



Universidade de Brasília
Faculdade de Economia, Administração, Contabilidade e Ciência da
Ciência da Informação e Documentação - FACE.
Departamento de Economia

MODELOS DE PREVISÃO DE FALÊNCIA EMPRESARIAL: UMA APLICAÇÃO DA REGRESSÃO LOGÍSTICA E DO *BACKWARD STEPWISE* AO CASO BRASILEIRO

PAULO ROBERTO FONTELES GUIMARÃES

Monografia de conclusão de curso
de Bacharel em Ciências
Econômicas. Universidade de
Brasília – UnB

Orientador: Prof. Dr. Moisés de Andrade Resende Filho

Brasília - DF
2017

GUIMARÃES, Paulo Roberto Fonteles.

Modelos de Previsão de Falência Empresarial: uma Aplicação da Regressão Logística e do *Backward Stepwise* ao caso brasileiro/Paulo Roberto Fonteles Guimarães. Brasília – DF, 2017, 68 págs.

Monografia (bacharelado) – Universidade de Brasília, Departamento de Economia, 2017.

Orientador: Prof. Dr. Moisés de Andrade Resende Filho, Departamento de Economia.

1. Introdução 2. Panorama Histórico dos Modelos de Previsão de Falência: 1930 a 2016 3. Metodologia 4. Desenvolvimento do Modelo 5. Resultados 6. Conclusões 7. Referências Bibliográficas.

MODELOS DE PREVISÃO DE FALÊNCIA
EMPRESARIAL: UMA APLICAÇÃO DA
REGRESSÃO LOGÍSTICA E DO BACKWARD
STEPWISE AO CASO BRASILEIRO

PAULO ROBERTO FONTELES GUIMARÃES

Banca Examinadora

.....
Prof. Dr. Rafael Terra de Menezes
Membro

.....
Prof. Dr. Moisés de Andrade Resende Filho
Orientador

Brasília - DF
2017

Agradecimentos

Em primeiro lugar, preciso agradecer a meu alicerce fundamental nesta vida, que é Deus, minha família e o amor que me rege. Por minha família, quero dizer não apenas meus pais, Paulo e Regina, e meu irmão Luiz, mas também Ana Catarina, que é o amor da minha vida. Sem estas pessoas, eu não teria a menor chance de sobreviver ao mundo e de me encontrar como profissional e como ser humano.

Aos meus pais, agradeço pela ajuda e pelo apoio que me deram ao longo da minha vida, todos os conselhos, todos os puxões de orelha e todas as motivações.

Ao meu irmão, agradeço por todos os momentos de companheirismo e de alegria, toda a participação e compaixão fornecidas a mim.

À Ana, agradeço por ter tornado minha vida muito melhor do que jamais pensei que pudesse ser e agradeço por toda a felicidade que me faz sentir, por todo o apoio incondicional que me dá e por me motivar a ser um profissional e um homem muito melhor. Agradeço por me fazer saber o que é amor verdadeiro.

Preciso agradecer também aos meus amigos, que sei que estão comigo sempre que precisar, para conselhos, risadas, apoio e companheirismo. Sei que com eles posso contar a qualquer hora.

Agradeço também a todos os professores que me ensinaram a nobre profissão de Economista, responsável por otimizar a alocação de recursos da sociedade e, desta forma, melhorar a vida da população. Em especial, gostaria de agradecer ao professor Moisés, que me orientou enormemente ao longo deste trabalho e, sem o qual, esta monografia não poderia ter sido feita e ao professor Rafael, por seus conselhos e sugestões.

Resumo

Este trabalho tem como objetivo principal aplicar o método estatístico conhecido como regressão logística, especificando um modelo capaz de prever o estado de inadimplência de empresas não financeiras brasileiras com, pelo menos, defasagem de um ano. Como objetivo secundário, o trabalho buscou identificar as variáveis explicativas mais relevantes para prever o evento de inadimplência. Para isso, fez-se um panorama histórico dos modelos mais reconhecidos na literatura sobre o tema e, então, encontrar a melhor especificação do modelo com a ajuda de técnicas e testes econométricos. Concluiu-se que é possível construir um modelo com relativos bons ajuste aos dados e desempenho preditivo.

Palavras-Chave: Insolvência, Regressão Logística, Logit, Backward Stepwise, Falência, Empresas não Financeiras

Abstract

The main purpose of this work is to apply the statistical method of logistic regression to specify a model to predict Brazilian non financial companies' insolvency state with, at least, a year lag. As a secondary objective, this work tries to identify which explanatory variables are more relevant to predict the bankruptcy event. In so doing, an historical panorama of the most important models in the literature on this topic was drawn and, then, the model was specified with the help of econometric techniques and tests. I concluded that it is possible to build a model with relative good fit to the data and predictive performance.

Keywords: Insolvency, Logistic Regression, Logit, Backward Stepwise, Bankruptcy, non Financial Companies

Sumário

1. Introdução.....	9
2. Panorama Histórico dos Modelos de Previsão de Insolvência: 1930 a 2016	11
2.1 Estudos Iniciais: 1930 a 1960	11
2.2 Estudos que Empregam Análise Discriminante	13
2.3 Estudos que Empregam o Modelo <i>Logit</i>	18
2.4 Estudos que Empregam Redes Neurais e Análise por Envoltória de Dados	20
2.5 Estudos que Empregam Outros Modelos de Risco	22
2.6 Estudos Brasileiros Seminais	23
2.7 Estudos Brasileiros Recentes	25
3. Metodologia	29
3.1 <i>Proxy</i> de Insolvência	29
3.2 Definição da Amostra	29
3.3 Seleção das Variáveis Explicativas.....	35
3.4 Modelo Econométrico.....	39
3.4.1 Regressão Logística	39
4. Desenvolvimento do Modelo.....	41
4.1 Método Manual	44
4.1 Método Automático	55
5. Resultados.....	58
6. Conclusões	60
7. Referências Bibliográficas	62

Lista de Figuras

Figura 1 – Quantidade de Dados Faltantes por Variável	39
Figura 2 – Representação da Matriz de Correlação das Variáveis Explicativas	40
Figura 3 – Curva ROC para o modelo com as variáveis PE.AT, LACUM.AT e PC.AT	58
Figura 4 – Curva ROC para o modelo com as variáveis PE.AT, LACUM.AT , MedioGrandePorte e GrandePorte	58

Lista de Gráficos

Gráfico 1 – Estado de Adimplência das empresas da amostra no quarto trimestre de 2015.	30
Gráfico 2 – Estado de Adimplência das empresas da amostra no terceiro trimestre de 2016.	31
Gráfico 3 – <i>Boxplot</i> da variável mediana da Receita Operacional Anual entre o terceiro trimestre de 2014 e primeiro trimestre de 2015, considerando apenas empresas não financeiras com dados existentes no sistema Económica.	33
Gráfico 4 – <i>Boxplot</i> da variável mediana da Receita Operacional Anual entre o terceiro trimestre de 2014 e o primeiro trimestre de 2015, sem as dez empresas de maior porte.	34

Lista de Tabelas

Tabela 1 – Seleção preliminar de variáveis explicativas contáveis a serem testadas no modelo	35
Tabela 2 – Seleção preliminar de variáveis explicativas não contáveis a serem testadas no modelo.	36
Tabela 3 – Análise Descritiva das Variáveis Explicativas	38
Tabela 4 – Resultados do teste de Mann-Whitney-Wilcoxon	43
Tabela 5 – Resultados do Modelo Nulo.	44
Tabela 6 - Resultados da Segunda Etapa	45
Tabela 7 - Resultados do Modelo com PE.AT	46
Tabela 8 - Resultados da Terceira Etapa	46
Tabela 9 - Resultados do Modelo com PE.AT e LACUM.AT	48
Tabela 10 - Resultados do Modelo com PE.AT e EBIT.AT	49
Tabela 11 - Resultados da Quarta Etapa	52
Tabela 12 - Resultados do Modelo com PE.AT, LACUM.AT e PC.AT	53
Tabela 13 - Resultados do Modelo com PE.AT, LACUM.AT, MedioGrandePorte, MedioGrandePorte*PE.AT e MedioGrandePorte*LACUM.AT	53
Tabela 14 – Resultados do Modelo Completo	55

Tabela 15 - Resultados do Modelo <i>Backward Stepwise</i>	56
Tabela 16 - Resultados do Modelo <i>Backward Stepwise</i> sem considerar as variáveis de porte	56

Lista de Quadros

Quadro 1 – Classificação do SEBRAE para porte empresarial, segundo faturamento anual, Brasil, 2016.	31
Quadro 2 – Classificação da Agência Nacional de Vigilância Sanitária para porte empresarial, segundo faturamento anual, Brasil, 2016.	32
Quadro 3 – Classificação do IBGE para porte empresarial, segundo número de funcionários, Brasil, 2016.	32
Quadro 4 – Classificação do BNDES para porte empresarial, segundo Receita Operacional Anual, Brasil, 2016.	32
Quadro 5 - <i>Confusion Matrix</i> para o Modelo com a Variável EBIT.AT Aplicado à Subamostra de Teste	50
Quadro 6 - <i>Confusion Matrix</i> para o Modelo com a Variável EBIT.AT Aplicado a Base de Dados	50
Quadro 7 - <i>Confusion Matrix</i> para o Modelo com a Variável EBIT.AT Aplicado ao Estado de Solvência no Terceiro Trimestre de 2016	50
Quadro 8 - <i>Confusion Matrix</i> para o Modelo com a Variável LACUM.AT Aplicado à Subamostra de Teste	50
Quadro 9 - <i>Confusion Matrix</i> para o Modelo com a Variável LACUM.AT Aplicado a Base de Dados	51
Quadro 10 - <i>Confusion Matrix</i> para o Modelo com a Variável LACUM.AT Aplicado ao Estado de Solvência no Terceiro Trimestre de 2016	51
Quadro 11 - Taxas de Acerto dos Modelos (Ponto de Corte = 0,35)	59
Quadro 12 - Taxas de Acerto dos Modelos para as Empresas Insolventes (Ponto de Corte = 0,35)	59

1. Introdução

A falência de uma empresa pode trazer impactos negativos para a sociedade, tanto diretamente, quando impacta aqueles ligados à empresa ou *stakeholders*, quanto indiretamente, ao gerar um efeito cascata àqueles agentes ligados indiretamente à empresa.

Do ponto de vista da teoria microeconômica, a falência de uma empresa indica que esta é ineficiente do ponto de vista social e, dessa forma, os recursos alocados em sua atividade produtiva devem passar a ser alocados de maneira diferente. No entanto, durante esse processo de realocação, gera-se um grande custo aos *stakeholders*.

Tendo o exposto acima em mente, variadas são as medidas tomadas por Governos no sentido de reduzir os custos associados a uma falência. No caso do Governo Brasileiro, uma medida relativamente importante é a Lei nº 11.101/05, que regula a recuperação judicial, extrajudicial e a falência do empresário e da sociedade empresária, também conhecida como Nova Lei de Falências. Particularmente, a Recuperação Judicial é um instrumento que permite à empresa reestruturar-se financeiramente e, portanto, anteceder a falência propriamente dita.

Alguns dos principais elementos previstos pela Lei nº 11.101/05, para o caso da Recuperação Judicial, são a concessão de prazos e condições especiais para pagamento das obrigações vencidas, a cisão, incorporação ou fusão, a alteração do controle societário, o aumento do capital social, a redução salarial e flexibilização da relação trabalhista, a venda de bens, emissão de valores mobiliários e constituição de Sociedade de Propósito Específico (SPE), tendo em vista a separação do patrimônio da empresa.

Segundo Araújo e Funchal (2009), o número de falências decretadas apresentou uma tendência de queda após a Lei nº 11.101/05, passando de uma média de 318 falências nos 12 meses anteriores à implementação da Lei nº 11.101/05 para uma média de 181, isto é, uma redução de, aproximadamente, 43%.

Nesse contexto, a predição de insolvência financeira de determinada empresa passou a ser um tema de bastante interesse nos campos da Economia, Finanças e Contabilidade, tendo em vista sua grande utilidade no contexto corporativo e do Mercado Financeiro. Desta forma, variados foram os modelos e técnicas aplicados à predição de insolvência financeira de empresas, tornando este tema também parte dos estudos em Estatística Computacional e Pesquisa Operacional.

Segundo Balcaen e Oohghe (2004), o processo de falência envolve diversas partes e pode ter um amplo custo social, de forma que a previsão de insolvência tornou-se de grande relevância não apenas para agentes privados, mas também para agentes públicos. Aos primeiros, em especial aos *stakeholders*, a capacidade de previsão de insolvência da firma de interesse pode contribuir para evitar perdas individuais, principalmente, como, por exemplo, perdas patrimoniais. Por sua vez, ao Governo, a previsão de insolvência

contribui, principalmente, para evitar o efeito cascata da falência de uma companhia, como a queda no número de empregos.

Dessa forma, a correta utilização de modelos de previsão de insolvência permite reduzir os custos associados ao processo de falência, uma vez que, identificada a elevada probabilidade de insolvência, os administradores podem mudar práticas de gestão, reformular a estrutura de capital, procurar fontes de recurso menos onerosas, entre outras atitudes. Além disso, esses modelos permitem que investidores tenham maior conhecimento a respeito da condição econômico-financeira das empresas, contribuindo para reduzir a assimetria informacional entre credores e devedores e entre administradores e acionistas, uma vez que, devedores e administradores, por estarem diretamente inseridos no contexto operacional e financeiro da firma, tendem a ter maior conhecimento sobre a situação financeira da empresa. Portanto, os modelos de previsão de insolvência permitem que investidores precifiquem mais corretamente valores mobiliários e ações.

Esta monografia está estruturada em cinco seções, além desta introdução. A seção 2 apresenta um panorama histórico dos principais trabalhos acerca dos modelos de previsão de falência. A seção 3 trata dos aspectos metodológicos utilizados na construção do modelo e a seção 4 discorre sobre o desenvolvimento do modelo em si. Finalmente, a seção 5 traz os principais resultados e conclusões do modelo desenvolvido.

2. Panorama Histórico dos Modelos de Previsão de Falência: 1930 a 2016

A intenção do presente capítulo é fazer um panorama dos principais estudos relacionados à temática do presente trabalho. Dessa forma, construiu-se uma linha histórica com base na técnica estatística utilizada pelos os principais estudos de 1930 ao presente, apontando suas principais inovações e falhas.

2.1. Estudos Iniciais: 1930 a 1960

O campo de estudo da previsão de falência, também conhecido como previsão de insolvência, ganhou bastante destaque nas últimas décadas, com aplicação de variadas técnicas e em diferentes regiões ao redor do mundo. Em realidade, os primeiros estudos sobre o tema datam da terceira dezena do século XXI.

Segundo Bellovary et al. (2007), em 1930, o *Bureau of Economic and Business Research*, da Universidade de Illinois, publicou um relatório analisando 24 índices financeiros de 29 firmas para tentar determinar características comuns a firmas em processo de falência. Neste estudo, calcularam-se os indicadores médios da amostra e compararam-se os indicadores observados individualmente com as médias das amostras, com o intuito de encontrar similaridades entre as empresas com dificuldades financeiras. O estudo encontrou oito índices considerados bons determinantes da fragilidade financeira de certa empresa. Tais índices foram: Capital Circulante Líquido / Ativos Totais, Lucros e Reservas / Ativos Totais, Patrimônio Líquido / Ativos Fixos, Ativos Fixos / Ativos Totais, Patrimônio Líquido / Ativos Totais, Vendas / Ativos Totais e Disponível / Ativos Totais.

Segundo Kurunmäki (2013), em 1931, Fitzpatrick (1931) observou treze indicadores financeiros de dezenove empresas em processo de falência e dezenove de empresas solventes, durante o período de 1920 a 1929, todas conhecidas nos Estados Unidos, indicando, dentre esses treze indicadores, quatro como sendo os melhores determinadores de falência futura.

No ano seguinte, Fitzpatrick (1932) comparou cada indicador financeiro de cada empresa com o de uma empresa bem sucedida, chegando a três indicadores como preceptores adequados de falência. Segundo Pereira e Martins (2015), os resultados de tal estudo apontam que o melhor preditor foi a razão Dívida Líquida / Patrimônio Líquido e sugerem que os indicadores financeiros de liquidez perdem importância para empresas com dívidas a longo prazo.

Ainda de acordo com Kurunmäki (2013), a década de 1930 foi marcada pelos estudos de Smith e Winakor (1930), que estudaram 21 indicadores e 183 processos de falência na década de 1920, apontando, como indicadores financeiros de relevância, Capital Circulante / Ativos Totais e Ativo Circulante / Passivo Circulante (Liquidez Corrente), e

Ramser e Foster (1931), que estudaram 173 empresas entre 1919 e 1925, indicando diferenças nos indicadores entre aquelas empresas que faliram e as saudáveis.

Em 1942, Merwin (1942) analisou um total de 582 empresas de pequeno porte em diferentes indústrias, comparando os 48 indicadores obtidos no estudo com as médias de cada indústria em que determinada firma estava inserida. Em seu estudo, Merwin observou que firmas apresentam sinais de fragilidade financeira entre quatro a cinco anos antes da falência propriamente dita e indica como indicadores de relevância, Capital Circulante Líquido / Ativos Totais, Liquidez Corrente e Patrimônio Líquido / Endividamento Total.

Em 1945, Chudson (1945) elegeu uma série de indicadores contábeis de empresas e apontou aqueles com maior ligação com empresas lucrativas, além de indicar a variação relativa de cada indicador entre indústrias diferentes e a sensibilidade de cada indicador com o crescimento da empresa.

Já em 1962, Jackendoff (1962) analisou indicadores financeiros ligados à mensuração da liquidez de curto prazo e da solvência de longo prazo. No estudo, o autor comparou indicadores entre empresas lucrativas e não lucrativas, chegando ao resultado já esperado de que, comparativamente a empresas não lucrativas, empresas lucrativas tem indicadores de liquidez maiores, maiores indicadores de Capital Circulante Líquido / Ativos Totais e indicadores menores de Endividamento / Patrimônio Líquido. Além disso, o autor afirma que há forte evidência de que firmas lucrativas tem um giro de ativos maior que firmas não lucrativas.

Segundo Pereira e Martins (2015), os artigos citados foram pioneiros no sentido de tentar encontrar quais seriam os preceptores do insucesso corporativo. Além disso, apontam, como fragilidades dos estudos: a falta de rigorosidade na definição do tamanho das empresas analisadas, a inexistência de padronização quanto à janela temporal utilizada, a grande variabilidade, entre os estudos, quanto ao correto número e tipo de indicadores utilizados, o comprometimento da qualidade das informações contábeis utilizadas, a não utilização de modelos de previsão propriamente dita, dada a baixa disponibilidade técnica e, finalmente, a instabilidade política e econômica do período do qual os dados são provenientes, como a Grande Depressão de 1929 e a Primeira e Segunda Guerras Mundiais.

Para Bellovary et al. (2007), tais estudos indicaram, de forma geral, como mais importantes, os indicadores de Capital Circulante / Ativos Totais e de Liquidez Corrente. Além disso, o autor atribui a tais estudos iniciais como essenciais para fundamentar os artigos dos anos seguintes.

Segundo Altman et al. (2013), a maior parte de tais estudos foram desenvolvidos para auxiliar no processo decisórios de concessão de crédito bancário.

