _4_0 = -1;
715:     _4_1 = -1;
716:     _4_2 = -1;
717:     _4_3 = -1;
718:     _dm_find = 1;
719:   end;
720:   end;
721:   end;
722:   if not _dm_find then do;
723:     _4_0 = .;
724:     _4_1 = .;
725:     _4_2 = .;
726:     _4_3 = .;
727:     substr(_warn_,2,1) = 'U';
728:     _DM_BAD = 1;
729:   end;
730:   end;
731:
732:   *** If missing inputs, use averages;
733:   if _DM_BAD > 0 then do;
734:     _P0 = 0.2900241741;
735:     _P1 = 0.7099758259;
736:     goto REGDR1;
737:   end;
738:
739:   *** Compute Linear Predictor;
```

```

740: drop _TEMP;
741: drop _LP0;
742: _LP0 = 0;
743:
744: *** Effect: GRP_FICO_orig_time ;
745: _TEMP = 1;
746: _LP0 = _LP0 + ( 0.51296160733899) * _TEMP * _1_0;
747: _LP0 = _LP0 + ( 0.27991536330134) * _TEMP * _1_1;
748: _LP0 = _LP0 + ( 0.06105378259046) * _TEMP * _1_2;
749: _LP0 = _LP0 + ( -0.13316716428952) * _TEMP * _1_3;
750:
751: *** Effect: GRP_LTV_orig_time ;
752: _TEMP = 1;
753: _LP0 = _LP0 + ( -0.3319324909476) * _TEMP * _2_0;
754: _LP0 = _LP0 + ( -0.02296870395815) * _TEMP * _2_1;
755: _LP0 = _LP0 + ( 0.05991431686603) * _TEMP * _2_2;
756: _LP0 = _LP0 + ( 0.30026168661542) * _TEMP * _2_3;
757:
758: *** Effect: GRP_LTV_time ;
759: _TEMP = 1;
760: _LP0 = _LP0 + ( -0.37158516692134) * _TEMP * _3_0;
761: _LP0 = _LP0 + ( -0.18560068889752) * _TEMP * _3_1;
762: _LP0 = _LP0 + ( -0.00295697959773) * _TEMP * _3_2;
763: _LP0 = _LP0 + ( 0.22449583021074) * _TEMP * _3_3;
764:
765: *** Effect: GRP_mat_time ;
766: _TEMP = 1;
767: _LP0 = _LP0 + ( -0.9877917346315) * _TEMP * _4_0;
768: _LP0 = _LP0 + ( -0.44439272084549) * _TEMP * _4_1;
769: _LP0 = _LP0 + ( 0.16634253811835) * _TEMP * _4_2;
770: _LP0 = _LP0 + ( 0.52925345649279) * _TEMP * _4_3;
771:
772: *** Naive Posterior Probabilities;
773: drop _MAXP _IY _P0 _P1;
774: _TEMP = -1.12831909867298 + _LP0;
775: if (_TEMP < 0) then do;
776: _TEMP = exp(_TEMP);
777: _P0 = _TEMP / (1 + _TEMP);
778: end;
779: else _P0 = 1 / (1 + exp(-_TEMP));
780: _P1 = 1.0 - _P0;
781:

```

```

782:  REGDR1:
783:
784:
785:  *** Posterior Probabilities and Predicted Level;
786:  label P_TARGET1 = 'Predicted: TARGET=1' ;
787:  label P_TARGET0 = 'Predicted: TARGET=0' ;
788:  P_TARGET1 = _P0;
789:  _MAXP = _P0;
790:  _IY = 1;
791:  P_TARGET0 = _P1;
792:  if (_P1 > _MAXP + 1E-8) then do;
793:  _MAXP = _P1;
794:  _IY = 2;
795:  end;
796:  I_TARGET = REGDRF[_IY];
797:  U_TARGET = REGDRU[_IY];
798:
799:  *****;
800:  ***** end scoring code for regression;
801:  *****;
802:  *-----
      *;
803:  * TOOL: Score Node;
804:  * TYPE: ASSESS;
805:  * NODE: Score;
806:  *-----
      *;
807:  *-----
      *;
808:  * Score: Creating Fixed Names;
809:  *-----
      *;
810:  LABEL EM_EVENTPROBABILITY = 'Probability for level 1 of
      TARGET';
811:  EM_EVENTPROBABILITY = P_TARGET1;
812:  LABEL EM_PROBABILITY = 'Probability of Classification';
813:  EM_PROBABILITY =
814:  max(
815:  P_TARGET1
816:  ,
817:  P_TARGET0
818:  );

```

```
819:  LENGTH EM_CLASSIFICATION $%dmnorlen;  
820:  LABEL EM_CLASSIFICATION = "Prediction for TARGET";  
821:  EM_CLASSIFICATION = I_TARGET;  
822:  
823:  run;
```