

UNIVERSIDADE FEDERAL DO PARANÁ

CARLOS HENRIQUE WROBEL

**MODELO LOGISTICO DE *CREDIT SCORE* APLICADO AO MERCADO DE
CESSÃO DE DIREITOS CREDITÓRIOS EMPRESARIAIS**

CURITIBA

2017

CARLOS HENRIQUE WROBEL

**MODELO LOGÍSTICO DE *CREDIT SCORE* APLICADO AO MERCADO DE
CESSÃO DE DIREITOS CREDITÓRIOS EMPRESARIAIS**

Artigo apresentado à disciplina Metodologia Científica como requisito à conclusão do curso MBA em Finanças, Centro de Pesquisa e Pós-Graduação em Administração da Universidade Federal do Paraná.

Orientadora: Prof^a Dr^a: Queila Regina Souza Matiz

CURITIBA

2017

MODELO LOGÍSTICO DE *CREDIT SCORE* APLICADO AO MERCADO DE CESSÃO DE DIREITOS CREDITÓRIOS EMPRESARIAIS

Carlos Henrique Wrobel

RESUMO

O mercado de compra de títulos de direitos creditórios (*Factoring*), bem como outros mercados que demandam mensuração de risco de crédito, também se beneficiam largamente de processos de avaliação automatizada de crédito. Neste contexto que se assenta a relevância do desenvolvimento e validação de técnicas de *credit scoring*. Este trabalho concentrou esforços na geração e validação de modelos *risk scoring* através de técnicas de regressão logística, tendo como alvo o mercado de compra e venda de direitos creditórios. Este trabalho não logrou êxito nesta tarefa, levantando hipóteses para tal.

Palavras-chave: Risco de crédito. Regressão logística. Modelos *Credit Score*. *Factoring*.

CREDIT SCORE LOGISTIC MODEL APPLIED TO THE FACTORING MARKET

ABSTRACT

The factoring market, as well as other markets that require credit risk measurement, largely benefits from automated credit assessment analysis. In this context, the development and validation of risk scoring techniques takes place. This article focused on development and validation of risk scoring models based on logistic regression techniques, targeting the factoring market . This work did not achieved success in this task, raising the hypotheses for such.

Keywords: Credit Risk. Logistic regression. Credit Scoring. Factoring.

1 INTRODUÇÃO

A concessão de crédito através da compra de direitos creditórios, também conhecido como *factoring* ou fomento comercial, é caracterizado pela compra de direitos creditórios originados em vendas de produtos ou de prestação serviços a prazo. Estes direitos creditórios são usualmente representados por títulos de crédito como duplicatas ou cheques (ANFAC).

Através de operações de compra de títulos de crédito, uma empresa de *factoring* provê liquidez à empresa até então detentora dos títulos, adiantando os recebimentos futuros. O tomador do crédito é representado pelo comprador dos produtos ou serviços prestados, que tem seu crédito transferido pela empresa fornecedora à empresa de *factoring* através do endosso destes títulos. Desta forma, a empresa que originalmente havia concedido o crédito para seu cliente, os transfere para a empresa terceira, a *factoring*, que se torna a nova detentora dos créditos.

Assim como em outras atividades de concessão de crédito, a compra de direitos creditórios expõe a detentora dos títulos a diversos riscos, entre eles o risco de crédito - a probabilidade do tomador do crédito não honrar as obrigações assumidas (CAPELLETTO e CORRAR, 2008, p. 8). A administração dos riscos é uma atividade constante nas empresas de *factoring* e a avaliação do risco de crédito dos sacados de títulos se torna importante, e neste contexto, também a técnica do *credit scoring*.

A avaliação de risco de crédito por meio de *credit score* se popularizou nos países desenvolvidos a partir da década de 1960, como forma de possibilitar uma avaliação de crédito rápida e consistente para o novo mercado de cartões de crédito para pessoas físicas. Segundo Thomas (2000), o volume elevado de novos pedidos de cartões de crédito tornou impossível a tarefa de avaliar o crédito sem uma automação da decisão de crédito.

Já no Brasil, a tentativa de mensuração do risco através de técnicas estatísticas com base em informações cadastrais ou comportamentais passou a atrair mais atenção principalmente a partir de 1994, com a estabilidade monetária conseguida com o Plano Real (VASCONCELLOS, 2002).

Em operações de compra de títulos de crédito empresariais, o agente que adquire ativos empresariais, os adquire de uma empresa que está na condição de

fornecedora ou prestadora de serviço em uma relação comercial com seu cliente, e será liquidado pela empresa cliente nesta situação de cliente-fornecedor, enquanto a empresa de *factoring* é uma terceira empresa nesta relação.