Os estudos acerca da previsão de insolvência entre 1930 e meados de 1960 são limitados no sentido de não apresentarem técnicas estatísticas de forma a determinar a causalidade de cada indicador financeiro, bem como o nível de significância de cada um. Além disso, o tamanho da amostra e a falta de rigorosidade metodológica em desses estudos também contribuíram para a fragilidade dos mesmos. Por outro lado, pode-se notar que o uso de indicadores contábeis como preditores de falência, método preconizado em tais estudos iniciais, determinaram a construção de um paradigma, visto que muitos estudos posteriores recorreram, exclusivamente, a tais instrumentos para a previsão de insolvência. Embora esta tenha sido a tendência geral, inclusive na maioria dos estudos mais recentes, a utilização de indicadores externos aos Demonstrativos Financeiros das empresas pode representar um ganho de eficiência aos modelos de previsão de falência.

2.2. Estudos que Empregam Análise Discriminante

Em 1966, o estudo de Beaver (1966) revolucionou o campo da previsão de insolvência corporativa, com a introdução de um modelo de análise discriminante univariada. O estudo fez uso de uma amostra de 158 empresas (79 insolventes e 79 solventes), entre os anos de 1954 e 1964, cujos ativos variavam de pouco mais de meio milhão de dólares a 45 milhões de dólares. Além de comparar os valores médios de 30 índices contábeis, da mesma forma que outros estudos antecedentes, tal estudo testou a capacidade individual de cada variável.

Como resultado, as variáveis mais importantes, enquanto preditoras de insolvência, foram respectivamente: Lucro Líquido / Endividamento Total, Lucro Líquido / Vendas, Lucro Líquido / Patrimônio Líquido, Fluxo de Caixa / Endividamento Total e, finalmente, Fluxo de Caixa / Ativos Totais.

Dois anos depois, Beaver (1968) analisou a capacidade preditiva de quatorze índices contábeis, a partir da mesma metodologia de seu estudo anterior. Dessa vez, o indicador de maior sucesso foi Fluxo de Caixa / Passivo Total.

Segundo Deakin (1972), o estudo de Beaver foi capaz de avaliar corretamente 78% companhias cinco anos antes da falência.

De acordo com Pereira e Martins (2015), os dois estudos de Beaver deixaram alguns legados importantes. Primeiramente, nem todos os indicadores utilizados possuem o mesmo nível de capacidade explicativa. Segundamente, prever empresas solventes é mais exato do que prever empresas insolventes.

Para Scott (1981), o segundo estudo de Beaver identificou os dois tipos de erro para modelos de previsão de insolvência, sejam eles o erro tipo I, que consiste em classificar erroneamente uma empresa insolvente como solvente e o erro tipo II, que, por sua vez,

consiste em classificar erroneamente uma empresa solvente como insolvente. Segundo Castro (2003), o estudo de Beaver considerou que o custo dos erros do Tipo I e do Tipo II seria o mesmo, o que seria incorreto, visto que a classificar erroneamente uma empresa insolvente como solvente é, de forma geral, mais custoso.

Seguidamente a Beaver (1966), outros estudos que fizeram uso da mesma técnica foram Tamari (1966) e Moses e Liao (1987).

Ainda em 1968, Altman (1968) produziu um estudo que se tornou seminal nos estudos de previsão de insolvência corporativa. Neste estudo, foi utilizada a técnica da análise discriminante multivariada. A amostra utilizada foi composta por 66 empresas, metade das quais classificadas como solventes e a outra metade, como insolventes. O valor dos ativos de tais empresas estava entre pouco menos de um milhão de dólares a quase 26 milhões de dólares.

Como *proxy* de condição de insolvência, o autor utilizou o evento da empresa em questão ter preenchido um pedido de falência de acordo com o *Chapter X do National Bankruptcy Act*, durante o período de 1946 a 1965. Foram eliminadas aquelas empresas consideradas pequenas, com ativo total inferior a US\$ 1 milhão e aquelas consideradas muito grandes.

Como forma de selecionar os indicadores financeiros utilizados, Altman (1986) fez uso de um primeiro filtro baseado na popularidade do indicador na literatura e na potencial relevância do mesmo para o estudo. Em seguida, o autor selecionou as variáveis com base em testes de significância estatística, na correlação entre cada uma (a fim de evitar multicolinearidade entre variáveis), na precisão preditiva na cada variável, medida pelo erro-padrão e, finalmente, no juízo pessoal.

No modelo final, batizado de Z-Score, Altman (1986) utilizou os seguintes índices: Capital Circulante Líquido / Ativo Total, Lucros Retidos / Ativo Total, EBIT / Ativo Total, Patrimônio Líquido “marcado a mercado” / Endividamento “marcado na curva” e Vendas / Ativo Total. O modelo calculado foi o seguinte:

$$Z = 0,012 X1 + 0,014X2 + 0,033X3 + 0,006X4 + 0,999X5 \quad (1)$$

Onde:

Z = Indicador de Propensão à Falência

X1 = Capital Circulante Líquido / Ativo Total;

X2 = Lucros Retidos / Ativo Total;

X3 = EBIT / Ativo Total;

$X4 = \text{Patrim\~{o}nio L\~{i}quido a valor de mercado} / \text{D\~{i}vida total a valor cont\~{a}bil};$

$X5 = \text{Vendas} / \text{Ativo Total}$

O autor ainda fez uso de testes de signific\~{a}ncia, a partir de um teste F de Snedecor, e notou que, com exce\~{c}\~{a}o do indicador Vendas / Ativo Total, as outras vari\~{a}veis foram conjuntamente significantes a um n\~{i}vel de confian\~{c}a de 0,001%. Al\~{e}m disso, o autor notou que as vari\~{a}veis medidoras de rentabilidade, foram aquelas com maior poder discriminante.

Como resultado, o autor concluiu que o modelo foi capaz de classificar corretamente as empresas em mais de 70% das vezes, inclusive quando avaliadas anos antes do momento de pedido de fal\~{e}ncia. Segundo Deakin (1978), para as classifica\~{c}\~{o}es feitas no ano anterior \~{a} fal\~{e}ncia, o \~{i}ndice de acerto elevou-se para 95%.

Castro (2003) destaca que o estudo de Altman (1968), ao incluir a an\~{a}lise multivariada, abriu espa\~{c}o para que diversas medidas fossem combinadas na previs\~{a}o de insolv\~{e}ncia. Neste sentido, o autor cita um exemplo do artigo de Altman, em que este afirma que uma empresa de baixa lucratividade, quando avaliada apenas por este aspecto, seria considerada insolvente, mesmo que apresentasse excelentes indicadores de liquidez.

Avaliando os resultados de ambos os estudos, Deakin (1972) reuniu os resultados de ambos e tra\~{c}ou o seguinte paralelo entre eles: cinco anos antes da fal\~{e}ncia, o estudo de Altman (1968) errou na classifica\~{c}\~{a}o de 64% das empresas, enquanto Beaver (1968) errou apenas 22%. Quatro anos antes, Altman (1968) errou 71% e Beaver (1968), 24%. Tr\~{e}s anos antes, os erros foram, respectivamente, 52% e 23%. Dois anos antes, 28% e 21%. Finalmente, no ano anterior, a an\~{a}lise discriminante multivariada de Altman errou apenas 5%, enquanto a an\~{a}lise discriminante univariada de Beaver errou em 13% das classifica\~{c}\~{o}es. Tal avalia\~{c}\~{a}o parece indicar uma maior efici\~{e}ncia do estudo de Altman (1968) para o curto prazo, embora o estudo de Beaver (1968) tenha tido porcentagem de erro menor para prazos mais longos. Ambos os estudos parecem indicar uma maior taxa de acerto para previs\~{o}es mais pr\~{o}ximas.

Segundo Castro (2013), o modelo de Altman contribuiu bastante com os estudos de previs\~{a}o de insolv\~{e}ncia, ao demonstrar que a an\~{a}lise discriminante m\~{u}ltipla, embora tenha tido resultados insatisfat\~{o}rios para per\~{i}odos maiores que dois anos, \~{e} mais adequada e robustas em suas conclus\~{o}es.

Em 1977, Altman (1977) construiu um novo modelo, batizado de *Zeta*, de forma a melhorar e aperfei\~{c}oar seu antigo modelo *Z-Score* da d\~{e}cada anterior. Segundo Altman (2000), o modelo original deveria ser mudado por cinco principais motivos: mudan\~{c}as no perfil financeiro das firmas e das fal\~{e}ncias, incorpora\~{c}\~{a}o de dados mais recente, constru\~{c}\~{a}o de um modelo mais geral, capaz de classificar empresas de qualquer setor, inclus\~{a}o de dados cont\~{a}beis mais recentes, de acordo com pr\~{a}ticas cont\~{a}beis mais

recentes e, finalmente, aplicar atualizações e avanços mais recentes na técnica da análise discriminante.

Os autores construíram uma amostra de 53 firmas insolventes, pareadas com outra amostra de 58 empresas saudáveis. Diferentemente do modelo *Z-Score*, a amostra deste segundo estudo incluiu firmas industriais e varejistas, no período de 1969 a 1975. Além disso, este segundo modelo incluiu apenas empresas de maior porte, com, no mínimo 20 milhões de dólares em ativos.

Foram escolhidas 27 variáveis, divididas nos seguintes grupos: lucratividade, cobertura e alavancagem, liquidez, capitalização e variáveis miscelâneas. Destas, apenas sete figuraram no modelo, escolhidas por meio de uma série de testes estatísticos. Estas foram: Lucros após Despesas Financeiras e Impostos / Ativos Totais, o desvio-padrão do Lucro, Serviço da Dívida, Lucros Retidos / Ativos Totais, Liquidez Corrente, Patrimônio Líquido / Ativos Totais e logaritmo dos Ativos Totais, como forma de representar o tamanho da firma.

Segundo Castro (2013), tendo em vista que o modelo *Zeta* foi desenvolvido em parceria com uma empresa para fins comerciais, não foi divulgada a equação final do mesmo. Mesmo assim, os resultados obtidos demonstram que o Modelo *Zeta* mostrou-se superior ao modelo *Z-Score*, principalmente nos anos 2 a 5 anos anteriores à falência.

No que tange a estes primeiros modelos, estes fizeram uso da técnica conhecida como análise discriminante, seja ela univariada, como no estudo de Beaver (1966) ou multivariada, como no estudo de Altman (1968).

De acordo com Maroco (2003), a análise discriminante tem como objetivos principais identificar aquelas variáveis com maior diferenciação entre dois ou mais grupos mutuamente exclusivos, bem como para desenvolver uma função discriminante que sirva para classificar a priori indivíduos em algum destes grupos.

Segundo Hair et al. (2005), a análise discriminante é aplicável aos casos em que a amostra pode ser dividida em grupos baseados em uma variável dependente qualitativa, isto é, não métrica. Dessa forma, o objetivo principal dessa técnica é prever a probabilidade de que um indivíduo pertença a um grupo particular da amostra, a partir de variáveis explicativas quantitativas. No contexto dos modelos de previsão de insolvência, usualmente, estes grupos resumem-se ao grupo das empresas solventes e ao grupo das empresas insolventes. A discriminação é atingida por meio da maximização da variância entre os grupos e, conseqüentemente, minimização da variância intragrupo.

No caso dos estudos sobre previsão de insolvência, a variável dependente é binária, indicando dois estados possíveis para cada indivíduo: solvente ou insolvente.

Ainda de acordo com Hair et al. (2005), pode-se testar a significância estatística da função discriminante por meio da medição da distância entre os centroides dos grupos.

Segundo Selau e Ribeiro (2009), uma vez verificada a significância do modelo, costuma-se criar matrizes de classificação para analisar o poder discriminatório da função. Nesse sentido, é necessária a criação de um escore de corte, de forma a classificar cada observação em determinado grupo. Os autores sugerem a obtenção do escore por meio de uma média ponderada dos centroides dos grupos, segundo a seguinte equação:

$$Z_c = \frac{N_A Z_B + N_B Z_A}{N_A + N_B} \quad (2)$$

Consoante Hair et al. (2005), certas condições são requeridas para a aplicação da análise discriminante, de forma que as mais importantes são a normalidade das variáveis independentes e a igualdade nas matrizes de dispersão e covariâncias dos grupos. A queda dessas hipóteses pode trazer problemas na estimação da função discriminante e no processo de classificação.

Por outro lado, Büyükoztürk e Çokluk-Bökeoğlu (2008), afirmaram que, atendidas as hipóteses da análise discriminante, o que, em alguns casos, pode ser facilmente obtido, a técnica constitui um instrumento poderoso matematicamente e conceitualmente.

Após os estudos pioneiros de Beaver (1968) e Altman (1968), o número de estudos de previsão de insolvência baseados em análise discriminante cresceu explosivamente, sendo esta técnica utilizada inclusive em anos mais recentes. Segundo Bellovary et al. (2007), de 1960 para 1970, o número de modelos deste tipo aumentou para 22, passando para 28 em 1980. A partir do final dos anos 80, a popularidade dos modelos baseados em análise discriminante declinou, com a maior utilização dos modelos *logit* e das redes neurais. Com isso, em 1990, ainda segundo Bellovary et al. (2007), o número diminuiu para 9 e, nos anos 2000, esse número foi de apenas dois.

De qualquer forma, ao longo do tempo, os novos modelos de análise discriminante passaram a levar em conta aplicações em setores específicos, índices financeiros inéditos, utilização de *softwares* e pacotes estatísticos (*SPSS*, *SAS*, *BMPD*, *LISREL*, *STATA*, etc.) (PEREIRA E MARTINS, 2015). Dentre estes, segundo levantamento realizado por Pereira e Martins (2015), aqueles com maior nível de eficiência, medida pela quantidade de empresas insolventes corretamente classificadas pelo modelo, são, em ordem decrescente: Mensah (1983), Marais (1980), Betts e Belhoul (1982), Patterson (2001), Hennway e Morris (1983), Norton e Smith (1979), Taffler (1977), Diamond Jr. (1976), Fulmer et al. (1984), Betts e Belhoul (1983), Gardiner et al. (1996) e Taffler (1980), todos com um nível de eficiência superior a 95%.

No entanto, cabe ressaltar que a simples ordenação pelo nível de eficiência não é suficiente para determinar que determinado modelo seja superior a outro ou que a análise discriminante é preferível a outros métodos, uma vez outros fatores devem ser levados em conta, tais como tamanho da amostra utilizada, setor econômico em que o modelo foi aplicado e com que quantidade de anos de antecedência o modelo foi capaz

de prever a insolvência. Além disso, é notável também o fato de que muitos modelos, embora capazes de classificar empresas como insolventes com um razoável grau de acerto, não mantêm o mesmo nível de eficiência ao classificar empresas como solventes ou vice-versa.

Ainda no que tange eficiência, Pereira e Martins (2015) citam também que existem limitações dos modelos advindas dos dados usados, o que pode sugerir que a capacidade preditiva dos modelos esteja sendo superestimada.

2.3. Estudos que Empregam o Modelo *Logit*

No final da década de 1970, alguns estudos de previsão de insolvência adotaram a análise *logit* (regressão logística), porém, apenas da década de 1980, estes passaram a substituir os modelos de análise discriminante, em termos de popularidade. Conforme citado anteriormente, a análise discriminante exige alguns pressupostos fortes, como a normalidade das variáveis independentes e a igualdade nas matrizes de dispersão e covariâncias de ambos os grupos, solventes e insolventes, o que acabou por fazer com que vários autores decidem pelo uso da regressão logística em detrimento da análise discriminante linear (BELLOVARY et al., 2007).

O modelo *logit* ganhou bastante popularidade nas mais variadas áreas científicas, entre as quais a Biologia, a Epidemiologia, a Medicina e a Economia (SANTOS, 2010).

Um dos primeiros e mais importantes estudos a utilizar a análise *logit* na previsão de falência corporativa foi o estudo de Ohlson (1980), que utilizou dados de empresas da segunda metade da década de 1970 e, segundo o próprio autor, fez uso de estimativas probabilísticas de falência desenvolvidas lógica e sistematicamente.

O autor elenca algum dos motivos para ter escolhido a análise *logit* e não a análise discriminante múltipla. Entre eles, ele cita: (i) os fortes pressupostos da análise discriminante, já citados anteriormente, (ii) a questão do resultado da análise discriminante ter pouca interpretação intuitiva e (iii) problemas relacionados com os procedimentos de pareamento dos grupos, isto é, o processo de escolher, para cada firma insolvente, uma solvente, segundo determinado critério. Segundo o autor, seria mais produtivo utilizar variáveis como preceptores e não como critérios de pareamento.

As variáveis utilizadas para o estudo foram: logaritmo de Ativos Totais / PIB, Passivo Exigível / Ativos Totais, Capital de Giro / Ativos Totais, Passivo Circulante / Ativo Circulante, uma variável binária igual a um se o Passivo Exigível for maior que os Ativos Totais e igual a zero, caso contrário, Lucro Líquido / Ativos Totais, Fluxo de Caixa Operacional / Endividamentos Totais, uma variável binária igual a um se o Lucro Líquido tiver sido negativo nos últimos dois anos e zero, caso contrário e uma variável responsável por medir mudanças no Lucro Líquido.

Segundo Rey (2003), entre outros estudos, o estudo de Ohlson (1980) foi pioneiro ao utilizar variáveis que consideraram o fluxo de caixa das empresas, que são indicadoras da eficiência operacional da companhia e da capacidade dos administradores da mesma. Segundo o mesmo autor, um importante motivo para que tais variáveis não terem sido utilizadas em estudos anteriores foi o avanço da Contabilidade ocorrida em meados da década de 1970 e a exigência de divulgação de relatórios financeiros de Fluxo de Caixa no período.

A utilização de variáveis de Fluxo de Caixa mostra-se como um recurso bastante interessante aos modelos de previsão de insolvência, uma vez que, ao utilizar somente variáveis do Balanço Patrimonial e da Demonstração de Resultados do Exercício, retrata-se a empresa como ela está no momento de apuração dos dados, enquanto que variáveis do Fluxo de Caixa permitem criar uma perspectiva quanto à situação econômico-financeira futura da firma. No entanto, segundo Casey e Bartczak (1985), existem algumas limitações preditivas na utilização do Fluxo de Caixa como variável explicativa.

O autor realizou testes *t* de *Student* para avaliar se as variáveis escolhidas são estatisticamente significantes e o resultado foi que as variáveis logaritmo de Ativos Totais / PIB, Passivo Exigível Total / Ativos Totais, Lucro Líquido / Ativos Totais, Fluxo de Caixa Operacional / Ativos Totais, Capital Circulante / Ativos Totais são todas individualmente significantes ao nível de 95% de confiança. O modelo final, que ficou batizado de *O-Score*, alcançou um nível de eficiência de 96% e é representado abaixo:

$$O = -1,32 - 0,407 \log \left(\frac{X_{1t}}{\text{PIB}} \right) + 6,03 \left(\frac{X_{2t}}{X_{1t}} \right) - 1,43 \left(\frac{X_{3t}}{X_{1t}} \right) + 0,0757 \left(\frac{X_{4t}}{X_{5t}} \right) - 1,72Y_1 - 2,37 \left(\frac{X_{7t}}{X_{1t}} \right) - 1,83 \left(\frac{X_{8t}}{X_{2t}} \right) + 0,285Y_2 - 0,521 \left(\frac{X_{7t} - X_{7t-1}}{|X_{7t}| + |X_{7t-1}|} \right) \quad (3)$$

Onde:

O = Probabilidade de Insolvência

X1 = Ativos Totais;

X2 = Passivo Exigível;

X3 = Capital de Giro;

X4 = Passivo Circulante;

X5 = Ativo Circulante;

Y1 = 1 se X2 > X1, 0 caso contrário;

X7 = Lucro Líquido;

X8 = Fluxo de Caixa Operacional;

$Y_2 = 1$ se $X_7 < 0$ pelos últimos dois anos, 0 caso contrário.

Segundo Wooldridge (2010), no modelo *logit*, emprega-se a função de ligação logística:

$$G(z) = \frac{\exp(z)}{1+\exp(z)} \quad (4)$$

que tem seu domínio no intervalo de zero a um.

Em razão da natureza não linear do modelo, não se pode estimá-lo por Mínimos Quadrados Ordinários (MQO). Dessa forma, em geral, utiliza-se o estimador de Máxima Verossimilhança (EMV). Cabe lembrar que, em boa parte das aplicações econométricas, MQO é preferível, tendo vista que, atendidas as hipóteses do modelo clássico de regressão linear (MCRL), MQO é o melhor estimador não viesado de todos os outros estimadores. Entretanto, a não linearidade citada anteriormente infringe uma das hipóteses do MCL (WOOLDRIDGE, 2010).

Cada parâmetro estimado por EMV pode ser testado por meio de testes *t*, o que permite selecionar quais razões contábeis deverão ser utilizadas no modelo estimado. Além disso, segundo Wooldridge (2010), pode-se também realizar testes de hipóteses múltiplas, como o teste da razão de verossimilhança (RV).

Além do estudo de Ohlson (1980), pode-se citar, segundo levantamento de Pereira e Martins (2015), em ordem decrescente de nível de eficiência medida pela quantidade de empresas insolventes corretamente classificadas pelo modelo: Meyer e Pifer (1970), Mensah (1983), Dambolena e Shulman (1988), Wilcox (1973), Platt et al. (1994), Forsyth (1991) e Lau (1987).

Ao contrário do que aconteceu com a análise discriminante linear múltipla, tais estudos figuram em uma faixa ligeiramente menor em termos de nível de eficiência, sendo todos com um nível de acerto superior, pelo menos, a 90%. No entanto, conforme citado anteriormente, a simples análise do nível de eficiência é incapaz de determinar assertivamente quais os melhores e os piores modelos.

2.4. Estudos que Empregam Redes Neurais e Análise por Envoltória de Dados

De acordo com Pereira e Martins (2015), na década de 1990, com a evolução computacional ocorrida no período e com o surgimento de técnicas estatísticas-computacionais conhecidas como de terceira geração, tais como algoritmos, programação computacional e inteligência artificial, os modelos de previsão de insolvência passaram a contar com duas técnicas mais modernas: a Análise por Envoltória de Dados e as Redes Neurais.

Em 1988, Messier e Hansen (1988) utilizaram Redes Neurais ID3 para prever regras de previsão para inadimplência de empréstimos e casos de falência. Foi utilizada uma amostra de treinamento de oito empresas insolventes e quinze empresas solventes australianas. Segundo Ko et al. (2001), neste estudo, as regras derivadas pelo mecanismo ID3 classificaram, corretamente, as empresas insolventes com altíssima eficiência e com 87,5% de acerto para os casos de inadimplência de empréstimos.

Em 1985, Frydman (1985) fez uso de uma amostra de 58 firmas insolventes juntamente com 142 firmas solventes, para determinar regras de previsão, a partir de um algoritmo recursivo de particionamento, corretamente classificando todas as empresas da amostra de treinamento.

Em 1993, Odom e Sharda (1993) utilizaram os mesmos indicadores financeiros utilizados por Altman (1968) e, a partir da aplicação de uma rede artificial de neurônios em uma amostra de 65 firmas falidas e 64 firmas não falidas. A rede utilizada foi uma rede *back-propagation* e conseguiu classificar, corretamente, todas as empresas.

Em 1994, Altman et al. (1994) desenvolveram um modelo fazendo uso de redes neurais. Segundo Castro (2013), tal estudo é especialmente importante por ter sido escrito por um dos maiores pesquisadores do campo da previsão de insolvência, Edward Altman, atualmente professor da *New York University*, o que acentua a tendência técnica dos modelos de previsão de falência. Neste estudo, os autores utilizaram mais de 1000 empresas de médio e pequeno porte e utilizaram uma rede com arquitetura de retropropagação. Foram utilizados dez indicadores financeiros, divididos em três grupos: estrutura de capital, liquidez e lucratividade.

Segundo Akkoç (2012), redes neurais são técnicas computacionais fundamentadas na estrutura neural de organismos inteligentes que são treinadas e adquirem conhecimento por meio da repetição. Tais redes neurais artificiais são desenvolvidas por meio de modelos matemáticos e neles, algumas suposições são mantidas, como, por exemplo, processamento de informações em unidades denominadas neurônios, transmissão de estímulos por meio de conexões, associadas a um peso e resposta da rede determinada pela função de ativação, na qual cada neurônio tem determinada contribuição. (FAUSETT, 1944). As redes apresentam a capacidade de aprender por meio de exemplos e de inferir resultados a partir disso, melhorando sua capacidade.

De acordo com Castro (2013), as redes neurais artificiais simulam simplificada o funcionamento do sistema nervoso humano, utilizadas, em processos de reconhecimento de padrões para associação e classificação, motivando seu uso em modelos de previsão de insolvência, para classificar cada firma como solvente ou não. Essa semelhança ao cérebro humano que motiva, inclusive, sua denominação e a denominação de seus instrumentos.

Segundo Gouvea et al. (2013), cada rede possui uma forma, denominada arquitetura e o método para determinação dos pesos, denominado algoritmo de aprendizado.

A arquitetura é dividida e classificada em três grupos: camada de entrada, na qual se apresenta os padrões à rede, camada intermediária, na qual é realizado o processamento, por meio dos pesos associados às conexões e, finalmente, a camada de saída, na qual se apresenta o resultado final.

O algoritmo de aprendizado são as regras utilizadas para solucionar um problema de aprendizado específico. Pode-se dividir e classificar os tipos de aprendizado em três tipos: aprendizado supervisionado, no qual a resposta esperada é indicada para a rede, aprendizado não supervisionado, no qual a rede baseia-se unicamente nos estímulos recebidos e, finalmente, aprendizado por reforço, em que existe a figura de um avaliador externo ao comportamento da rede.

Segundo Trippi e Lee (1996), as redes neurais não precisam atender hipóteses muito restritivas tais como a normalidade das variáveis independentes do modelo de análise discriminante, e a multicolinearidade é um problema bem menor. Por outro lado, não é possível discriminar o peso específico de cada um dos índices financeiros para a resposta do modelo, o que constitui uma das maiores críticas aos modelos de previsão de insolvência baseados em redes neurais.

Segundo levantamento realizado por Pereira e Martins (2015), em ordem decrescente por nível de eficiência, isto é, porcentagem de empresas insolventes corretamente classificadas, podemos citar os seguintes estudos: Frydman et al. (1985), Messier Jr. E Hansen (1988), Guan (1993), Theodossiou (1993), Andres et al. (2005), Mckee (1995), Wilson e Sharda (1994), Rudorfer (1995), Wilson et al. (1995) e Dwyer (1992).

2.5. Estudos que Empregam Outros Modelos de Risco

Embora não seja o foco do presente trabalho, cabe citar alternativas para assessorar o problema da antecipação do risco de crédito: os modelos de *CreditMetrics*, *CreditRisk+* e o Modelo KMV.

Segundo Santos e Santos (2009), o modelo *CreditMetrics*, desenvolvido pelo banco de investimentos J.P. Morgan, tem a função primária de estimar a distribuição de valor de determinado portfólio de ativos sujeitos a variações de valor, incluindo a própria inadimplência, e o modelo *CreditRisk+* foca o risco de *default* por meio da classificação do devedor em categorias de risco, cada qual associada a determinada probabilidade de inadimplência.

Por sua vez, o modelo KMV, fruto do melhoramento do modelo de Merton (1974) parte do pressuposto que o preço das ações em mercado sinaliza expectativas de agentes / a própria firma, de forma que o risco de crédito é dado pela volatilidade do valor de mercado dos ativos da firma. Nesse contexto, os ativos da determinada firma são

tratados como ativos subjacentes de um derivativo, mais especificamente de uma opção europeia, de forma que, em caso de inadimplência, os mesmo são repassados ao credor.

Baseando-se na condição de mercado eficiente, o modelo KMV determina a probabilidade de *default*, por meio da estimação do valor dos ativos, da volatilidade do retorno sobre os ativos, do ponto de *default*, da distância ao *default* e da probabilidade de *default* (SANTOS e SANTOS, 2009).

De acordo com Chen e Chu (2014), o modelo KMV parte da mesma lógica de estimação do modelo de precificação de opções Black-Scholes-Merton (1997).

2.6. Estudos Brasileiros Seminais

No caso brasileiro, os modelos de previsão de insolvência tornaram-se mais conhecidos devido ao modelo conhecido como Termômetro de Kanitz (1974). Kanitz (1974) analisou, aproximadamente, 5.000 demonstrações contábeis de empresas brasileiras, escolhendo, aleatoriamente 21 empresas falidas entre 1972 e 1974. Partindo do pressuposto da análise discriminante múltipla, o autor selecionou 21 empresas solventes, referentes aos mesmos anos. Com isso, o mesmo determinou a seguinte fórmula:

$$Y = (0,05 RP + 1,65LG + 3,55LS) - (1,06LC + 0,33GE) \quad (5)$$

Onde:

Y = Propensão à Falência

RP = Rentabilidade do Patrimônio (Lucro Líquido / Patrimônio Líquido);

LG = (Ativo Circulante somado ao Ativo Realizável a Longo Prazo / o Passivo Circulante somado ao Passivo Exigível a Longo Prazo);

LS = Liquidez Seca (Ativo Circulante menos Estoques / Passivo Circulante);

LC = Liquidez Corrente (Ativo Circulante / Passivo Circulante);

GE = Grau de Endividamento (Passivo Circulante somado ao Passivo Exigível a Longo Prazo / Ativo Total).

Em 1976, foi desenvolvido por Elizabetsky (1976) um modelo de análise discriminante múltipla com o objetivo de aprimorar o processo de concessão de crédito. Para tanto, foi elaborada uma amostra de 373 empresas, das quais 274 eram classificadas como solventes e 99 como insolventes. Como *proxy* de insolvência, utilizou-se o atraso em pagamentos. A equação final criada foi a seguinte:

$$Z = 1,93X_1 - 0,20X_2 + 1,02X_3 + 1,33X_4 - 1,12X_5 \quad (6)$$

Onde:

Z = Propensão à Falência

X_1 = Lucro Líquido / Vendas;

X_2 = Disponível / Ativo Permanente;

X_3 = Contas a Receber/ Ativo Total;

X_4 = Estoque/Ativo Total;

X_5 = Passivo Circulante/Ativo Total.

Em 1978, foi desenvolvido por Matias (1978) mais um modelo para previsão de insolvência baseado na análise discriminante. Assim como o modelo de Elizabetsky (1976), o modelo de Matias foi elaborado para facilitar o processo decisório de concessão de crédito (MATIAS, 1978).

A equação do modelo final é a seguinte:

$$Z = 23,792X_1 - 8,260X_2 - 8,868X_3 - 0,764X_4 + 0,535X_5 + 9,912X_6 \quad (7)$$

Onde:

Z = Propensão à Falência

X_1 = Patrimônio Líquido / Ativo Total;

X_2 = Empréstimos / Ativo Circulante;

X_3 = Fornecedores / Ativo Total;

X_4 = Ativo Circulante / Ativo Total;

X_5 = Margem Operacional (Lucro Operacional /Lucro Bruto);

X_6 = Disponível / Ativo Total.

Em 1979, foi elaborado no Brasil o modelo de Altman et al. (1979), baseado no modelo ZETA do próprio autor. Para tanto, foram analisadas 58 firmas, 23 insolventes e 35 solventes (ALTMAN et al., 1979). A equação encontrada foi a seguinte:

$$Z = -1,44 + 4,03X_1 + 2,25X_2 + 0,14X_3 + 0,42X_4 \quad (8)$$

Onde:

Z = Propensão à Falência

X_1 = Lucros Acumulados / Ativo Total;

$X_2 = \text{Lucros antes de Juros e Imposto de Renda} / \text{Ativo Total};$

$X_3 = \text{Patrimônio Líquido} / \text{Passivo Exigível Total};$

$X_4 = \text{Vendas} / \text{Ativo Total}.$

Em 1982, Pereira (1982) desenvolveu um modelo com o intuito de facilitar o processo de concessão de crédito. Seu modelo apresentou um nível de eficiência de classificação de 90% para empresas solventes e de 86% para empresas insolventes. A equação final é a descrita a seguir:

$$Z = 0,722 - 5,124E23 + 11,016L19 - 0,342L21 - 0,048L26 + 8,605R13 - 0,004R29 \quad (9)$$

Onde:

Z = Propensão à Falência

E23 = Duplicatas Descontadas/Duplicatas a Receber;

L19 = Estoque (Final)/Custo das Mercadorias Vendidas;

L21 = Fornecedores/Vendas;

L26 = Estoque Médio/Custo das Mercadorias Vendidas;

R13 = (Lucro Operacional + Despesas Financeiras)/(Ativo Total – Investimento Médio)

R29 = Exigível Total/(Lucro Líquido + 0,1Imobilizado Médio + Saldo Devedor da Correção Monetária)

Podemos notar que as variáveis utilizadas em Pereira (1982) são um pouco menos usuais e não são facilmente encontradas em outros estudos do tema.

2.7. Estudos Brasileiros Recentes

Mais recentemente, nos últimos dez anos, outros estudos brasileiros abordaram o tema de Previsão de Insolvência, trazendo técnicas estatísticas mais modernas e levando em consideração procedimentos contábeis mais atualizados.

Como exemplo, podemos citar o artigo de Guimarães e Moreira (2008). Estes autores construíram um modelo baseado em análise discriminante múltipla, a partir de uma amostra de 116 empresas de capital aberto de 17 setores diferentes, no período de 1994 a 2003. As empresas foram divididas em dois grupos para cada setor, cada grupo composto por 35 empresas. Além disso, foi definida uma amostra de controle para, posteriormente, testar o modelo.

Como *proxy* de insolvência, os autores basearam-se no estudo de Requião (1998), no que seria chamado falência presumida: evento em que a empresa passa a ter Passivo a Descoberto, isto é, Patrimônio Líquido negativo.

O modelo estimado foi o seguinte:

$$Y = 0,526 + 1,6791\text{EstA} - 1,8309\text{EstC} + 0,012 \text{RCMCPd} + 2,0812 \text{EBITCT} \quad (10)$$

Onde:

Y = Propensão à Falência

EstA = Capital de Giro Líquido/Ativo Total;

EstC = (Passivo Circulante + Exigível a Longo Prazo)/Patrimônio Líquido;

EBITCT = Lucros Antes de Juros e Imposto de Renda/(Passivo Circulante + Exigível a Longo Prazo);

RCMCPd = (Patrimônio Líquido – Ativos de Baixa Liquidez)^{1/3}

O modelo proposto atingiu alto índice de acerto: 88,6%.

Ainda em 2008, Assaf e Brito (2008) construíram um modelo baseado em Regressão Logística a partir de uma amostral de 60 empresas, 30 insolventes que se tornaram concordatárias ou falidas entre 1994 e 2004 e 30 solventes. Para categorizar as empresas da amostra, os autores utilizaram dados do cadastro da Comissão de Valores Mobiliários – CVM.

Os autores testaram 25 índices frequentemente utilizados em modelos de previsão de insolvência, divididos em seis categorias: liquidez, rentabilidade, de atividade, de estrutura, de análise dinâmica e de fluxo de caixa. Os índices foram calculados com base nos demonstrativos do penúltimo exercício anterior ao ano do evento de *default*. Não foram consideradas demonstrações do último exercício, pois estas poderiam estar refletindo a situação concordatária da empresa, o que prejudicaria a qualidade do modelo.

Foram empregados testes t de Student para verificar se as diferenças das médias são estatisticamente significativas. O resultado do teste indica que, ao nível de confiança de 5%, não pode ser rejeitada a hipótese nula de igualdade de médias para os índices Retorno do Ativo (ROA) Giro do Ativo, Margem Operacional, Estoque/Ativo Total, Necessidade de Capital de Giro, Fluxo de Caixa Operacional / Ativo, Fluxo de Caixa Operacional/Exigível Total e Fluxo de Caixa Operacional / Endividamento Financeiro.

As variáveis definitivas foram selecionadas pelo método de *forward stepwise* e pelo teste do *Likelihood Ratio*. O modelo estimado foi o seguinte:

$$\ln\left(\frac{p}{1-p}\right) = -4,740 - 4,528X1 + 18,433X2 - 14,080X3 - 11,028X4 \quad (11)$$

Onde:

p = Probabilidade de Insolvência

X1 = Lucros Retidos/Ativo;

X2 = Endividamento Financeiro;

X3 = Capital de Giro Líquido;

X4 = Saldo de Tesouraria / Vendas.

Do grupo de solventes, 28 empresas foram classificadas corretamente e 2 erradamente, representando um acerto de 93,3%. Em relação ao grupo de insolventes, houve 26 classificações corretas e 4 erradas, correspondendo a uma taxa de acerto de 86,7%.

A validação do modelo foi feita por meio do método de múltiplas sub-amostras denominado *jackknife* de Lacherbruch (1967). Segundo os autores, o percentual de acerto acumulado foi de 88,3%. Além disso, também foi construída uma Curva ROC, segundo Oliveira e Andrade (2002), que indicou um excepcional poder de discriminação do modelo.

O artigo construiu um modelo bastante útil para prever insolvência financeira de empresas e chama atenção pelo elevado nível técnico e qualitativo. A utilização da regressão logística parece ser muito apropriada pelo motivo de ser menos restritiva que a análise discriminante múltipla. A utilização dos diversos testes para seleção de variáveis, testes de ajuste do modelo e testes de validação também foram muito bem empregados.

A alta taxa de acerto do modelo parece indicar uma elevada qualidade preditiva do mesmo.

Entretanto, dois fatores aparentaram ser um pouco mais negativos, a saber, a pequena amostra utilizada, embora os autores justifiquem tal condição, e a não seleção de uma amostra teste para realizar o *back-testing* do modelo, a exemplo de outros artigos.

Em 2009, Guimarães e Alves (2009) aplicaram o modelo Z de Altman (1968) e construíram um modelo próprio de Regressão Logística em uma amostra de 706 empresas, todas do setor de planos de saúde.

O artigo utiliza dados financeiros de companhias em 2004 e seu estado de solvência em 2005, com o propósito de modelar uma Regressão Logística para estimar a probabilidade de insolvência de uma empresa em um ano. O modelo é, então, validado em uma amostra de dados em 2005 e seu estado de solvência em 2006 e, finalmente,

comparado com o modelo Z de Altman. Foram consideradas insolventes empresas cujo valor de Patrimônio Líquido é inferior ou igual à zero.