Esta situação de distanciamento entre a empresa de *factoring* e a empresa-cliente (sacada do título), tem, como efeito prático, uma redução da capacidade de obtenção de informações por parte da empresa de *factoring* sobre o sacado - alvo da avaliação de risco. Muitas vezes, as informações obtidas acerca dos sacados dos títulos se restringem àquelas obtidas através de serviços de *Credit bureau* como o Serasa Experian®, BoaVista SCPC® e outras empresas prestadoras destes serviços.

Este artigo objetivou avaliar a performance e sugerir um modelo de *credit scoring* baseado em análise de regressão logística para o mercado de compra de direitos creditórios empresariais.

Foram aplicadas técnicas de regressão logística e inferência de rejeitados para elaboração de modelos *credit scoring* a partir de informações de um banco de dados de liquidações de títulos, de informações cadastrais e comportamentais providas por relatórios de *credit bureaus* sobre os sacados dos créditos, verificando a capacidade de predição do modelo sugerido.

As informações foram obtidas de um banco de dados cedido por uma securitizadora de créditos empresariais, contendo também informações cadastrais e comportamentais providas pelo Serasa Experian®. Foi realizada inferência de rejeitados pela reclassificação de títulos rejeitados como inadimplentes. Não obteve-se sucesso nas tentativas de estimação de um modelo eficiente para a inadimplência a partir dos dados levantados, com ou sem a inferência de rejeitados.

2 O USO DE MODELOS CREDIT SCORE PARA CONCESSÃO DE CRÉDITO

O tema modelagem para avaliação de risco de crédito para concessão de crédito é recorrentemente tratado na literatura, normalmente envolvendo a aplicação de técnicas estatísticas para geração de modelos com base em dados contábeis – a exemplo de Luis Damasceno et al (2008), Assaf Neto e Silva Brito (2008), Xavier Gonçalves e Soto Urbina (2009) e outros.

Para Vicente(2001):

“*Credit score*’ pode ser definido como o processo de atribuição de pontos às variáveis de decisão de crédito mediante aplicação de técnicas estatísticas. Trata-se de processo que define a probabilidade de que um cliente com certas características, pertença ou não a um grupo possuidor de outras determinadas características, consideradas desejáveis (hipótese quem que se aprova um limite de crédito) ficando a critério da instituição dar alçada operacional ou não para o gestor atribuir o crédito. Esta técnica estabelece uma regra de discriminação de um determinado cliente solicitante de crédito.” (VICENTE, 2001, p.49).

Um sistema de *credit score* é, desta forma, um sistema que pode simplificar e padronizar o evento de concessão de crédito, com uma variável de corte previamente definida.

2.1 VANTAGENS DE UTILIZAÇÃO DE MODELOS *CREDIT SCORE*

Silva Brito e Assaf Neto (2008, p. 28), ressaltam o expressivo nível de acerto e apelo prático obtidos por modelos de *credit score*, apesar de críticas com relação à sua fundamentação teórica.

Conforme destacaram Parkinson e Ochs¹ (1998 citados por NETO e BRITO, 2008, p 6.), são vantagens do uso de sistemas de *risk scoring*: a possibilidade de revisões de crédito consistentes, a organização de informações, a diminuição da subjetividade na avaliação, uma maior compreensão do processo de concessão do crédito, uma maior eficiência do processo, além de possibilidade de agregar informações fornecidas por terceiros.

Ou seja, a possibilidade de verificar empiricamente a relevância de cada informação disponível e construir modelos adequados ao fim que se almeja, além de possibilitar uma maior eficiência na utilização da informação, abre caminho para uma padronização dos níveis de risco aceitos e uma maior velocidade na avaliação.

Caouette, Altman e Narayanan (1999), listam vantagens sobre a utilização deste método de avaliação de risco:

“Os modelos de *credit scoring* oferecem muitas vantagens. São objetivos e consistentes, características desejáveis para qualquer instituição,

¹ PARKINSON, Kenneth L. & OCHS, Joyce R. **Using credit screening to manage credit risk.** Business Credit, 1998.

especialmente para aquelas que não possuem uma forte cultura de crédito. Se desenvolvidos apropriadamente, eles podem eliminar práticas discriminatórias nos empréstimos. (...) As metodologias usadas para construir esses modelos são comuns e bem entendidas, assim como as abordagens usadas para avaliá-los. (...) Uma instituição é capaz de proporcionar melhor serviço ao consumidor pela sua habilidade de aprovar ou negar um pedido de empréstimo rapidamente.” (CAOQUETTE, ALTMAN e NARAYANAN, 1999, p.188).