Foram selecionadas diversas variáveis e eliminadas aquelas com correlação muito alta entre si, a fim de evitar a multicolinearidade dos dados. São aplicados testes de Wald, de Log-Likelihood ratio e de Hosmer-Lemeshow, que indicaram, respectivamente, uma relação significativa entre o conjunto de variáveis selecionadas e o evento de insolvência e uma aderência aceitável das estimativas do modelo.

Segundo os autores, a taxa de acerto do modelo de Altman foi de 68,77%, enquanto o do modelo desenvolvido, de 71,60%.

O artigo apresentou um caso prático muito útil na análise de previsão de insolvência de empresas. A comparação com o modelo de Altman no final do artigo também significa um aspecto bastante interessante: a aplicação de um modelo especificamente para um setor possivelmente gera resultados mais fidedignos do que a aplicação de um modelo geral.

Os modelos discutidos no presente estudo, de uma maneira geral, são denominados modelos de classificação, tendo em vista que a variável resposta Y é qualitativa. Segundo James et al. (2000), os três tipos de modelos mais utilizados para a finalidade de modelos de classificação são a Regressão Logística e a Análise Discriminante Linear Múltipla. Em termos de modelos mais computacionalmente intensivos, temos modelos aditivos generalizados, *trees*, *random forests* e *boosting* e *support vector machines*.

Um questionamento pertinente diz respeito a quais seriam os motivos que favoreceriam a Análise Discriminante Linear Múltipla em relação à Regressão Logística ou vice-versa. Quando existem mais de dois grupos para a variável dependente, a Análise Discriminante Linear Múltipla é absolutamente preferível à Regressão Logística. No entanto, quando a variável dependente é composta por dois grupos, existem vários motivos pelos quais Regressão Logística torna-se preferível. Como já citado anteriormente, a Análise Discriminante Linear Múltipla depende de alguns pressupostos considerados relativamente fortes, tais como a normalidade das variáveis explicativas e a igualdade das matrizes de variância-covariância dos grupos, enquanto que a Regressão Logística não depende de supostos tão rígidos, o que faz tal método seja mais aplicável em muitos casos. Além disso, mesmo quando atendidos os pressupostos da Análise Discriminante Linear Múltipla, a Regressão Logística pode ser preferível por sua similaridade com a Regressão Linear Múltipla, tendo em vista que a Regressão Logística apresenta testes estatísticos diretos, possibilidade de incorporação de efeitos não lineares, entre outros (HAIR et al., 1998). Finalmente, a Regressão Logística fornece, como resultado, um valor entre 0 e 1, o que pode ser interpretado como uma probabilidade, o que torna o modelo mais intuitivo. (HOSMER e LEMESHOW, 2000).

3. Metodologia

3.1. Proxy de insolvência

Como forma de determinar quais empresas da amostra seriam consideradas insolventes para fins de classificação da variável resposta na amostra, foram consideradas duas possíveis metodologias. A primeira é a classificação das empresas segundo a Comissão de Valores Mobiliários (CVM). Entretanto, essa opção apresenta uma limitação, que é a existência de poucas empresas em estado de insolvência. Considerando aquelas firmas classificadas como “Em Liquidação Judicial”, “Em Liquidação Extrajudicial”, “Liquidação Extrajudicial”, “Em Recuperação Judicial ou Equivalente”, “Falida” e “Paralisada”, o total de firmas insolventes foi de apenas 19.

Segundo Assaf e Brito (2008), o fato de existir um pequeno número de empresas insolventes na amostra é esperado, tendo em vista de que o número de *defaults* em empresas de grande porte é reduzido. Cabe ressaltar que, embora esta não seja a metodologia aplicada no estudo, segundo Hair et al. (2005), quando um dos grupos utilizados na Análise Discriminante Linear for inferior a 20, a análise pode ficar comprometida.

A segunda opção é a existência de Passivo a Descoberto, isto é, de Patrimônio Líquido Negativo no final do exercício social de 2015, desde que tal condição tenha se repetido por dois ou mais trimestres antecedentes. Embora seja possível que uma empresa consiga manter-se operacional mesmo quando suas exigibilidades superam seus ativos, vários autores tomam a condição de Passivo a Descoberto como indicador de insolvência, tendo em vista que, uma vez vencidas suas obrigações com fontes financiadoras terceiras, os ativos da empresa não seriam suficientes para fazer frente aos compromissos financeiros, o que caracterizaria a insolvência (JANOT, 1999; RUBENS REQUIÃO, 1998; GUIMARÃES e ALVES, 2009).

Decidiu-se pela segunda opção, uma vez que, seguindo esta classificação, o grupo de empresas insolventes passaria a ser composto por 43 empresas.

3.2. Definição da Amostra

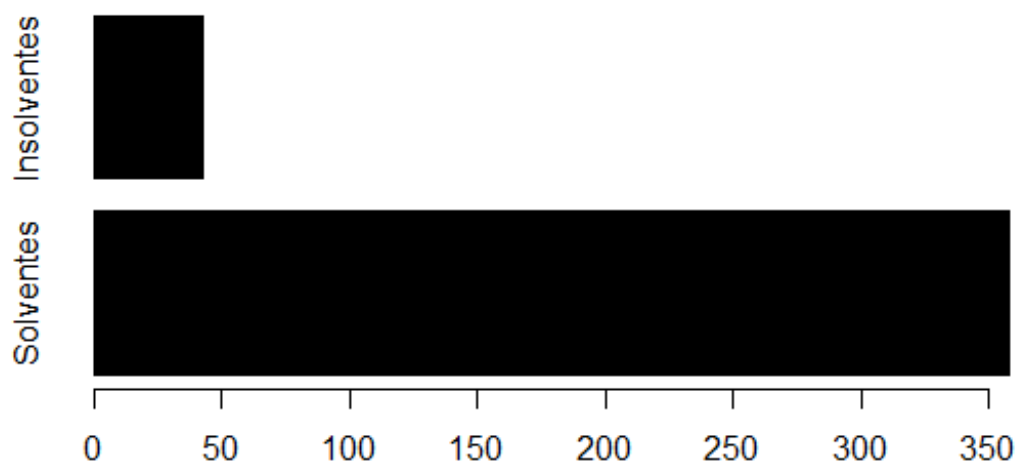
Para realização do estudo, construiu-se uma amostra inicial de 615 empresas, todas retiradas do *software* Economática. Elegeram-se, como ano foco do estudo, o fim do exercício de 2015, isto é, os demonstrativos financeiros consolidados do quarto trimestre do ano de 2015. O ano de 2015 foi escolhido por ser o último com exercício social fechado no momento de elaboração do estudo.

Eliminaram-se da amostra inicial as empresas que, segundo classificação da própria Economática, são instituições financeiras ou prestam algum tipo de serviço financeiro,

como seguradoras. O motivo da eliminação de tais empresas reside no fato de suas demonstrações financeiras serem diferentes das de empresas não financeiras, além do fato de que os próprios indicadores contábeis utilizados na análise de empresas financeiras diferirem daqueles utilizados em empresas não financeiras.

Além desse procedimento de filtragem, decidiu-se eliminar aquelas empresas que não possuíam dados para o horizonte de tempo da análise, seja por não estarem operando ou por, simplesmente, não terem seus dados disponibilizados no sistema da Economática. Portanto, a amostra de estudo passou a ser constituída por um total de 401 companhias. Destas, 43 empresas são classificadas como insolventes, enquanto 358 são classificadas como solventes conforme Gráfico 1.

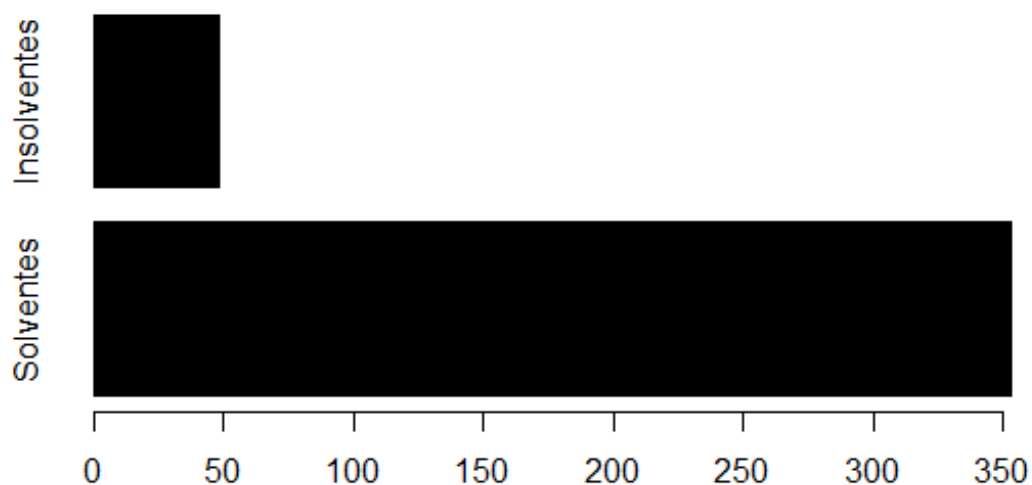
Gráfico 1 – Estado de Inadimplência/Adimplência das 401 empresas da amostra no quarto trimestre de 2015.



Fonte: Elaboração própria com base em dados da Economática.

Já no terceiro trimestre de 2016, o número de empresas insolventes passa a ser 48 empresas enquanto o número de empresas solventes passa a ser de 353, conforme Gráfico 2. Decidiu-se pela utilização deste período por ser o mais recente possível no momento da realização deste estudo e, desta forma, seria possível utilizar dados colhidos em 2014 para testar a previsibilidade do modelo com, aproximadamente, dois anos de antecedência.

Gráfico 2 – Estado de Adimplência/Inadimplência das empresas da amostra no terceiro trimestre de 2016.



Fonte: Elaboração própria com base em dados da Economática.

Existe uma forte assimetria de porte empresarial na amostra, dada, principalmente, pela existência de um pequeno número de empresas com porte muito maior do que das outras.

Existem diversas classificações oficiais de porte, que utilizam diversos critérios para classificação de empresas. Entre elas, a do Serviço de Apoio a Micro e Pequena Empresa (SEBRAE), que, a partir da Lei das Micro e Pequenas Empresas – Lei Complementar 123/06 -, classifica empresas segundo seu faturamento anual, a do Instituto Brasileiro de Geografia e Estatística (IBGE), que classifica empresas de acordo com o número de funcionários, a da Agência Nacional de Vigilância Sanitária (ANVISA), que classifica de acordo com o faturamento e o Banco Nacional de Desenvolvimento Econômico e Social (BNDES), que também classifica de acordo com o faturamento.

As classificações citadas são expostas abaixo.

Quadro 1 – Classificação do SEBRAE para porte empresarial, segundo faturamento anual, Brasil, 2016.

SEBRAE	
Faturamento Anual	Porte
Até R\$ 60 mil	Microempreendedor Individual
Até R\$ 360 mil	Microempresa
Até R\$ 3,6 milhões	Empresa de Pequeno Porte

Fonte: Serviço Brasileiro de Apoio às Micro e Pequenas Empresas - SEBRAE. Elaboração Própria.

Quadro 2 – Classificação da Agência Nacional de Vigilância Sanitária para porte empresarial, segundo faturamento anual, Brasil, 2016.

Vigilância Sanitária	
Faturamento Anual	Porte
Até R\$ 20 milhões	Médio Porte
Acima de R\$ 20 milhões	Grande Porte

Fonte: Agência Nacional de Vigilância Sanitária – ANVISA. Elaboração Própria.

Quadro 3 – Classificação do IBGE para porte empresarial, segundo número de funcionários, Brasil, 2016.

IBGE		
Nº de Empregados	Setor	Porte
entre 100 e 499	Indústria	Médio Porte
entre 50 e 99	Comércio/Serviços	Médio Porte
mais de 500	Indústria	Grande Porte
mais de 100	Comércio/Serviços	Grande Porte

Fonte: Instituto Brasileiro de Geografia e Estatística – IBGE. Elaboração Própria.

Quadro 4 – Classificação do BNDES para porte empresarial, segundo Receita Operacional Anual, Brasil, 2016.

BNDES	
Faturamento Anual	Porte
Menor que R\$ 2,4 milhões	Microempresa
Entre R\$ 2,4 milhões e R\$ 16 milhões	Pequeno Porte
Entre R\$ 16 milhões e R\$ 90 milhões	Médio Porte
Entre R\$ 90 milhões e R\$ 300 milhões	Médio Grande Porte
Maior que R\$ 300 milhões	Grande Porte

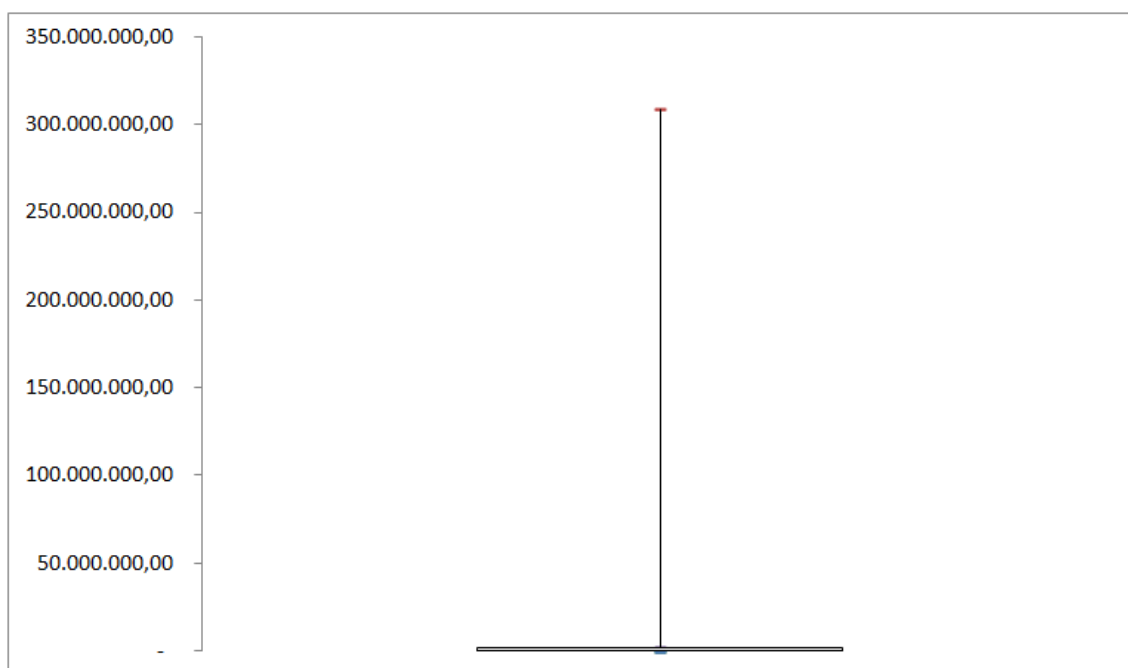
Fonte: Banco Nacional do Desenvolvimento Econômico e Social – BNDES. Elaboração Própria.

Entre as classificações apresentadas, a do BNDES é mais abrangente e segmentada e, além disso, parte de uma instituição com elevada capilaridade no ambiente corporativo brasileiro, o que, em tese, traria maior capacidade crítica ao seu critério de classificação. Portanto, as menções a porte corporativo na presente monografia seguirão essa metodologia.

Dessa forma, para conceber o porte empresarial de determinada empresa, definiu-se a variável Receita Operacional Anual, que representa a quantidade de recursos financeiros obtidos, durante um ano, a partir das principais atividades operacionais de determinada empresa, já desconsiderando descontos concedidos e alguns impostos.

Ilustrando a assimetria descrita acima, com o Gráfico 3 do tipo *boxplot*, pode-se observar como se comporta a distribuição da mediana da variável Receita Operacional, Anual entre o terceiro trimestre de 2014 e o primeiro trimestre de 2015.

Gráfico 3 – *Boxplot* da variável mediana da Receita Operacional Anual entre o terceiro trimestre de 2014 e primeiro trimestre de 2015, considerando apenas empresas não financeiras com dados existentes no sistema Economática.

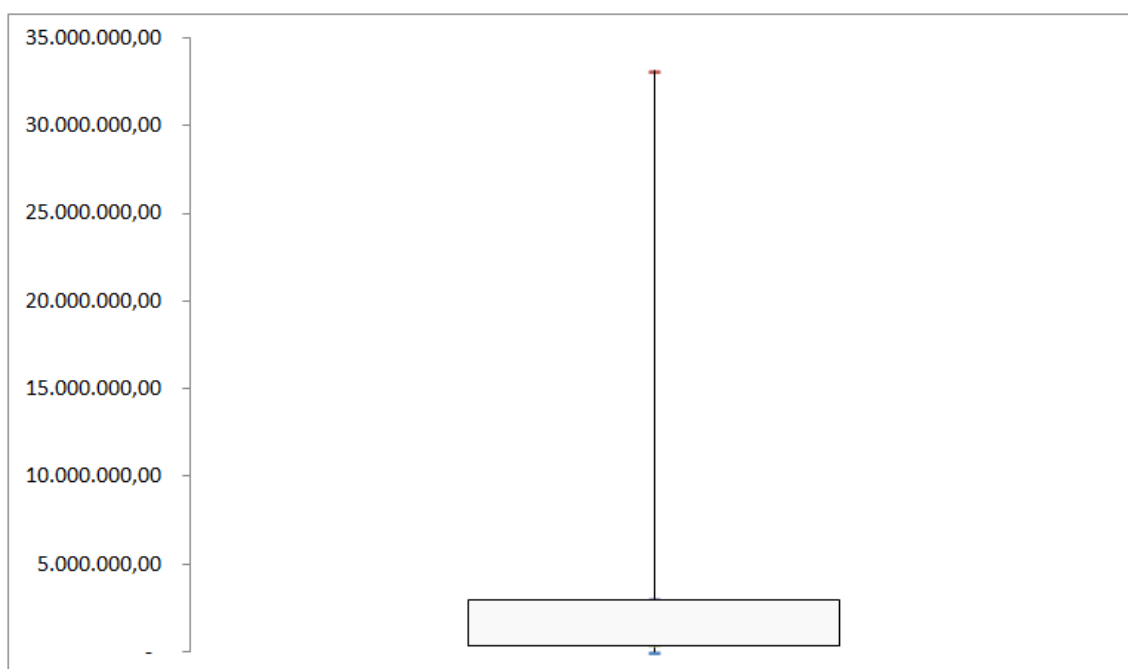


Fonte: Elaboração Própria.

Como já citado anteriormente, a principal razão de haver elevada assimetria na amostra é a presença de um pequeno número de empresas com elevada Receita Operacional Anual. A título de ilustração, ao se somar as medianas da Receita Operacional Anual das dez empresas com maior porte, chegaríamos a um valor correspondente a, aproximadamente, 88% da soma das medianas de todas as outras empresas. Tais dez empresas são, em ordem crescente, Telefônica Brasil, Ambev, Gerdau, Gerdau Met, Braskem, Companhia Brasileira de Distribuição (Pão de Açúcar), Ultrapar, Vale, JBS e Petrobrás.

Eliminando-se tais dez empresas, a amostra resultante passaria a seguir a distribuição descrita no Gráfico 4 abaixo, do tipo *Boxplot*.

Gráfico 4 – *Boxplot* da variável mediana da Receita Operacional Anual entre o terceiro trimestre de 2014 e o primeiro trimestre de 2015, sem as dez empresas de maior porte.



Fonte: Elaboração Própria.

Cogitou-se a possibilidade de eliminação de mais empresas da amostra, como, por exemplo, daquelas de menor porte, uma vez que um dos focos de utilização do modelo, que é a precificação de títulos corporativos, tem maior aplicabilidade em empresas de médio-grande porte, que são aquelas com maior acesso ao mercado de capitais. No entanto, decidiu-se pela manutenção das empresas de menor porte, e para controlar a possível diferença que o porte empresarial pode exercer sobre a saúde financeira, decidiu-se pré-selecionar variáveis *dummy* para controlar o porte.

3.3. Seleção das Variáveis Explicativas

Para selecionar as variáveis explicativas mais relevantes para a construção do modelo de previsão de insolvência, foram utilizados, principalmente, os trabalhos de Sanvicente e Minardi (1998), Pereira e Martins (2015) e Bellovary (2007).

Em 1998, Sanvicenti e Minardi (1998) pré-selecionaram 14 variáveis independentes. Como método de seleção, os autores escolheram cinco variáveis utilizadas nos estudos de Altman et al. (1977) e Sanvicenti e Bader (1996). Estas foram: $(\text{Ativo Circulante} - \text{Passivo Total})/\text{Ativo Total}$, $(\text{Patrimônio Líquido} - \text{Capital Social})/\text{Ativo Total}$, $(\text{Lucro Operacional} - \text{Despesas Financeiras} + \text{Receitas Financeiras})/\text{Ativo Total}$, $\text{Valor Contábil do Patrimônio Líquido}/\text{Valor Contábil do Exigível Total}$ e $\text{Receita Líquida}/\text{Ativo Total}$.

Em seguida, foram escolhidas pelos autores duas variáveis com base em medidas de liquidez referentes ao fenômeno de *overtrading*: (Ativo Circulante Operacional – Passivo Circulante Operacional)/Ativo Total e (Ativo Circulante Financeiro – Passivo Circulante Financeiro)/Ativo Total.

Segundo Assaf Neto e Tibúrcio Silva (1995), o fenômeno do *overtrading*, um dos casos do chamado Efeito Tesoura, é definido pela situação em que a empresa apresenta relativo grau de crescimento e não tem fontes próprias ou de longo prazo para financiar tal crescimento. Com isso, as empresas acabam por tomar empréstimos de curto prazo, comprometendo a situação econômico-financeira da empresa.

Uma das variáveis foi escolhida com base no estudo da agência de classificação de risco Standard & Poor's (1997): Lucro Operacional Antes de Juros e Imposto de Renda /Despesas Financeiras ou, mais facilmente, *Earnings Before Interest and Taxes (EBIT)*/Despesas Financeiras.

Finalmente, foram escolhidas pelos autores seis variáveis de fluxo de caixa, consideradas essenciais em estudos de previsão de insolvência (PEREIRA e MARTINS, 2015).

As variáveis que, segundo Pereira e Martins (2015), apresentaram maior poder preditivo foram os indicadores de liquidez, seguido pelos indicadores de lucros retidos, rentabilidade, endividamento e cobertura de juros.

Por sua vez, Pereira e Martins (2015) e Bellovary (2007) fizeram um levantamento com os indicadores contábeis que mais figuraram em estudos de previsão de insolvência.

Baseados nos três estudos citados nesta seção foram eleitas 29 variáveis contábeis explicativas, exibidas na Tabela 1 abaixo. Cabe ressaltar que as três últimas variáveis são variáveis *dummy* igual a 1 se a empresa pertence ao porte descrito e igual a 0, caso contrário.

Tabela 1 – Seleção preliminar de variáveis explicativas contábeis a serem testadas no modelo

#	Variável	Abreviação
1	Log da Receita Total	LGRT
2	Lucro Líquido/Ativo Total	LL.AT
3	Ativo Circulante/Passivo Circulante	AC.PC
4	Capital de Giro/Ativo Total	CG.AT
5	Lucros Acumulados/Ativo Total	LACUM.AT
6	EBIT/Ativo Total	EBIT.AT
7	(Ativo Circulante - Estoques)/Passivo Circulante	AC.EST.PC
8	Dívida Total/Ativo Total	DT.AT
9	Ativo Circulante/Ativo Total	AC.AT
10	Caixa/Ativo Total	CX.AT

11	Passivo Exigível/Ativo Total	PE.AT
12	Fluxo de Caixa Operacional/Ativo Total	FCO.AT
13	Passivo Circulante/Ativo Total	PC.AT
14	Fluxo de Caixa Operacional/Passivo Exigível	FCO.PE
15	Dívida Total/Patrimônio Líquido	DT.PL
16	Fluxo de Caixa Operacional/Dívida Total	FCO.DT
17	EBIT/Despesas Financeiras	EBIT.DFIN
18	Passivo Exigível/Patrimônio Líquido	PE.PL
19	(Ativo Circulante - Estoques)/Ativo Total	AC.EST.AT
20	Patrimônio Líquido/Ativo Total	PL.AT
21	Empréstimos de Longo Prazo/Ativo Total	EMLP.AT
22	Caixa/Passivo Circulante	CX.PC
23	Fluxo de Caixa Operacional/Passivo Circulante	FCO.PC
24	Capital de Giro/Patrimônio Líquido	CG.PL
25	Log do Ativo Total	LGAT
26	Grande Porte	Grande.Porte
27	Médio-Grande Porte	MedioGrandePorte
28	Médio Porte	MedioPorte

Fonte: Elaboração própria.

Todos os indicadores financeiros foram calculados a partir das medianas dos dados para o primeiro trimestre de 2015, o quarto trimestre de 2014 e o terceiro trimestre de 2014, de forma a não utilizar dados muito próximos ao evento de insolvência (OLHSON, 1980).

Além das variáveis explicativas proveniente de demonstrativos financeiros, decidiu-se pré-selecionar cinco variáveis externas ao contexto contábil. Estas foram: a quantidade de negócios que as ações da empresa tiveram durante o ano de 2015, a quantidade de títulos da empresa circulados durante o ano de 2015, o volume negociado de suas ações durante o ano de 2015, a participação das ações da empresa em relação ao Ibovespa e, finalmente, uma variável *dummy* igual a 1 para empresas com ações listadas na Bovespa e igual a zero, caso contrário.

Tais variáveis são citadas na Tabela 2 abaixo:

Tabela 2 – Seleção preliminar de variáveis explicativas não contábeis a serem testadas no modelo

#	Variáveis Não Contábeis	Abreviação
1	Participação no Ibovespa	ParInd
2	Quantidade de Títulos	Qtits
3	Quantidade de Negócios	Qnegs
4	Volume Negociado	Vol
5	Listagem na Bovespa	Bolsa

Fonte: Elaboração própria.

Além das 34 variáveis explicativas, temos a variável resposta, “Insolvente”, igual a 1 se empresa for insolvente no quarto trimestre de 2015 e igual a 0, caso contrário, e a variável “Insolvente.2016”, igual a 1 se a empresa for insolvente no terceiro trimestre de 2016 e igual a 0, caso contrário.

A partir das variáveis dispostas acima, realizou-se uma análise descritiva, de forma a analisar a distribuição de cada variável na amostra, conforme Tabela 3 abaixo.

Tabela 3 – Análise Descritiva das Variáveis Explicativas

Variável	Média	Mínimo	1º Quartil	Mediana	3º Quartil	Máximo
ParInd	0,14	0,00	0,00	0,00	0,00	7,16
Qtits	86.288,00	0,00	0,00	126,00	21.865,00	7.767.888,00
Qnegs	192.651,00	0,00	0,00	262,00	69.969,00	7.223.334,00
Vol	1.250.949,00	0,00	0,00	1.062,00	275.362,00	96.291.487,00
Bolsa	0,74	0,00	0,00	1,00	1,00	1,00
GrandePorte	0,73	0,00	0,00	1,00	1,00	1,00
MedioGrandePorte	0,13	0,00	0,00	0,00	0,00	1,00
MedioPorte	0,10	0,00	0,00	0,00	0,00	1,00
LGRT	13,45	4,09	12,42	13,59	14,90	19.624,00
LL.AT	0,04	-4,05	-0,02	0,02	0,06	12,16
AC.PC	1,78	0,00	0,83	1,27	1,99	24,72
CG.AT	-0,09	-56,35	-0,03	0,05	0,19	0,84
LACUM.AT	-0,60	-64,66	-0,03	0,00	0,01	0,19
EBIT.AT	0,07	-0,96	0,02	0,06	0,11	4,39
AC.EST.PC	1,51	0,00	0,68	1,06	1,59	23,60
DT.AT	0,58	0,00	0,19	0,35	0,48	46,23
AC.AT	0,34	0,00	0,16	0,34	0,50	0,95
LL.PL	-0,13	-69,67	-0,01	-0,13	0,19	15,89
CX.AT	0,07	0,00	0,01	0,04	0,10	0,77
PE.AT	1,11	0,03	0,50	0,65	0,79	63,70
FCO.AT	0,10	-0,20	0,02	0,06	0,11	12,62
PC.AT	0,42	0,00	0,13	0,22	0,33	56,62
FCO.PE	0,18	-0,25	0,03	0,09	0,21	13,95
DT.PL	1,63	-12,03	0,22	1,63	1,70	113,11
FCO.DT	1,08	-1,26	0,05	1,08	0,36	155,24
EBIT.DFIN	2,27	-2.414,20	0,28	2,27	2,66	2.244,61
PE.PL	6,57	-20,92	0,79	1,58	2,98	958,46
AC.EST.AT	0,27	0,00	0,13	0,25	0,36	0,95
PL.AT	-0,11	-62,70	0,21	0,35	0,50	0,97
EMLP.AT	0,46	0,00	0,08	0,22	0,36	46,08
CX.PC	0,48	0,00	0,06	0,48	0,47	21,75
FCO.PC	0,79	-6,02	0,07	0,27	0,65	118,08
CG.PL	-0,23	-134,21	-0,05	0,18	0,57	6,13

LGAT	14,31	7,69	13,30	14,42	15,56	20,52
------	-------	------	-------	-------	-------	-------

Fonte: Elaboração própria.

Conforme se pode perceber pela análise da Tabela 3, algumas variáveis possuem certa assimetria, tendo em vista o intervalo interquartil entre a mediana e o terceiro quartil e entre este e o valor máximo, principalmente. Este é o caso, por exemplo, das variáveis Qtits, Qnegs, LGRT, PE.PL, AC.PC, entre outras.

Com isso, pode-se perceber que existem algumas empresas que destoam significativamente da maioria, seja positivamente, com índices de liquidez e rentabilidade bem acima do terceiro quartil, ou negativamente, com índices de alavancagem, por exemplo, bem acima do terceiro quartil.

3.4. Modelo Econométrico

A presente seção trata sobre o modelo abordado no presente estudo: a Regressão Logística. Para abordar o tema, serão utilizados os seguintes materiais: *Applied Logistic Regression* (Hosmer e Lemeshow, 2000), *Applied Multivariate Techniques* (Sharma, 1996), *Análise Multivariada de Dados* (Hair et al., 1998) e *An Introduction to Statistical Learning* (JAMES et al., 2000).

3.4.1 Regressão Logística

Para apresentar o modelo de Regressão Logística, assume-se que há uma variável latente y^* (não observável) que representa a propensão de que a variável resposta y ocorra, tal que:

$$y_i^* = \mathbf{x}_i' \boldsymbol{\beta} + \epsilon_i \quad (12)$$

Onde:

\mathbf{x}_i é o vetor das K variáveis explicativas do modelo;

$\boldsymbol{\beta}$ é um vetor de K parâmetros;

ϵ_i é a realização de um erro aleatório identicamente e independentemente distribuído com média zero e distribuição logística padronizada com variância $\pi/3$ (GREENE, 2012: p. 686).

Com isso, as observações de y_i são tais que $y_i=1$, se $y_i^* > 0$ e $y_i=0$, caso contrário, ou seja:

$$y_i = 1(y_i^* > 0) \equiv 1(\mathbf{x}_i' \boldsymbol{\beta} + \epsilon_i > 0) \quad (13)$$

A estimativa de máxima verossimilhança de $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}} = \text{argmax} \ln L(\tilde{\boldsymbol{\beta}})$, é a solução do problema de maximização, com respeito a $\tilde{\boldsymbol{\beta}}$, da função log-verossimilhança:

$$\ln L(\tilde{\beta}) = \sum_{i \in S} \ln \Lambda(\mathbf{x}_i' \tilde{\beta}) + \sum_{i \notin S} \ln(1 - \Lambda(\mathbf{x}_i' \tilde{\beta})) \quad (14)$$

onde $\Lambda(\cdot)$ é a função de distribuição acumulada logística, $i \in S$ denota uma observação pertencente ao conjunto S de sucessos (ou seja, o conjunto em que $y=1$) e $i \notin S$ denota uma observação não pertencente ao conjunto S .

Em suma, o estimador de máxima verossimilhança encontra a estimativa de β que maximiza a verossimilhança de se obter a amostra de dados observada (HOSMER e LEMESHOW, 2000).

Vários aspectos da Regressão Logística são similares à Regressão Linear Múltipla, como por exemplo, a mensuração da precisão das estimativas por meio do desvio-padrão, a utilização de testes estatísticos para inferência, entre outros (JAMES et al., 2000).

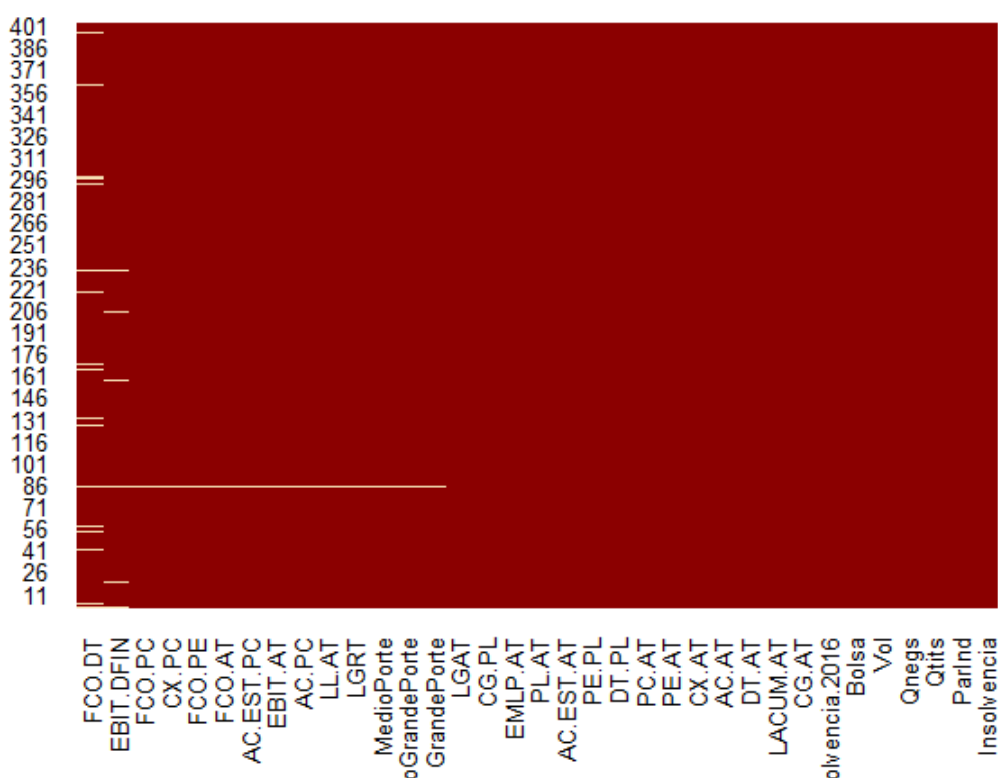
Além da Regressão Logística propriamente dita, serão utilizados no presente estudo o R^2 de McFadden, também conhecido como pseudo R^2 , o teste de Wald, o teste de Hosmer-Lemeshow, o teste de Mann-Whitney-Wilcoxon, o teste de Shapiro-Wilk e, finalmente, o Critério de Informação de Akaike. (HOSMER E LEMESHOW, 2000; SHARMA, 1996).

4. Desenvolvimento do Modelo

O modelo foi estimado com o software R, no qual foram realizados diversos procedimentos.

Em um primeiro momento, por meio do pacote “Amelia”, elaborou-se a Figura 1 abaixo, conhecida com “missmap”. Basicamente, ela ilustra os dados faltantes para cada uma das variáveis.

Figura 1 – Quantidade de dados faltantes por variável



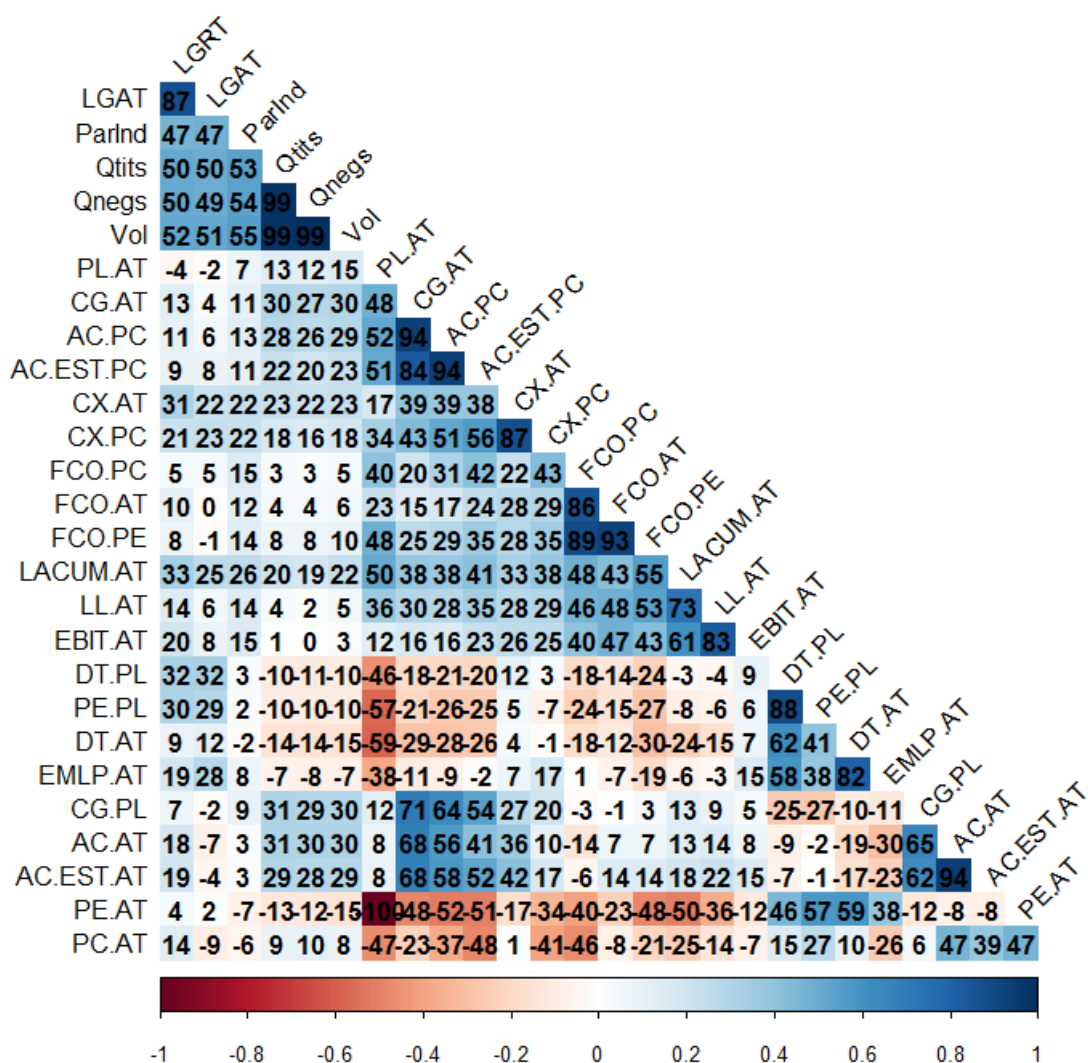
Fonte: Elaboração própria.