Uma avaliação intuitiva e subjetiva, além de mais lenta, dificulta a manutenção de estritos padrões de aceitação de risco e dificulta sua avaliação de eficiência a posteriori. Silva (2000), reforça a agilidade que uma avaliação por *credit score* pode proporcionar a uma instituição que concede crédito:

“A agilidade que o banco ou a empresa que concede crédito ganha é altamente valiosa, pois ao invés do analista ficar examinando e concluindo sobre cada um dos índices, poderão dedicar seus tempos a outros assuntos relevantes e que não possam ser sistematizados”. (SILVA, 2000, p. 306)

2.2 DESVANTAGENS E LIMITAÇÕES DE UTILIZAÇÃO DE MODELOS *CREDIT SCORE*

Caouette, Altman e Narayanan (1999) destacam a degradação temporal que um modelo *credit score* pode sofrer, por divergências entre a população base para a construção do modelo e a população em que o modelo terá aplicação.

Silva (2000), cita ainda aspectos geográficos como limitação, afirmando que aspectos únicos de uma região demandaria também um modelo exclusivo, o que exigiria amostras de dados muito grandes e específicas de cada região geográfica.

Alves (2008), ressalta a uma limitação quanto ao viés usualmente presente nas amostras utilizadas como base para a elaboração dos modelos, já que esta normalmente é constituída apenas por proponentes aptos ao crédito. Características sistematicamente presentes em proponentes rejeitados não seriam avaliadas corretamente pela análise, gerando um modelo incapaz de prever o comportamento destes proponentes. Nestes casos é exigida a aplicação de técnicas de inferência de rejeitados de forma a tentar mitigar o viés.

3 REGRESSÃO LOGÍSTICA

Na regressão logística, os coeficientes da regressão são estimados a partir do conjunto de dados pelo método da máxima verossimilhança, método que encontra valores para os coeficientes que maximizam a probabilidade de se obter resultados como os observados na amostra de dados utilizada para a estimação (HOSMER e LEMESHOW, 1989).

Modelos baseados em regressões logísticas são amplamente difundidos pelo mercado no desenvolvimento de modelos para *credit score* (ALVES, 2008). Majer (sem data), também cita como amplamente difundido para modelagem de variáveis dependentes binárias.

Por produzir resultados variando entre 0 e 1, podem ser facilmente interpretados como a probabilidade da ocorrência de um evento de inadimplência (ALVES, 2008).

Os modelos de regressão logística assumem a forma da expressão que se segue:

$$\ln\left(\frac{p}{1-p}\right) = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \dots + b_nX_n$$

Também citado por Alves (2008), como vantagens da regressão logística, está a não exigência de suposições fortes, como normalidade dos erros, além da já colocada interpretação direta da probabilidade da ocorrência do evento.

Foi realizada a estimação dos parâmetros e testes de ajuste e performance dos modelos com o apoio do software R, também com a inclusão de inferência de rejeitados, verificando a possibilidade de melhoria na estimação dos parâmetros e ajustes do modelo.

3.1 INFERÊNCIA DE REJEITADOS

A inferência de rejeitados é uma forma de incluir informações sobre o risco de inadimplência de ocorrências que foram previamente rejeitados ainda na análise de crédito (LOUZADA, 2013). Tipicamente, os dados disponíveis para a elaboração de

modelos de *credit scoring* são apenas aqueles de propostas de crédito aceitas, já que as informações referentes à inadimplência ou adimplência das propostas rejeitadas são desconhecidas.

A inclusão de informações através de técnicas de inferência de rejeitados visa a redução do viés existente na amostra com observações quanto a adimplência, já que esta inclui apenas a informação de proponentes aceitos, e a população a ser analisada pelo modelo será a população total dos proponentes.

As técnicas de inferência de rejeitados mais comumente utilizadas com objetivo de reduzir o vício amostral são as técnicas de reclassificação, ponderação e parcelamento (LOUZADA, 2013).

A inferência dos rejeitados neste exercício foi realizada pela reclassificação de rejeitados na análise de crédito como inadimplentes. Esta técnica de inferência de rejeitados, ao assumir a certeza de inadimplência dos proponentes rejeitados, pode causar um viés no modelo desenvolvido (ALVES, 2008).