A intuição na Figura 1 é bastante simples: em amarelo, estão representadas observações que apresentam dados faltantes para as variáveis descritas no eixo x.

Conforme se pode notar, as variáveis FCO.DT e EBIT.DFIN foram as que apresentaram maior número de dados faltantes. Por esse motivo, tais variáveis foram eliminadas. Para as outras variáveis, adotou-se o procedimento de eliminar as observações com dados faltantes, reduzindo a amostra para 400 empresas.

Calculou-se uma matriz de correlação entre cada uma das variáveis, a partir do coeficiente de correlação de postos de Spearman, que é uma medida não-paramétrica da correlação entre duas variáveis. A visualização de tal matriz é facilitada pela Figura 2.

Figura 2 – Representação da matriz de correlação das variáveis explicativas



Fonte: Elaboração própria.

A Figura 2 é bastante intuitiva: cada quadrado representa a correlação entre duas variáveis, de forma que a tonalidade da cor indica o grau de correlação. Quanto mais vermelho, mais próximo de -1 é o coeficiente de correlação e, quanto mais azul, mais próximo de 1 é o coeficiente de correlação. Em cada um dos quadrados, existe um número que representa, em porcentagem, o valor do coeficiente de correlação de Spearman.

A visualização da correlação entre duas variáveis ajuda evitar problemas de multicolinearidade, isto é, forte correlação entre variáveis explicativas. Neste sentido, é interessante manter, no modelo final, variáveis explicativas que não apresentem uma correlação muito elevada entre si, desde que não exista nenhuma razão teórica para a manutenção de variáveis correlacionadas entre si, que, embora possam não ser, individualmente, significantes, possam ser, conjuntamente, significantes.

Neste momento, baseado em Scalzer et al. (2016), realizou-se o teste não paramétrico de Mann-Whitney-Wilcoxon, a partir de um nível de significância de 5%, sob a hipótese nula deste teste, as funções de distribuição dos dados são iguais para ambos os grupos do estudo, isto é, empresas solventes e insolventes.

Segundo Scalzer et al (2016), o resultado do teste de Mann-Whitney-Wilcoxon permite descrever quais variáveis são, comumente, diferentes entre os grupos de empresas solventes e insolventes. Dessa forma, a aplicação do teste teve a intenção de gerar uma perspectiva sobre quais variáveis, possivelmente, são boas classificadoras. No entanto, embora, para algumas variáveis, não tenha sido possível rejeitar a hipótese nula de igualdade das funções de distribuição, estas não foram removidas da lista de candidatas aos modelos. Os p-valores dos testes de Mann-Whitney-Wilcoxon estão na Tabela 4 abaixo em ordem crescente.

Tabela 4 – Resultados dos testes de Mann-Whitney-Wilcoxon.

Variável	p-valor
PE.AT	2,20E-16
LACUM.AT	2,20E-16
PL.AT	2,20E-16
LL.AT	1,48E-13
CX.PC	8,45E-11
CG.AT	9,34E-11
AC.EST.PC	7,21E-10
AC.PC	2,65E-09
DT.PL	3,05E-08
LGAT	1,40E-07
LGRT	3,17E-07
PE.PL	4,55E-07
CX.AT	4,65E-07
FCO.PE	6,63E-07
GrandePorte	1,52E-06
EBIT.AT	1,64E-06
PC.AT	1,49E-05
FCO.PC	1,83E-05
DT.AT	1,09E-04
MedioGrandePorte	7,68E-03
ParInd	1,52E-02
FCO.AT	3,27E-02
MedioPorte	1,09E-01
Bolsa	1,11E-01
CG.PL	2,63E-01
AC.EST.AT	2,88E-01
LL.PL	3,04E-01

Vol	4,78E-01
AC.AT	5,97E-01
Qtits	6,43E-01
EMLP.AT	7,03E-01
Qnegs	9,12E-01

Fonte: Elaboração própria.

A hipótese nula de igualdade das funções de distribuição não pôde ser rejeitada, a um nível de significância de 5%, apenas para as variáveis situadas abaixo da variável FCO.AT na Tabela 4.

A partir da amostra inicial, dividiu-se, aleatoriamente, a base de dados em duas subamostras: uma utilizada para desenvolver o modelo e outra para testá-lo. A amostra de desenvolvimento, ou de treino, ficou composta por 340 observações, enquanto a de teste, por 60 observações, isto é, 15% da amostra total.

A seleção das variáveis explicativas do modelo final foi efetuada segundo duas possíveis metodologias: manual e automática como a seguir.

4.1. Método Manual

Em um primeiro momento, estimou-se o modelo nulo ou ingênuo, isto é, aquele que apenas inclui o intercepto. Os resultados deste modelo são descritos na Tabela 5 abaixo.

Tabela 5 – Resultados do Modelo Nulo

Coefficientes	Estimativa	Erro-Padrão	Wald p-valor
(Intercepto)	-2,2644	0,1857	<2e-16***
Códigos de Signif.	0,001 '***'	0,01 '**'	0,1 '.'
AIC: 214,13		McFadden:	0

Número de Iterações de Fisher: 5

Fonte: Elaboração própria.

Conforme se pode notar, o intercepto é estatisticamente significativo, a um nível de significância de 5% e, portanto, foi mantido no modelo.

O segundo passo do método manual consistiu na construção de modelos individuais univariados para cada uma das variáveis explicativas, computando os valores de R² de McFadden, do Critério de Informação de Akaike, dos p-valores para o teste de Wald e, finalmente, para o teste de Hosmer-Lemeshow, cuja hipótese nula é a de não existência de diferenças significativas entre os valores preditos e observados, ao longo

dos decis da amostra, o que pode indicar um bom ajuste do modelo aos dados (HOSMER e LEMESHOW, 2000; ASSAF NETO e BRITO, 2008).

A Tabela 6 abaixo ilustra os resultados desta primeira etapa.

Tabela 6 – Resultados da Segunda Etapa

Variável	p-valor de Wald	AIC	R² de McFadden	p-valor de Hosmer-Lemeshow	p-valor de Mann-Whitney-Wilcoxon
PL.AT	3,47E-05	54,36	76,25%	0,00	2,20E-16
PE.AT	3,47E-05	54,36	76,25%	0,00	2,20E-16
LACUM.AT	4,28E-09	130,20	40,50%	0,01	2,20E-16
PC.AT	1,48E-08	140,84	35,49%	0,44	1,49E-05
CG.AT	4,53E-07	151,40	30,51%	0,35	9,34E-11
DT.AT	1,91E-03	190,35	0,12%	0,02	1,09E-04
CX.PC	1,18E-03	193,23	10,79%	0,00	8,45E-11
FCO.PE	1,07E-04	195,54	9,70%	0,00	6,63E-07
LGAT	1,52E-05	196,27	9,36%	0,93	1,40E-07
GrandePorte	3,09E-05	198,59	8,27%	1,00	1,52E-06
LGRT	4,96E-05	199,57	7,81%	0,49	3,17E-07
CX.AT	5,27E-03	203,59	5,91%	0,00	4,65E-07
Vol	8,61E-02	206,44	4,57%	0,00	4,78E-01
ParInd	9,88E-01	207,47	4,09%	1,00	1,52E-02
LL.AT	2,66E-03	207,60	4,02%	0,00	1,48E-13
Qnegs	7,04E-02	207,64	4,00%	0,00	9,12E-01
Qtits	9,40E-02	207,73	3,96%	0,00	6,43E-01
MedioGrandePorte	2,12E-03	207,86	3,90%	1,00	7,68E-03
EMLP.AT	1,03E-01	209,09	3,31%	0,00	7,03E-01
PE.PL	2,41E-01	212,03	0,02%	0,00	4,55E-07
FCO.AT	6,48E-02	212,46	1,73%	0,96	3,27E-01
Bolsa	9,00E-02	212,64	1,65%	1,00	1,11E-01
FCO.PC	1,56E-01	212,65	1,64%	0,00	1,83E-05

MedioPorte	5,42E-02	212,92	1,51%	NA	1,09E-01
DT.PL	1,18E-01	213,37	1,30%	0,00	3,05E-08
CG.PL	1,38E-01	213,78	1,11%	0,00	2,63E-01
AC.PC	2,57E-01	214,14	0,94%	0,00	2,65E-09
AC.AT	6,51E-01	215,93	0,10%	0,87	5,97E-01
EBIT.AT	7,45E-01	216,04	4,39%	0,00	1,64E-06
AC.EST.AT	8,23E-01	216,08	0,02%	0,16	2,88E-01
AC.EST.PC	2,05E-01	239,81	1,10%	0,00	7,21E-10

Fonte: Elaboração própria.

Conforme pode ser observado na Tabela 6, as variáveis PE.AT e PL.AT foram as que apresentaram menor AIC e maior R² de McFadden, seguidas pela variável LACUM.AT, sendo ambas estatisticamente significantes a um nível de significância de 5%. Além disso, para estas variáveis, foi possível rejeitar, a um nível de significância de 5%, a hipótese nula do teste de Hosmer-Lemeshow e a hipótese nula do teste de Mann-Whitney-Wilcoxon.

A variável PE.AT e a variável PL.AT, somadas, igualam a 1, uma vez que PE + PL = AT. Por este motivo, arbitrariamente, escolheu-se a variável PE.AT.

Os resultados do modelo com esta variável estão na Tabela 7 abaixo.

Tabela 7 – Resultados do Modelo com PE.AT

Coefficientes	Estimativa	Erro-Padrão	Wald p-valor
(Intercepto)	-20,993	4,739	9,43E-06***
PE.AT	21,422	5,174	3,47E-05***
Códigos de Signif.	0,001 '***'	0,01 '**'	0,1 '!
AIC: 54,36		McFadden:	76,25%

Número de Iterações de Fisher: 12

Fonte: Elaboração própria.

Pode-se perceber que a magnitude do coeficiente estimado para a variável PE.AT foi bastante relevante. Isto pode ser fruto da assimetria existente nesta variável, conforme pode ser percebido pela análise da Tabela 3. Existe a possibilidade que as observações contidas no intervalo interquartil entre o terceiro quartil desta variável e seu valor de máximo sejam, em maioria, empresas classificadas como insolventes, influenciado o valor do coeficiente estimado.

A terceira etapa é testar a inclusão de uma segunda variável explicativa ao modelo. A variável PL.AT não foi considerada, tendo em vista sua correlação perfeitamente negativa com a variável PE.AT.

Para todas as outras variáveis, foram calculadas as mesmas medidas que na segunda etapa, dispostas na Tabela 8. Desta vez, o coeficiente de correlação entre a variável PE.AT e cada uma das variáveis também foi exibido, com o intuito de se avaliar a possibilidade de existência de multicolinearidade, o que não foi o caso.

Os resultados foram demonstrados na Tabela 8 abaixo.

Tabela 8 – Resultados da Terceira Etapa

Variável	p-valor de Wald	AIC	R ² de McFadden	p-valor de Hosmer-Lemeshow	p-valor de Mann-Whitney-Wilcoxon	Correlação com PE.AT
EBIT.AT	1,72E-02	47,08	80,63%	0,01	1,64E-06	-12%
LACUM.AT	7,48E-03	50,30	79,11%	0,25	2,20E-16	-50%
PC.AT	6,55E-02	52,87	77,90%	0,23	1,49E-05	47%
FCO.PE	1,36E-01	54,02	77,36%	1,00	6,63E-07	-48%
MedioGrandePorte	1,37E-01	54,21	77,27%	0,01	7,68E-03	-
Bolsa	2,21E-01	54,44	77,16%	0,34	1,11E-01	-
FCO.AT	2,53E-01	54,97	76,91%	0,41	3,27E-01	-23%
ParInd	9,96E-01	55,00	76,90%	0,38	1,52E-02	-7%
EMLP.AT	3,15E-01	55,29	76,76%	0,00	7,03E-01	38%
PE.PL	7,40E-01	55,30	76,76%	0,00	4,55E-07	57%
AC.AT	3,06E-01	55,34	76,74%	0,00	5,97E-01	-8%
FCO.PC	2,51E-01	55,56	76,63%	0,39	1,83E-05	-40%
CG.PL	4,99E-01	55,56	76,63%	0,01	2,63E-01	-12%
MedioPorte	8,71E-01	55,81	76,52%	0,35	1,09E-01	-
AC.EST.AT	4,44E-01	55,82	76,51%	0,00	2,88E-01	-8%
Qtits	5,57E-01	55,86	76,50%	0,61	6,43E-01	-13%
CG.AT	4,86E-01	55,88	76,49%	0,60	9,34E-11	-48%
CX.PC	5,55E-01	55,92	76,46%	0,65	8,45E-11	-34%

Qnegs	5,80E-01	55,98	76,43%	0,58	9,12E-01	-12%
Vol	5,92E-01	55,99	76,43%	0,35	4,78E-01	-15%
AC.EST.PC	6,27E-01	56,04	76,41%	0,63	7,21E-10	-51%
LGAT	5,80E-01	56,06	76,40%	0,63	1,40E-07	2%
GrandePorte	6,05E-01	56,10	76,38%	0,00	1,52E-06	
AC.PC	7,61E-01	56,26	76,31%	0,61	2,65E-09	-52%
LGRT	7,77E-01	56,29	76,30%	0,39	3,17E-07	4%
DT.PL	7,86E-01	56,29	76,29%	0,00	3,05E-08	46%
DT.AT	7,91E-01	56,29	76,29%	0,00	1,09E-04	59%
CX.AT	9,48E-01	56,36	76,26%	0,37	4,65E-07	-17%
LL.AT	2,00E-16	2168,60	-919,46%	0,00	1,48E-13	-36%

Fonte: Elaboração própria.

Conforme pode ser percebido na Tabela 8 acima, as variáveis com melhor desempenho foram EBIT.AT e LACUM.AT. Em termos de qualidade de ajuste, a variável EBIT.AT demonstrou-se melhor, com um AIC menor e um R² de McFadden maior em relação ao modelo com LACUM.AT. Além disso, a variável EBIT.AT apresentou um coeficiente de correlação com a variável PE.AT inferior ao da variável LACUM.AT. Por outro lado, para a variável LACUM.AT, pôde-se rejeitar, a um nível de significância inferior, as hipóteses nulas dos testes de Wald e de Mann-Whitney-Wilcoxon. Para ambas as variáveis, não foi possível rejeitar, a um nível de significância de 5%, a hipótese nula do teste de Hosmer-Lemeshow, o que pode indicar que ambos os modelos são bons.

De qualquer forma, dadas às proximidades dos valores de AIC e de R² de McFadden para ambos os modelos, decidiu-se por exibir, nas Tabelas 9 e 10 abaixo, os resultados encontrados para os dois modelos.

Tabela 9 – Resultados do Modelo com PE.AT e LACUM.AT

Coefficientes	Estimativa	Erro-Padrão	Wald p-valor
(Intercepto)	-22,6991	5,7039	6,9E-05***
PE.AT	22,8275	6,151	0,000206***
LACUM.AT	-1,2805	0,4787	0,007477**
Códigos de Signif.	0,001 '***'	0,01 '**'	0,1 '.'
AIC: 50,30		McFadden:	79,11%

Número de Iterações de Fisher: 12

Fonte: Elaboração própria.

Tabela 10 – Resultados do Modelo com PE.AT e EBIT.AT

Coefficientes	Estimativa	Erro-Padrão	Wald p-valor
(Intercepto)	-27,4775	7,2586	0,000153***
PE.AT	28,2613	7,8319	0,000308***
EBIT.AT	1,9361	0,8128	0,017217.

Códigos de Signif.	0,001 '***'	0,01 '**'	0,1 '.'
--------------------	-------------	-----------	---------

AIC: 47,08	McFadden:	80,63%
------------	-----------	--------

Número de Iterações de Fisher: 15

Fonte: Elaboração própria.

O valor estimado para o coeficiente de EBIT.AT apresentou um sinal contrário ao que se esperava, inicialmente. Tendo em vista que a variável é um indicador de rentabilidade da empresa, esperava-se que a mesma contribuísse para uma menor probabilidade de insolvência. Por outro lado, o valor estimado para o coeficiente da variável LACUM.AT apresentou um sinal condizente com o esperado. Os Lucros Acumulados de uma empresa são aqueles restantes após a distribuição de dividendos e possíveis investimentos realizados. Portanto, volumes elevados de Lucros Acumulados somente se tornam possível quando a empresa apresenta resultados positivos.

Dados os resultados das medidas de AIC e R^2 de McFadden e de ter sido possível encontrar que ambas as variáveis são estatisticamente significantes, para decidir entre os dois modelos, averiguou-se a taxa de acerto das previsões do modelo frente aos valores efetivamente observados para a amostra de teste, para a base como um todo e para a situação de solvência das empresas no terceiro trimestre de 2016. Para tanto, utilizou-se um ponto de corte de 0,35: aquelas empresas para as quais o valor predito pelo modelo da variável resposta for acima de 0,35 são classificadas como insolventes e, caso contrário, classificadas como solventes. O motivo de ter se escolhido o ponto de corte de 0,35 foi baseado no fato de que o custo de se classificar, erroneamente, uma empresa insolvente como solvente ser maior do que o custo de se classificar, erroneamente, uma empresa solvente como insolvente, tendo em vista que, ao se cometer o primeiro erro descrito acima, decisões de investimento podem resultar em perdas financeiras.

A representação dos resultados para este tipo de método é organizada em uma matriz conhecida como *confusion matrix*, depostas nos Quadros de 5 a 10 abaixo.

Quadro 5 – Confusion Matrix para o Modelo com a Variável EBIT.AT Aplicado à Subamostra de Teste

		Observado	
		0	1
Previsto	0	49	1
	1	0	10

Fonte: Elaboração própria.

O índice de acerto para as empresas insolventes foi de 90,90% e, para as empresas, solventes, de 100%. Globalmente, o índice de acerto foi de 98,33%.

Quadro 6 – Confusion Matrix para o Modelo com a Variável EBIT.AT Aplicado a Base de Dados

		Observado	
		0	1
Previsto	0	354	6
	1	3	37

Fonte: Elaboração própria.

O índice de acerto para as empresas insolventes foi de 86,05% enquanto que, para as empresas solventes, o índice foi de 99,16%. Globalmente, o índice de acerto foi de 97,75%.

Quadro 7 – Confusion Matrix para o Modelo com a Variável EBIT.AT Aplicado ao Estado de Solvência no Terceiro Trimestre de 2016

		Observado	
		0	1
Previsto	0	345	13
	1	7	35

Fonte: Elaboração própria.

O índice de acerto para as empresas insolventes foi de 72,92% e, para as empresas solventes, de 98,01%. No geral, o acerto foi de 95%.

A apuração dos resultados pareceu indicar que, mesmo utilizando um ponto de corte de 0,35, é mais fácil classificar, corretamente, empresas solventes do que empresas insolventes.

Quadro 8 – Confusion Matrix para o Modelo com a Variável LACUM.AT Aplicado à Subamostra de Teste

		Observado	
		0	1
Previsto	0	49	1

	1	0	10
--	---	---	----

Fonte: Elaboração própria.

Pode-se observar, pelo Quadro 8, que o desempenho do modelo com a variável LACUM.AT na classificação das empresas da subamostra de teste foi o mesmo que o do modelo com a variável EBIT.AT.

Quadro 9 – *Confusion Matrix* para o Modelo com a Variável LACUM.AT Aplicado a Base de Dados

		Observado	
		0	1
Previsto	0	354	6
	1	3	37

Fonte: Elaboração própria.

Assim como o que ocorreu com a subamostra de teste, o desempenho do modelo com a variável LACUM.AT aplicado na classificação das empresas da base como um todo foi o mesmo que com o modelo com a variável EBIT.AT.

Finalmente, temos a última *confusion matrix*.

Quadro 10 – *Confusion Matrix* para o Modelo com a Variável LACUM.AT Aplicado ao Estado de Solvência no Terceiro Trimestre de 2016

		Observado	
		0	1
Previsto	0	348	12
	1	4	36

Fonte: Elaboração própria.

Dessa vez, o desempenho do modelo com a variável LACUM.AT foi superior: o índice de acerto para as empresas insolventes foi de 75% enquanto que, para as empresas solventes, foi de 98,86%. No geral, o índice de acerto do modelo foi de 96%. Mais uma vez, foi possível observar que, mesmo a um nível de corte de 0,35, a correta classificação das empresas solventes parece mais fácil.

Dado o exposto acima, decidiu-se aplicar a quarta etapa ao modelo com a variável explicativa LACUM.AT.

Finalmente, na quarta etapa, testou-se a inclusão de mais uma variável explicativa ao modelo. Os resultados estão apresentados nas Tabelas 11 a seguir.

Tabela 11 – Resultados da Quarta Etapa

Variável	p-valor de Wald	AIC	R ² de McFadden	p-valor de Hosmer-Lemeshow	p-valor de Mann-Whitney-Wilcoxon
LL.AT	1,32E-01	48,01	81,14%	0,00	1,48E-13
EBIT.AT	1,31E-01	48,21	81,04%	0,00	1,64E-06
PC.AT	1,08E-01	49,66	80,36%	0,02	1,49E-05
MedioGrandePorte	1,22E-01	49,93	80,23%	0,39	7,68E-03
AC.AT	1,78E-01	50,48	79,97%	0,08	5,97E-01
PE.PL	6,98E-01	50,97	79,74%	0,32	4,55E-07
AC.EST.AT	2,39E-01	51,05	79,71%	0,07	2,88E-01
ParInd	9,96E-01	51,39	79,55%	0,98	1,52E-02
Bolsa	3,92E-01	51,45	79,52%	0,18	1,11E-01
CG.PL	4,93E-01	51,55	79,47%	0,35	2,63E-01
MedioPorte	9,18E-01	51,79	79,36%	0,21	1,09E-01
LGRT	6,03E-01	52,02	79,25%	0,12	3,17E-07
CX.AT	6,02E-01	52,04	79,24%	0,15	4,65E-07
Qtits	6,91E-01	52,10	79,21%	0,29	6,43E-01
FCO.PE	6,82E-01	52,13	79,20%	0,99	6,63E-07
FCO.PC	6,93E-01	52,16	79,19%	0,29	1,83E-05
Qnegs	7,45E-01	52,18	79,17%	0,26	9,12E-01
LGAT	7,31E-01	52,19	79,17%	0,14	1,40E-07
Vol	7,55E-01	52,19	79,17%	0,23	4,78E-01
EMLP.AT	7,39E-01	52,19	79,17%	0,31	7,03E-01
DT.AT	7,67E-01	52,22	79,16%	0,94	1,09E-04
AC.PC	8,14E-01	52,26	79,14%	0,23	2,65E-09
DT.PL	8,93E-01	52,29	79,12%	0,23	3,05E-08
CX.PC	8,99E-01	52,29	79,12%	0,27	8,45E-11
FCO.AT	9,04E-01	52,29	79,12%	0,25	3,27E-01
CG.AT	0,95E-01	52,30	79,12%	0,25	9,34E-11

GrandePorte	9,76E-01	52,30	79,12%	0,25	1,52E-06
AC.EST.PC	9,94E-01	52,30	79,12%	0,25	7,21E-10

Fonte: Elaboração própria.

Conforme se pode observar na Tabela 11, não foi possível, a um nível de significância de 5%, rejeitar a hipótese nula de insignificância para nenhuma das variáveis. Além disso, a inclusão de uma variável adicional pouco melhora as medidas de AIC e R² de McFadden. Tal resultado parece-nos indicar que, uma vez controlados os efeitos de PE.AT e LACUM.AT, a inclusão de mais um indicador financeiro ou mesmo de uma variável não contábil não contribui para melhorar o ajuste do modelo aos dados de maneira significativa.

Entretanto, se optarmos pela inclusão de uma variável explicativa mesmo assim, as variáveis que, efetivamente, reduzem o AIC e aumentam o R² de McFadden, além de terem os menores p-valores para o teste de Wald, são as variáveis LL.AT, EBIT.AT, PC.AT e MedioGrandePorte. No entanto, para as duas primeiras, existe uma forte correlação (acima de 0,7) com a variável LACUM.AT.

Dessa forma, nas Tabelas 12 e 13 abaixo, estão demonstrados os resultados para os modelos que acrescentam, respectivamente, as variáveis PC.AT e MedioGrandePorte. Além disso, na Tabela 13, estimaram-se também as variáveis de interação PE.AT*MedioGrandePorte e LACUM.AT*MedioGrandePorte, para se captar o efeito que as variáveis PE.AT e LACUM.AT geram para empresas de Médio-Grande Porte.

Tabela 12 – Resultados do Modelo com PE.AT, LACUM.AT e PC.AT

Coefficientes	Estimativa	Erro-Padrão	Wald p-valor
(Intercepto)	-24,5577	6,3399	0,000107***
PE.AT	23,1752	6,5336	0,000390***
LACUM.AT	-1,1793	0,4757	0,013176.
PC.AT	3,9184	2,441	0,108441
Códigos de Signif.	0,001 '***'	0,01 '**'	0,1 '!
AIC: 49,66		McFadden:	80,36%

Número de Iterações de Fisher: 11

Fonte: Elaboração própria.

Tabela 13 – Resultados do Modelo com PE.AT, LACUM.AT, MedioGrandePorte, PE.AT*MedioGrandePorte e LACUM.AT*MedioGrandePorte

Coefficientes	Estimativa	Erro-Padrão	Wald p-valor
(Intercepto)	-18,8579	5,0604	0,000194***

PE.AT	18,3316	5,5283	0,000913***
LACUM.AT	-1,1888	0,4615	0,009991**
MedioGrandePorte	-301,7378	43.209,7133	0,994428
PE.AT*MedioGrandePorte	314,5209	42.426,4468	0,994476
LACUM.AT*MedioGrandePorte	-96,9357	17.089,1619	0,995474
Códigos de Signif.	0,001 '***'	0,01 '**'	0,1 ''
AIC: 50,34		McFadden:	81,92%

Número de Iterações de Fisher: 23

Fonte: Elaboração própria.

Pode-se notar que nenhuma das variáveis acrescentadas ao modelo com PE.AT e LACUM.AT é, individualmente, estatisticamente significativa, a um nível de significância de 5%. As variáveis acrescentadas no modelo da Tabela 13 apresentaram erros-padrões estimados elevados, resultando nos altos p-valores encontrados para elas.

Além disso, o AIC de nenhum dos modelos das Tabelas 12 e 13 mostrou mudanças muito significativas em relação ao modelo com PE.AT e LACUM.AT.

Embora não sejam estatisticamente significantes, os sinais dos coeficientes estimados de todas as variáveis condizem com o esperado. Empresas com um volume muito alto de passivo circulante tendem a ter problemas de liquidez, por não conseguirem casar os fluxos de ativos e passivos de curto prazo. Além disso, uma vez que empresas de maior porte apresentam, de maneira geral, maior acesso ao mercado de capitais e a melhores práticas de gestão, além de outras vantagens, é de se esperar que o fato da empresa ser de Médio-Grande Porte reduza a probabilidade de a mesma ser insolvente.

Em relação aos termos de interação, esperava-se que, para uma empresa de porte maior, um acréscimo aos valores de PE.AT e LACUM.AT contribuíssem para aumentar e diminuir a probabilidade de insolvência, respectivamente, assim como ocorre no caso geral, sem a captação desse tipo de efeito de interação. Os coeficientes estimados confirmaram essa expectativa.

Foi realizado um teste de Wald para testar a significância conjunta da variável MedioGrandePorte e das variáveis de interação PE.AT*MedioGrandePorte e LACUM.AT*MedioGrandePorte. O p-valor deste teste foi bastante próximo a 1, assim como a significância individual de cada uma das variáveis. Portanto, não foi possível rejeitar a hipótese nula de não significância.

Dessa forma, realizou-se um novo teste de Wald para testar a significância conjunta das variáveis de interação PE.AT*MedioGrandePorte e LACUM.AT*MedioGrandePorte. Mais uma vez, observou-se um p-valor bastante próximo de 1, levando à não rejeição da hipótese nula de não significância.

Uma vez constatados os resultados dos dois testes de significância conjunta descritos acima, optou-se pela não inclusão das variáveis MedioGrandePorte e das variáveis de interação PE.AT*MedioGrandePorte e LACUM.AT*MedioGrandePorte.

4.2. Método Automático

O método selecionado para o procedimento automático é conhecido como *Backward Stepwise* e consiste em um algoritmo que elimina, a cada passo do processo, uma das variáveis explicativas do Modelo Completo, isto é, o Modelo com todas as variáveis explicativas, até que se alcance determinado critério. No caso específico, o *output* do algoritmo é o modelo com menor AIC.

No entanto, o algoritmo não leva em consideração, por exemplo, variáveis explicativas que sejam muito correlacionadas entre si, o que poderia resultar em um modelo final com erros-padrões inflados e com problemas nas estimativas. Portanto, para o modelo completo, do qual o algoritmo começa o procedimento de eliminação, foram escolhidas variáveis em ordem decrescente da lista apresentada na Tabela 6, eliminando-se aquelas variáveis explicativas com forte correlação (acima de 0,7) com alguma das variáveis explicativas já escolhidas.

Com isso, o modelo completo estimado foi o descrito na Tabela 14.

Tabela 14 – Resultados do Modelo Completo

Coeficientes	Estimativa	Erro-Padrão	Wald p-valor
(Intercepto)	-2,76E+01	3,13E+01	0,37953
PE.AT	1,99E+01	6,35E+00	0,00172**
LACUM.AT	-1,71E+00	9,89E-01	0,08352.
PC.AT	2,71E+00	3,35E+00	0,418130
CG.AT	2,91E-01	3,33E+00	0,930570
DT.AT	1,42E+00	2,42E+00	0,558070
CX.PC	6,11E-01	1,59E+00	0,701050
FCO.PE	-2,42E+00	7,88E+00	0,759100
LGAT	-4,03E-01	6,28E-01	0,520790
Vol	5,55E-07	6,93E-07	0,423400
ParInd	-6,76E+01	1,32E+04	0,995900
GrandePorte	1,11E+01	2,93E+01	0,703810
MedioGrandePorte	1,14E+01	2,92E+01	0,695900
PE.PL	2,27E-02	4,59E-02	0,620560
Códigos de Signif.	0,001 '****'	0,01 '***'	0,1 ' !'
AIC: 61,576		McFadden:	84,17%

Número de Iterações de Fisher: 23

Fonte: Elaboração própria.

Conforme pode ser observado na Tabela 14, a maioria das variáveis explicativas não é estatisticamente significativa. Além disso, a inclusão do grande número de variáveis contribuiu para elevadas estimativas para os erros-padrão das variáveis. Além disso, dado o elevado número de variáveis, foram necessárias 23 iterações de Fisher para estimar o modelo.

Após realizado o procedimento do *Backward Stepwise*, o modelo encontrado foi o seguinte:

Tabela 15 – Resultados do Modelo *Backward Stepwise*

Coeficientes	Estimativa	Erro-Padrão	Wald p-valor
(Intercepto)	-3,19E+01	3,07E+01	0,297654
PE.AT	2,17E+01	6,15E+00	0,000424***
LACUM.AT	-1,91E+00	6,31E-01	0,002458**
GrandePorte	1,00E+01	2,96E+01	0,735150
MedioGrandePorte	1,15E+01	2,97E+01	0,697485
Códigos de Signif.	0,001 '****'	0,01 '***'	0,1 ' !'
AIC: 47,22		McFadden:	82,45%

Número de Iterações de Fisher: 12

Fonte: Elaboração própria.

Diferentemente do ocorrido com o modelo descrito na Tabela 13, as variáveis de porte apresentaram, embora estatisticamente insignificantes a um nível de significância de 5%, sinais contrários aos esperados para as estimativas dos coeficientes.

Além disso, assim como já constatado antes, uma vez controlados os efeitos de LACUM.AT e PE.AT, a inclusão de novas variáveis não acrescentou em muito à qualidade de ajuste do modelo.

Caso queiramos eliminar as variáveis de porte, dado os sinais das estimativas encontradas, podemos realizar o procedimento de *Backward Stepwise* sem incluir as variáveis de porte no modelo completo e, com isso, o resultado seria o da Tabela 16.

Tabela 16 – Resultados do Modelo *Backward Stepwise* sem considerar as variáveis de porte

Coeficientes	Estimativa	Erro-Padrão	Wald p-valor
(Intercepto)	-24,5577	6,3399	0,000107***
PE.AT	23,1752	6,5336	0,000390***
LACUM.AT	-1,1793	0,4757	0,013176.

PC.AT	3,9184	2,441	0,108441
Códigos de Signif.	0,001 '****'	0,01 '***'	0,1 '!''
AIC: 49,66		McFadden:	80,36%

Número de Iterações de Fisher: 11

Fonte: Elaboração própria.

O modelo estimado foi, exatamente, o mesmo encontrado pelo método manual, escrito na Tabela 12.

5. Resultados

Após a estimação dos modelos, os mesmos foram utilizados para construir curvas ROC (Receiver Operating Characteristic), para que se fosse possível representar, graficamente, o desempenho preditivo dos modelos enquanto classificadores binários.

Construiu-se a curva ROC para os dois modelos encontrados ao final do método manual e para os dois modelos encontrados ao final do método automático. Como existe um modelo em comum, foram construídas três curvas ROC, exibidas a seguir.

Figura 3 – Curva ROC para o Modelo com PE.AT, LACUM.AT e PC.AT

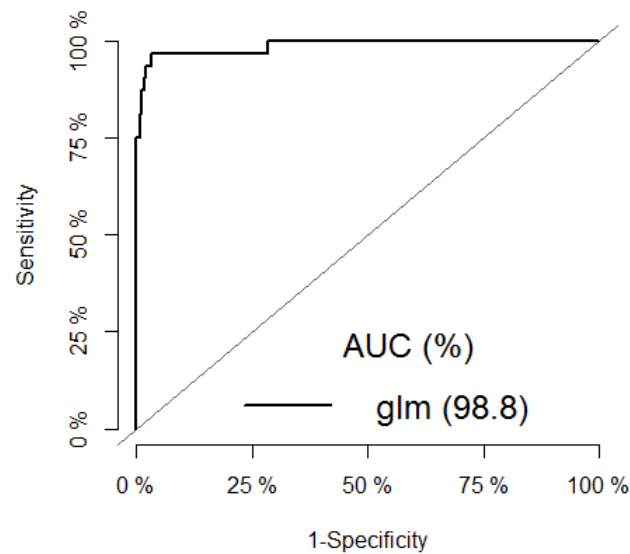
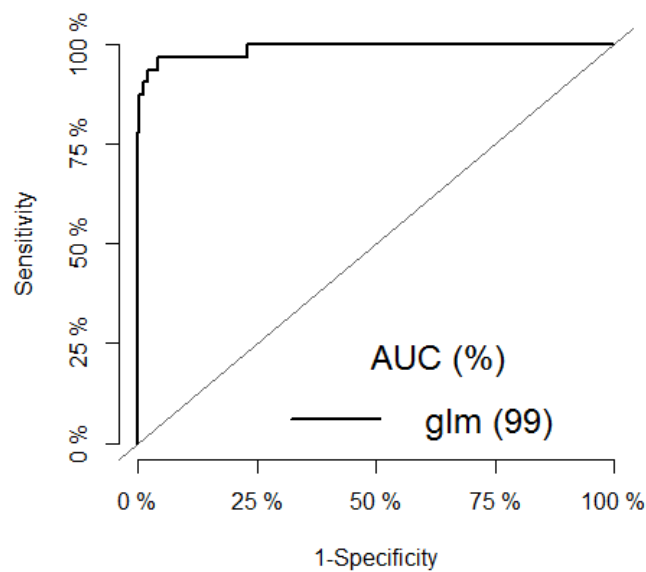


Figura 4 – Curva ROC para o Modelo Manual com PE.AT, LACUM.AT, MedioGrandePorte e GrandePorte



Conforme pode ser observado pela análise das curvas ROC, ambos os modelos apresentaram um elevado desempenho preditivo. Comparativamente a linha diagonal, que representa uma classificação binária aleatória, os modelos apresentaram bastantes melhoras. Além disso, as medições das áreas embaixo das curvas (AUC) indicam que há uma elevada precisão no sentido de separar o grupo das empresas solventes e insolventes.

Abaixo, está representada, no Quadro 11, a taxa de acerto dos modelos para a amostra de teste, para a base como um todo e para os dados referentes ao terceiro trimestre de 2016.

Quadro 11 – Taxas de Acerto dos Modelos (Ponto de Corte = 0,35)

Modelo	Teste	Base	2016
PE.AT+LACUM.AT+PC.AT	98,33%	97,75%	96,00%
PE.AT+LACUM.AT+MedioGrande Porte +GrandePorte	95,00%	97,75%	96,00%

Fonte: Elaboração própria.

Conforme pode-se observar, não houve muita diferença de um modelo para o outro. Os dois tiveram alto índice de acerto, com um desempenho um pouco pior para o modelo com as duas variáveis de porte e um desempenho um pouco melhor para o modelo PE.AT, LACUM.AT e PC.AT.

Para o Quadro 12, calculou-se a taxa de acerto apenas das empresas insolventes, isto é, o percentual de empresas insolventes corretamente previstas como tal. Manteve-se o ponto de corte de 0,35 com o intuito de se minimizar o erro de se classificar, incorretamente, uma empresa insolvente como solvente.

Quadro 12 – Taxas de Acerto dos Modelos para as Empresas Insolventes (Ponto de Corte = 0,35)

Modelo	Teste	Base	2016
PE.AT+LACUM.AT+PC.AT	90,91%	88,37%	77,08%
PE.AT+LACUM.AT+MedioGrande Porte +GrandePorte	72,73%	83,72%	72,92%

Fonte: Elaboração própria.