3.2 DESENVOLVIMENTO DOS MODELOS

3.2.1 CONJUNTOS DE DADOS

Para a construção dos modelos foi levantada uma base de dados de liquidação de títulos adquiridos entre os dias dois de maio de 2016 e oito de maio de 2017, além de, para fins de inferência de rejeitados, selecionados os títulos rejeitados neste mesmo período, contendo 2.272 ocorrências.

Foram formadas duas bases de dados, uma contendo apenas os proponentes aceitos, e outra, contendo adicionalmente os proponentes rejeitados – classificados como inadimplentes. Estes conjuntos de dados contêm informação sobre o prazo de pagamento, além de informações fornecidas pelo Serasa Experian® no momento da decisão de aceitação ou não do título de crédito.

Estas bases de dados então foram separadas em bases de treinamento e de validação, com 60% e 40% dos total de ocorrências, respectivamente, para avaliação dos modelos gerados.

3.2.2 VARIÁVEL DEPENDENTE

Para a variável dependente, um título foi classificado como inadimplente quando apresentava um atraso superior a 5 dias corridos. Este prazo de definição de inadimplência decorre da obrigação de recompra dos títulos imposta ao cedente dos títulos pela securitizadora, além de se tratar de medida conveniente para aferição do risco assumido - o risco de não conseguir impor ao seu cliente a recompra de títulos em atraso. O número de classificados como inadimplentes somou 36 ocorrências.

Foi atribuído o valor 1 para títulos declarados como inadimplentes e 0 para os adimplentes, ou seja, liquidados com um prazo de até 5 dias de atraso.

Títulos recomprados antes do prazo de vencimento foram excluídos do conjunto de dados, por se tratarem primordialmente de títulos com algum vício de origem, seja no lastro dos produtos e/ou serviços prestados/vendidos, falhas na emissão dos títulos como data de vencimento ou número de identificação do título incorretos, que não terão pagamento pontual por motivos outros que não condição de crédito ou liquidez da empresa sacada.

3.2.3 VARIÁVEIS EXPLICATIVAS

As variáveis explicativas para a elaboração do modelo são informações cadastrais e de comportamento de pagamento informados pelo Serasa Experian®, apresentadas a seguir, representando sempre valores referenciados à condição da empresa sacada dos títulos de crédito avaliados:

Número de ocorrências – Pendências financeiras: representa o número de pendências financeiras indicadas, usualmente por fornecedores da empresa indicada, para registro no banco de dados de *bureaus de crédito*. Espera-se uma resposta positiva da inadimplência a esta variável.

Valor acumulado – Pendências financeiras: valor monetário acumulado das pendências financeiras registradas. Espera-se uma resposta positiva da inadimplência a esta variável.

Número de ocorrências – Refins: refere-se ao número de pendências financeiras de origem bancária registradas nos *bureaus de crédito*. Espera-se uma resposta positiva da inadimplência a esta variável.

Valor acumulado – Refins: valor acumulado de pendências financeiras bancárias. Espera-se uma resposta positiva da inadimplência a esta variável.

Número de ocorrências – Protestos: refere-se ao número de títulos protestados em cartório informados por cartórios aos *bureaus de crédito*. Espera-se uma resposta positiva da variável dependente.

Valor acumulado – Protestos: valor monetário acumulado de títulos protestados em cartório. Também se espera uma resposta positiva da variável dependente a este valor.

Número de ocorrências – Cheques devolvidos não reapresentáveis: número de registros de cheques devolvidos não reapresentáveis nos bancos de dados de *bureaus de crédito*. Espera-se uma resposta positiva da variável dependente.

Valor acumulado – Cheques devolvidos não reapresentáveis: valor acumulado de cheques devolvidos não reapresentáveis registrados nos bancos de dados de *bureaus de crédito*. Espera-se uma resposta positiva da variável dependente.

Número de ocorrências – Dívidas: número de registros de dívidas vencidas nos bancos de dados de *bureaus de crédito*. Espera-se uma resposta positiva da variável dependente.

Valor acumulado – Dívidas: valor acumulado de dívidas vencidas nos bancos de dados de *bureaus de crédito*. Espera-se uma resposta positiva da variável dependente.

Número de relacionamentos com fornecedores e clientes: número de empresas relacionadas com a empresa avaliada. Espera-se uma resposta negativa da inadimplência a esta variável, avaliando que quão mais pulverizada é a rede de clientes e fornecedores de uma empresa, menor a possibilidade de problemas em seu fluxo de pagamentos decorrentes de um evento de uma inadimplência de um de seus clientes.

Tempo médio de relacionamento com fornecedores e clientes: Tempo médio de relacionamento da empresa avaliada com seus clientes e fornecedores. Espera-se também uma resposta negativa da variável dependente, ao avaliar-se que longos relacionamentos com clientes e fornecedores derivam também de boas relações financeiras com clientes e fornecedores.

Percentual de pagamentos pontuais: Percentual da quantidade de pagamentos registrados como pontuais. Espera-se um coeficiente negativo para esta variável.

Percentual de pagamentos pontuais no mercado - últimos 12 meses: Percentual de pagamentos pontuais no mercado nos últimos 12 meses. Espera-se um coeficiente negativo para esta variável.

Percentual de pagamentos pontuais no mercado - últimos 3 meses: Percentual de pagamentos pontuais no mercado nos últimos 3 meses. Espera-se um coeficiente negativo para esta variável.

Percentual de pagamentos pontuais para factorings - últimos 12 meses: Percentual de pagamentos pontuais no para *factorings* nos últimos 12 meses. Espera-se um coeficiente negativo para esta variável.

Percentual de pagamentos pontuais para factorings - últimos 3 meses: Percentual de pagamentos pontuais no para *factorings* nos últimos 3 meses. Espera-se um coeficiente negativo para esta variável.

Idade da empresa: Idade, em anos, da empresa avaliada.

O quadro 1 apresenta a notação utilizada para cada uma das variáveis elencadas.

QUADRO 1 – VARIÁVEIS EXPLICATIVAS

Notação	Variável
X ₁	Número de ocorrências – Pendências financeiras
X ₂	Valor acumulado – Pendências financeiras
X ₃	Número de ocorrências – Refins
X ₄	Valor acumulado – Refins
X ₅	Número de ocorrências – Protestos
X ₆	Valor acumulado – Protestos
X ₇	Número de ocorrências – Cheques devolvidos não rerepresentáveis
X ₈	Valor acumulado – Cheques devolvidos não rerepresentáveis
X ₉	Número de ocorrências – Dívidas
X ₁₀	Valor acumulado – Dívidas
X ₁₁	Número de relacionamentos com fornecedores e clientes
X ₁₂	Tempo médio de relacionamento com fornecedores e clientes

X ₁₃	Percentual de pagamentos pontuais
X ₁₄	Percentual de pagamentos pontuais no mercado - últimos 12 meses
X ₁₅	Percentual de pagamentos pontuais no mercado - últimos 3 meses
X ₁₆	Percentual de pagamentos pontuais para factorings - últimos 12 meses
X ₁₇	Percentual de pagamentos pontuais para factorings - últimos 3 meses
X ₁₈	Idade da empresa

FONTE: O autor (2017)

3.2.4 SELEÇÃO DAS VARIÁVEIS

O método de seleção de variáveis principal adotado foi o *forward stepwise*, procedimento que começa ao acrescentar a um modelo inicial a variável independente que melhor explica a variável dependente, e então acrescenta novas variáveis independentes ao modelo progressivamente, por ordem de contribuição de razão de verossimilhança ao modelo. Este procedimento de seleção não acrescenta mais variáveis independentes quando novas variáveis independentes não produzem acréscimo significativo à razão de verossimilhança (SILVA, 2006).

Os modelos finais propostos, para o conjunto de dados sem a inferência de rejeitados e com a inferência de rejeitados são descritos pelas seguintes funções:

Sem inferência de rejeitados:

$$\ln\left(\frac{p}{1-p}\right) = 0,362 - 0,076 \cdot X_{14} - 0,426 \cdot X_{17}$$

E com inferência de rejeitados:

$$\ln\left(\frac{p}{1-p}\right) = 1,67 - 3,25 \cdot 10^{-4} \cdot X_1 - 2,68 \cdot 10^{-4} \cdot X_{11} - 8,1 \cdot 10^{-1} \cdot X_{13} + 1,74 \cdot 10^{-1} X_{14} - 6,87 \cdot 10^{-1} X_{16} - 2,84 \cdot 10^{-1} X_{17}$$

4 AVALIAÇÃO DOS MODELOS

Para avaliar os modelos gerados, usamos como indicador de avaliação a Curva ROC (*Receiver Operating Characteristic Curve*) e a estatística Kolmogorov-Smirnov (KS).

4.1 Curva ROC

A curva ROC (*Receiver Operating Characteristic Curve*) é uma ferramenta utilizada para avaliação de desempenho de modelos baseados em regressão logística.

Esta técnica parte do princípio de construção de várias matrizes de confusão. As matrizes de confusão são matrizes construídas a partir da definição de um ponto de corte (0,5 para os exemplos dos quadros 2 e 3 a seguir), em que resultados do modelo superiores à linha de corte são classificados como inadimplentes e abaixo deste valor, classificadas como adimplentes.

Estes resultados são então contrastados com a amostra de validação, e então revela-se a quantidade de classificações corretas que o modelo gerou.

Quando as ocorrências observadas e previstas são coincidentes, estas ocorrências são contabilizadas.

A seguir os resultados de matriz de confusão para as bases de dados com inferência de rejeitados:

QUADRO 2 – MATRIZ DE CONFUSÃO - COM INFERÊNCIA DE REJEITADOS

Matriz	Observado		
	Adimplente	Inadimplente	Total
Previsão			
Adimplente	882	70	952
Inadimplente	10	13	23
Total	892	83	975

FONTE: O autor (2017).

A seguir os resultados de matriz de confusão para as bases de dados sem a inferência de rejeitados:

QUADRO 3 – MATRIZ DE CONFUSÃO - SEM INFERÊNCIA DE REJEITADOS

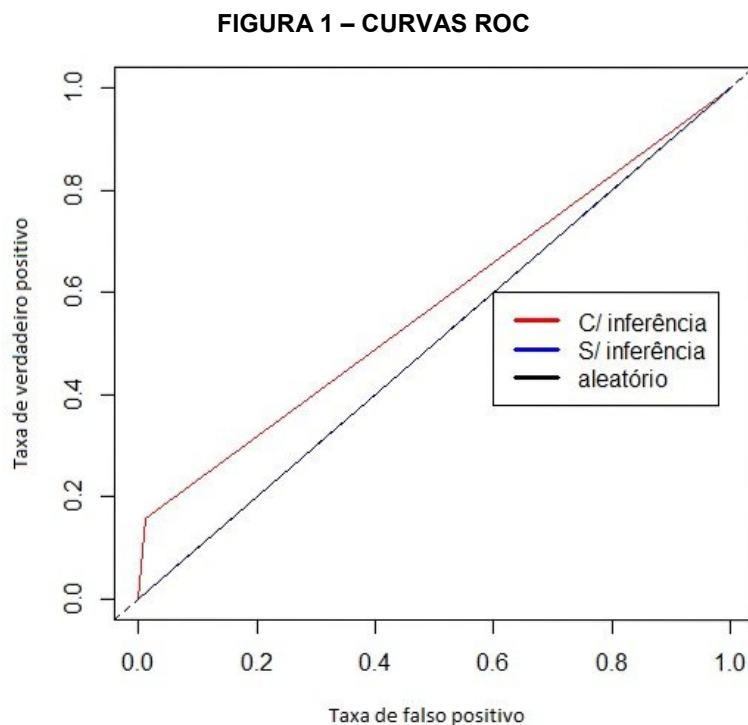
Matriz	Observado		
	Adimplente	Inadimplente	Total
Previsão			
Adimplente	897	12	909
Inadimplente	0	0	0
Total	897	12	909

FONTE: O autor (2017).

As curvas ROC são construídas a partir da variação da linha de corte, obtendo as diferentes classificações para cada ponto da linha de corte, que são utilizados então para a construção da curva. Os valores do eixo horizontal são compostos pela proporção de inadimplentes classificados incorretamente como inadimplentes pelo modelo, enquanto os valores no eixo vertical são compostos pela proporção de inadimplentes classificados corretamente pelo modelo (ZWEIG e CAMPBELL, 1993).

Desta forma, um modelo que discrimine perfeitamente inadimplentes e adimplentes, teria uma reta partindo da origem para o ponto extremo superior esquerdo, enquanto um modelo incapaz de distinguir corretamente as ocorrências tende a descrever uma reta partindo da origem ao ponto extremo superior direito da área gráfica.

A seguir o desenho da curva ROC obtida pelos resultados do modelo originado a partir da base de dados com inferência de rejeitados:



FONTE: O autor (2017).

4.2 Estatística Kolmogorov-Smirnov

A estatística de Kolmogorov-Smirnov (KS) é uma medida de performance de classificação dos modelos, ao tentar determinar se duas distribuições se diferenciam significativamente.

A frequência relativa acumulada da pontuação atribuída aos inadimplentes deverá se distanciar ao máximo em relação a frequência relativa acumulada da pontuação obtida aos proponentes adimplentes em um modelo eficiente na classificação (PEREIRA, 2004).

A estatística KS tem resultados variando entre 0 e 1, em que modelos capazes de diferenciar melhor proponentes adimplentes de inadimplentes, teriam resultados da estatística se aproximando de 1, enquanto modelos incapazes de diferenciá-los tem resultado para este teste com valores se aproximando de 0.

Para os modelos gerados a partir da base de dados com inferência de rejeitados, o resultado obtido é de 0,1454157, enquanto para o modelo sem a inferência de rejeitados, este resultado é de 0 – o modelo não conseguiu classificar corretamente nenhuma ocorrência. Estes são valores que, de acordo com Alves (2008), denotam baixa capacidade de diferenciar proponentes adimplentes e inadimplentes por parte dos modelos gerados.

5 CONCLUSÕES

O objetivo deste estudo foi desenvolver modelos de *risk scoring*, com ou sem a inferência de rejeitados proposta, que se adequem às necessidades de empresas que atuam no mercado de compra e venda de direitos creditórios empresariais. O trabalho foi realizado utilizando técnicas de regressão logística, com validação através da aplicação de técnica de construção de Curva ROC e verificação da estatística Kolmogorov-Smirnov para os modelos gerados.

Os modelos gerados não alcançaram níveis mínimos aceitáveis de capacidade de predição do parâmetro de inadimplência proposto, fato verificável através das matrizes de confusão, curvas ROC e estatísticas KS obtidas.

Esta incapacidade, pode, por hipótese, ter origem na reduzida amostra de inadimplência contida na base de dados cedida, haja visto o número de apenas 36

ocorrências para esta situação identificados. Esta limitação da amostra foi imposta por condição da empresa que se dispôs a ceder as informações, haja visto que opera a apenas dois anos no mercado, tendo um número reduzido de títulos liquidados na base de dados.

Outra possibilidade para a incapacidade de geração de um modelo com base nos dados cadastrais fornecidos seria a de que, dadas as restrições impostas para a condição da variável dependente – títulos classificados como inadimplentes com apenas 5 dias de atraso – seria esta uma condição muito danosa à capacidade de geração de modelos com base nas variáveis dependentes disponibilizadas. Existem ocorrências de empresas com condições de crédito identificadas positivamente pelos relatórios de *bureaus de crédito* que, por eventuais falhas operacionais – não vinculadas à condição de crédito – tiveram títulos identificados como “inadimplentes”. Estas ocorrências não são possíveis de se identificar no histórico de liquidações disponibilizado, mas são ocorrências declaradas como recorrentes pela empresa que nos cedeu a base de dados, fato este que, em conjunto com o número reduzido de ocorrências de inadimplentes pode ter colaborado significativamente para a incapacidade de geração de modelos eficientes na identificação de inadimplentes e adimplentes com base nas variáveis independentes selecionadas.

REFERÊNCIAS

- ALVES, MAURO CORREIA. Estratégias para o desenvolvimento de modelos de *credit score* com inferência de rejeitados. 2008. **Biblioteca digital USP**. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/45/45133/tde-14102008-204609/publico/DissertacaoMauro.pdf>> Acesso em: 03 de junho de 2017.
- ANFAC. Securitizadoras. Disponível em: <<http://www.anfac.com.br/v3/securitizadoras.jsp>>. Acesso em: 08 de outubro de 2016.
- CAOQUETTE, JOHN B.; ALTMAN, EDWARD I.; NARAYANAN, PAUL. **Gestão do risco de crédito**. 1ª ed. Rio de Janeiro: Qualitymark, 1999.
- CAPELLETTO, L. R.; CORRAR, L. J. Índices de Risco Sistêmico para o Setor Bancário. **Revista Contabilidade & Finanças - USP**, v. 19, n. 47, art. 2, p. 6-18, 2008. Disponível em: <<http://www.spell.org.br/documentos/download/6343>>.

Acesso em: 03 de junho de 2017.

DAMASCENO, DANILO LUIS; ARTES, RINALDO; ACCIOLY FONSECA M., ANDREA MARIA. Determinação de rating de crédito de empresas brasileiras com a utilização de índices contábeis. 2008. **Revista de Administração da Universidade de São Paulo**.

Disponível em: <http://200.232.30.99/busca/artigo.asp?num_artigo=1350>.

Acesso em: 08 de outubro de 2016.

GONÇALVES, FABRÍCIO X.; URBINA, LIGIA M. S.. Avaliação do risco de crédito empresarial. **Divisão de informação e Documentação – ITA**.

Disponível em: < <http://www.bibl.ita.br/xvencita/MEC05.pdf> >.

Acesso em: 09 de outubro de 2016.

HOSMER, D.; LEMESHOW, S. Applied logistic regression. **New York: John Wiley & Sons**. 1989.

Disponível em: <http://resource.heartonline.cn/20150528/1_3kOQSTg.pdf>

Acesso em: 19 de junho de 2017.

LOUZADA, CARLOS D. F.. Métodos Estatísticos para Análise de dados de crédito. 2013. **6th Brazilian Conference on Statistical Modelling in Insurance and Finance**.

Disponível em:

<http://www.mwstat.com/franciscoulouzada/Books_files/Livro_BICSMIF%202013-TamanhoReduzido.pdf>.

Acesso em: 17 de junho de 2017.

MAJER, IZABELA. *Working Paper No. 10-06: Application scoring: logit model approach and the divergence method compared*. **Department of applied Econometrics Working Papers. Warsaw School of Economics**.

Disponível em: <[https://ssl-](https://ssl-kolegia.sgh.waw.pl/pl/KAE/struktura/IE/struktura/ZES/Documents/Working_Papers/awp10-06.pdf)

[kolegia.sgh.waw.pl/pl/KAE/struktura/IE/struktura/ZES/Documents/Working_Papers/awp10-06.pdf](https://ssl-kolegia.sgh.waw.pl/pl/KAE/struktura/IE/struktura/ZES/Documents/Working_Papers/awp10-06.pdf)>.

Acesso em: 17 de junho de 2017.

NETO, A.; BRITO, G. A. S. Modelo de classificação de risco de crédito de empresas. 2008. **Sistema integrado de bibliotecas Universidade de São Paulo**.

Disponível em:

<http://www.producao.usp.br/bitstream/handle/BDPI/6154/art_BRITO_Modelo_de_classificacao_de_risco_de_credito_2008.pdf?sequence=1>.

Acesso em: 08 de outubro de 2016.

PEREIRA, GUSTAVO H. DE ARAUJO. Modelos de Risco de Crédito de clientes: Uma aplicação a Dados Reais. 2004. **Sistema integrado de bibliotecas Universidade de São Paulo**.

Disponível em: <<http://www.teses.usp.br/teses/disponiveis/45/45133/tde-28122004-224257/publico/diss1GP.pdf>>.

Acesso em: 17 de junho de 2017.

PEREIRA, G. A.; MILANI, LETÍCIA L.; CIRILLO, MARCELO Â.. Uso de alguns estimadores ridge na análise estatística de experimentos em entomologia. 2014.

Rev. Ceres, Viçosa , v. 61, n. 3, p. 338-342.

Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-737X2014000300006&lng=en&nrm=iso>.

Acesso em: 19 de junho de 2017.

SILVA, CLEIDE A. M. Exploração de métodos de seleção de variáveis pela técnica de regressão logística para análise de dados epidemiológicos. 2006. **Repositório da Produção Científica e Intelectual da Unicamp.**

Disponível em:

<<http://repositorio.unicamp.br/bitstream/REPOSIP/313543/1/Silva%2C%20Cleide%20Aparecida%20M.pdf>>.

Acesso em: 17 de junho de 2017.

THOMAS, LYN C. *A survey of credit and Behavioural Scoring: Forecasting financial risk of lending to consumers.* **International Journal of Forecasting 16 (2000) 149-172.**

Disponível em:

<<http://pages.ucsd.edu/~aronatas/project/academic/A%20survey%20of%20credit%20and%20behavioural%20scoring%20Forecasting%20fina.pdf>>.

Acesso em: 15 de abril de 2017.

VASCONCELLOS, MAURÍCIO S.. Proposta de método para análise de concessões de crédito a pessoas físicas. **Biblioteca digital USP – Teses e dissertações**, vol. 72, Nº 6, p. 584-590, dezembro de 1964.

Disponível em: <<http://www.teses.usp.br/teses/disponiveis/12/12138/tde-18112004-125442/publico/dissertacao.pdf>>.

Acesso em: 10 de setembro de 2016.

SILVA, JOSÉ P.. **Gestão e análise do risco de crédito.** 3ª ed. São Paulo: Atlas, 2000.

VICENTE, ERNESTO F. R.. A estimativa do risco na constituição da PDD. 2001.

Biblioteca digital USP.

Disponível em: <<http://www.teses.usp.br/teses/disponiveis/12/12136/tde-31012002-000545/publico/DissertacaoEFernando.pdf>>

Acesso em: 03 de junho de 2017.

ZWEIG, M. H.; CAMPBELL, G.. *Receiver-operating characteristic (ROC) plots.*

Clinical Chemistry. Vol. 39, Issue 4. p. 564. Abril de 1993.

Disponível em: <<http://clinchem.aaccjnls.org/content/39/4/561>>.

Acesso em: 22 de junho de 2017.