Mais uma vez, o modelo com melhor performance foi o modelo com PE.AT, LACUM.AT e PC.AT, que apresentou melhores taxas de acerto. Por sua vez, o modelo com as variáveis de porte apresentaram taxas de acerto piores, especialmente no que tange a classificação da amostra de teste.

6. Conclusões

Tendo analisado todos os modelos apresentados nesta monografia, algumas conclusões específicas podem ser levantadas.

No que tange as variáveis explicativas, a variável PE.AT ou a variável PL.AT são as que se mostraram com maior potencial de ajuste do modelo aos dados, sendo ambas estatisticamente significativas, a um nível de significância de 5%. Tendo em vista que estamos modelando, como variável resposta, o evento de se ter um Patrimônio Líquido negativo, ambas as variáveis refletem, direta ou indiretamente, o valor da variável resposta com uma defasagem em relação ao quarto trimestre de 2015. O fato das variáveis PE.AT e PL.AT, individualmente, apresentarem um bom ajuste do modelo aos dados pode indicar que existe uma tendência de empresas manterem seu estado de adimplência no horizonte de um ano.

Em segundo lugar, as variáveis de rentabilidade também se mostraram estatisticamente significantes, o que condiz com o esperado. Dentre as variáveis dessa categoria, a variável LACUM.AT, frequentemente utilizada em modelos de previsão de insolvência, indica o quanto dos lucros gerados não foi repassado aos acionistas em forma de dividendos e a variável EBIT.AT indica a geração de lucro operacional, isto é, aquele, efetivamente, gerado pela atividade produtiva e que desconsidera, portanto, as despesas fiscais e financeiras.

Em relação ao desempenho dos modelos, pareceu haver boa capacidade preditiva, especialmente ao que tange as empresas solventes. Tal fato parece indicar que é mais difícil prever a insolvência do que a solvência, embora tal condição possa ser específica à amostra utilizada e não à população como um todo.

Especificamente, o modelo que apresentou melhor desempenho preditivo foi o modelo com PE.AT, LACUM.AT e PC.AT, embora com pouca diferença em relação ao modelo automático, para o caso geral, e relativo melhor desempenho para classificar empresas insolventes.

Em relação ao ajuste dos modelos aos dados, os modelos também apresentaram bastante similaridade. No entanto, ao contrário do que ocorreu com o desempenho preditivo, o modelo com maior R^2 de McFadden e menor AIC foi o modelo com PE.AT, LACUM.AT, MedioGrandePorte e GrandePorte.

Entretanto, cabe ressaltar que o presente estudo pode ser melhorado, inclusive no que tange a metodologia e a amostra utilizadas. A utilização de uma amostra com maior número de empresas insolventes e o emprego de outras técnicas estatísticas de classificação mais avançadas, tais como *Support Vector Machine* e *Decision Trees* são pontos a serem explorados em estudos futuros. Embora não tão citada na literatura, a regressão *Least Absolute Shrinkage and Selection Operator* (LASSO) também pode ser

empregada em estudos futuros, de forma a melhor selecionar a combinação de variáveis explicativas do modelo.

Além disso, os anos de 2015 e 2016 foram anos marcados pela crise política e econômica, o que pode ter afetado os resultados dos modelos. A replicação do estudo para outros anos pode ser mais um ponto positivo para análises futuras.

No entanto, para todos os modelos, pareceu haver evidência da validade da utilização de um modelo baseado em índices contábeis para prever a insolvência de uma empresa. Ademais, ao contrário do que é feito por muitos analistas, pelo menos estatisticamente, a utilização de um número muito elevado de indicadores parece não exercer contribuições adicionais, uma vez mantidas as variáveis mais relevantes.

7. Referências Bibliográficas

AKKOÇ, S. An empirical comparison of conventional techniques, neural networks and the three stage hybrid Adaptive Neuro Fuzzy Inference System (ANFIS) model for credit scoring analysis: The case of Turkish credit card data. **European Journal of Operational Research**, v. 222, p.168–178, 2012.

ALTMAN, E. I. Financial Ratios, Discriminate Analysis and the Prediction of Corporate Bankruptcy. **The Journal of Finance**, v. 23, n. 4, p. 589-609, set. 1968.

ALTMAN, E. I.; BAYIDA, T.; RIBEIRO, L.M.D Assessing Potential Financial Problems for Firms in Brazil. **Journal of International Business Studies**, p. 9-24, 1979.

ALTMAN, E. I.; HALDEMAN, R.; NARAYANAN, P. Zeta Analysis. **Journal of Banking and Finance**, p. 29-54, 1977.

ALTMAN, E. I.; HARTZELL, L; PECK, M. **Emerging Markets Corporate Bonds** A Scoring System. New York: Salomon Brothers Inc, 1995.

ALTMAN, E.; GIANNOZZI, A.; ROGGI, O.; SABATO, G. Building SME Rating: Is It Necessary for Lenders to Monitor Financial Statements of the Borrowers?" **Bancaria**, n. 10, 2013.

ALTMAN, E.; MARCO, G.; VARETTO, F. Corporate Distress Diagnosis: Comparisons Using Linear Discriminant Analysis and Neural Networks. **Journal of Banking and Finance**, v. 18, p. 505-529, 1994.

ALTMAN, E.I. The Success of Business Failure Prediction Models. **Journal of Banking and Finance**, v. 8, n.2, p. 171-198, 1984.

ALTMAN. E. I. Predicting Financial Distress of Companies: Revisiting the Z-Score and Zeta Models. In: BELL, A.R.; BROOKS, C; PROKOPCZUK, M. **Handbook of Research Methods and Applications in Empirical Finance**. New York: Edward Elgar Publishing. p. 428-456.

ARAÚJO, A; FUNCHAL, B.. A nova lei de falências brasileira: primeiros impactos. **Revista de Economia Política**, v. 29, n. 3, p. 191-212, set. 2009.

ASSAF, A.N.; BRITO, G.A.S. Modelo de Classificação de Risco de Crédito de Empresas. **Revista Contabilidade & Finanças**, v. 19, p. 18-29, 2008.

BALCAEN, S; OOGHE, H. 35 years of studies on business failure: an overview of the classical statistical methodologies and their related problems. **The British Accounting Review**, v. 38, p. 63-93, 2004.

BEAVER, W. Alternative accounting measures as predictors of failure. **The Accounting Review**, p. 112-122, 1968.

- BEAVER, W.H. Financial Ratios as Predictors of Failure. **Empirical Research in Accounting: Selected Studies. Supplement to Journal of Accounting Research**, p. 71-111, jan. 1966.
- BELLOVARY, J. *et al* . A Review of bankruptcy prediction studies: 1930-present. **Journal of Financial Education**, v. 33, p. 1-42, inverno de 2007.
- BLACK, F.; SCHOLES, M. The Pricing of Option and Corporate Liabilities. **Journal of Political Economy**, v. 81, n. 3, p. 637-654, 1973.
- BRASIL. Decreto-Lei n. 11.101, de 9 de fevereiro de 2005. **Diário Oficial [da] República Federativa do Brasil**, Brasília, DF, 9 fev. 2005. Seção 1, p.1.
- BRITO, G.A.S.; ASSAF NETO, A. A Model for the Classification of Companies Credit Risk. **Revista Contabilidade & Finanças**, v. 19, n. 46, p. 18-29, 2008.
- BRITO, G. A. S.; ASSAF NETO, A.; CORRAR, L. J. Sistema de classificação de risco de crédito: uma aplicação a companhias abertas no Brasil. **Revista Contabilidade & Finanças**, v. 20, n. 51, p. 28-43, 2009.
- BÜYÜKÖZTÜRK, S.; ÇOKLUK-BÖKEOĞLU, Ö. Discriminant Function Analysis: Concept and Application. **Eurasian Journal of Education Research**, v. 33, p. 73-92, 2008.
- CASEY, C.; BARTCZAK, N. Using Operating Cash Flow to Predict Financial Distress: Some Extensions. **Journal of Account Research**, v. 23, p. 384-401, 1985.
- CASTRO JUNIOR. **Previsão de insolvência de empresas brasileiras usando análise discriminante, regressão logística e redes neurais**, 2003. Dissertação de Mestrado - Programa de Pós-Graduação em Administração, São Paulo, SP: Universidade de São Paulo.
- CHEN, Y.; CHU, G. Estimation of Default Risk Based on KMV Model: An Empirical Study for Chinese Real Estate Companies. **Journal of Financial Risk Management**, v. 3, p. 40-49, 2014.
- CHUDSON, W. **The Pattern of Corporate Financial Structure**. New York: National Bureau of Economic Research, 1945.
- DAMBOLENA, I.; KHOURY, S. Ratio Stability and Corporate Failure. **The Journal of Finance**, v. 35, v. 4, p. 1017-1026, 1980.
- DEAKIN, E. B. A discriminant analysis of predictors of business failure. **Journal of Accounting Research**, p. 167-179, 1972.
- DIAMOND, JR., H. **Pattern Recognition and the Detection of Corporate Failure**. 1976. Tese (Doutorado). New York University, New York, 1976.

DWYER, M. **A Comparison of Statistical Techniques and Artificial Neural Network Models in Corporate Bankruptcy Prediction**. 1992. Tese (Doutorado). University of Wisconsin, Madison. 1992.

ELIZABETSKY, R. **Um modelo matemático para a decisão no banco comercial. Trabalho de formatura** – Departamento de Engenharia de Produção, Escola Politécnica da Universidade de São Paulo. São Paulo: POLI – USP, 1976.

FAUSETT, L. **Fundamentals of neural networks: architectures, algorithms, and applications**. Englewood-Cliffs: Prentice Hall, 1994.

FITZPATRICK, P.A. A Comparison of the Ratios of the Successful Industrial Enterprises with those of Failed Companies. **The Accountants Publishing Company**, 1932.

FORSYTH, T. **A Study of the Ability of Financial Ratios to Predict Corporate Failure and the Relationship between Bankruptcy Model Probability Assessments and Stock Market Behavior**. Tese (Doutorado) – University of Alabama, Alabama, 1991.

FRYDMAN, H.; ALTMAN, E.; KAO, D. Introducing Recursive Partitioning for Financial Classification: the Case of Financial Distress. **The Journal of Finance**, v. 40, n. 1, p. 269-291, 1985.

FULMER, J *et al.* A Bankruptcy Classification Model for Small Firms. **The Journal of Commercial Bank Lending**, v. 66, n. 11, p. 25-37, 1984.

GARDINER, L.; OSWALD, S.; JAHERA, J. Prediction of Hospital Failure: A Post PPS Analysis. **Hospital & Health Services Administration**, v. 41, n. 4, p. 441-460, 1996.

GOUVEA, M.A. *et al.* Análise de Risco de Crédito com Aplicação de Regressão Logística e Redes Neurais. **Contabilidade Vista & Revista**, v. 24, n. 4, p. 96-123, 2013.

GREENE, William. **Econometric Analysis**, 7th edition. New York: Pearson, 2012.

GUAN, Q. **Development of Optimal Network Structures for Back-Propagation-Trained Neural Networks**. Tese (Doutorado). University of Nebraska, Nebraska, 1993.

GUIMARÃES, A.; MOREIRA, T.B.S. Previsão de Insolvência: Um Modelo Baseado em Índices Contábeis com Utilização da Análise Discriminante. **Revista de Economia Contemporânea**, v. 12, p. 151-178, 2008.

GUIMARÃES, A.L.S.; ALVES, W.O. Prevendo a Insolvência de Operadoras de Planos de Saúde. **RAE**, v. 49, p. 459-471, 2009.

- HAIR Jr., J.F. *et al.* **Análise Multivariada de Dados**. 5. ed. Porto Alegre: Bookman, 2005.
- HOSMER, D.W.; LEMESHOW, S. **Applied Logistic Regression**. 2. ed. Estados Unidos: Wiley, 2000.
- JACKENDOFF, N. **A Study of Published Industry Financial and Operating Ratios**. Philadelphia: Temple University, Bureau of Economic and Business Research, 1962.
- JAMES, G. *et al.* **An Introduction to Statistical Learning with Applications in R**. 6. ed. Estados Unidos: Springer, 2000.
- JANOT, M. M. **Modelos de previsão de insolvência bancária no Brasil: aplicação de diferentes modelos entre 1995 e 1998**. 1999. 94 p. Dissertação (Mestrado) - Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 1999.
- KANITZ, S. C. **Como prever falências**. São Paulo: Abril, 1974.
- KO, L. *et al.* Prediction of Corporate Financial Distress: An Application of the Composite Rule Induction System. **The International Journal of Digital Accounting Research**, v. 1, n. 1, p. 69-85, 2001.
- KURUNMÄKI, L.; MILLER, P. Calculating failure: The making of a calculative infrastructure for forgiving and forecasting failure. **Business History**, v. 55, p. 1100-1118, 2013.
- LACHENBRUSCH, P.A. An Almost Unbiased Method of Obtaining Confidence Intervals for the Probability of Misclassification in Discriminant Analysis. **Biometrics**, v. 23, p. 639-645, dezembro de 1967.
- LAU, A. A Five-State Financial Distress Prediction Model. **Journal of Accounting Research**, v. 25, n. 1, p. 127-138, 1987.
- MAROCO, J. **Análise Estatística com Utilização do SPSS**, 3. ed. Lisboa: Sílabo, 2003.
- MATIAS, A. B. **Contribuição às técnicas de análise financeira: um modelo de concessão de crédito**. Trabalho de Formatura – Faculdade de Economia, Administração e Contabilidade. São Paulo: FEA – USP, 1978
- MCKEE, T. Predicting Bankruptcy via Induction. **Journal of Information Technology**, v. 10, n. 1, p. 26-36.
- MENSAH, Y. M. An Examination of the Stationarity of Multivariate Bankruptcy Prediction Models: A Methodological Study. **Journal of Accounting Research**, v. 22, n. 1, p. 380-395, 1983.

- MERTON, R. On the Pricing of Corporate Debt: the Risk Structure of Interest Rates. **The Journal of Finance**, Berkeley, v. 29, n. 2, p. 449-470, 1974.
- MERWIN, C. L. Financing Small Corporations in Five Manufacturing Industries, 1926–1936, **National Bureau of Economic Research**, v. 22, 1942.
- MESSIER, W.F. Jr; HANSEN, J.V. Inducing Rules for Expert Systems Development: An Example Using Default and Bankruptcy Data. **Management Science**, p. 1404-1415, 1988.
- MEYER, P.; PIFER, H. Prediction of Bank Failures. **Journal of Finance**, v. 25, n. 4, p. 853-868, 1970.
- MOSES, D.; S. LIAO. On Developing Models for Failure Prediction. **Journal of Commercial Bank Lending**, v. 69, p. 27-38, março 1987.
- ODOM, M. AND R. SHARDA. A neural network model for bankruptcy prediction. Article in *Neural Networks in Finance and Investing: Using Artificial Intelligence to Improve Real-World Performance*. **Chicago,IL: Probus Publishing Co.** 1993.
- OHLSON, J. A. Financial ratios and the probabilistic prediction of bankruptcy, **Journal of Accounting Research**, v. 18, v. p. 109-131, primavera de 1980.
- ONUSIC, L.M.; CASA NOVA, S.P.C. A Utilização Conjunta das Técnicas de Análise por Envoltória de Dados e Regressão Logística no Estudo de Insolvência de Empresas: Um Estudo Exploratório. **30º Encontro da ANPAD**, 2006.
- PATTERSON, D. **Bankruptcy Prediction: A Model for the Casino Industry**. 2001. Tese (Doutorado) – University of Nevada, Las Vegas, 2001.
- PEREIRA DA SILVA, José. **Administração de crédito e previsão de insolvência**. São Paulo: Atlas, 1982
- PEREIRA, V.S.; MARTINS, V.F. Estudos de previsão de falências – uma revisão das publicações internacionais e brasileiras de 1930 a 2015. **Revista Contemporânea de Contabilidade**, v.12, n. 26, p. 163-196, 2015.
- RAMSER, J. R.; FOSTER, L. O. A Demonstration of Ratios Analysis. **Bureau of Business Research**, University of Illinois, Bulletin n. 40, 1931.
- REQUIÃO, R. **Curso de Direito Falimentar**. 2. ed. São Paulo: Saraiva, 1998.
- REY, D. Stock Market Predictability: Is it There? A Critical Review. **Work paper, University of Basel**, WWZ/Department of Finance, 2003.
- ROYSTON J. P. An Extension of Shapiro and Wilk's W Test for Normality to Large Samples. **Journal of the Royal Statistical Society**, v. 31, n. 2, p. 115-124, 1982.

RUDORFER, G. Early Bankruptcy Prediction Using Neural Networks. **APL Quote Quad**, v. 25, n. 4, p. 171-176, 1995.

SANTOS, J.O.; SANTOS, J.A.R. O Modelo KMV e sua Utilidade no Processo de Análise do Risco de Crédito. **Revista de Gestão USP**, São Paulo, v. 16, n. 2., p. 73-92, abril-junho 2009.

SANTOS, J. O. **Análise de Crédito: Empresas, Pessoas Físicas e Agronegócio e Pecuária**. 3. ed. São Paulo: Atlas, 2010.

SANVICENTE, A.Z.; MINARDI, A.M.A.F. Identificação de Indicadores Contábeis Significativos para a Previsão de Concordata de Empresas. **Finance Lab Working Papers**, São Paulo. Outubro de 1998.

SCALZER *et al.* Insolvência Empresarial: um Estudo sobre as Distribuidoras de Energia Elétrica Brasileiras, **Revista Contemporânea de Contabilidade**, v. 12, n. 27, 2015.

SCOTT, W. R. **Financial Accounting Theory**. 2. ed. Prentice Hall: New Jersey, 1981.

SELAU, L.P.R.; RIBEIRO, J.L.D. Uma Sistemática Para Construção e Escolha de Modelos de Previsão de Risco de Crédito. **Gestão & Produção**, v. 16, n.3, p. 398-413, jul.-set. 2009.

SHARMA, S. **Applied Multivariate Techniques**. 1. ed. John Wiley & Sons, Inc.: Estados Unidos, 1996.

TAFFLER, R. Empirical Models for the Monitoring of UK Corporations. **Journal of Banking and Finance**, v. 8, n. 2, p. 199-227, 1984.

TAMARI, M. Financial Ratios as a Means of Forecasting Bankruptcy. **Management International Review**, v. 6, n. 4, p. 15-21, 1966.

THEODOSSIOU, P. Alternative Models for Assessing Financial Condition of Business in Greece. **Journal of Business Finance & Accounting**, v. 18, n. 5, p. 697-720, 1991.

TRIPPI, R.R.; LEE, J.K. **Artificial Intelligence in Finance & Investing: State-of-the-Art Technologies for Securities Selection and Portfolio Management**. Chicago: Irwin, 1996.

WILSON, R.; SHARDA, R. Bankruptcy Prediction Using Neural Networks. **Decision Support Systems**, v. 11, n. 5, p. 545-557, 1994.

WILCOX, J. A Prediction of Business Failure Using Accounting Data. **Journal of Account Research**, v. 11, p. 163-179, 1973.

WILSON, N. *et al.* Neural Networks Simulation and the Prediction of Corporate Outcomes: Some Empirical Findings. **International Journal of the Economics of Business**, v. 2, n. 1, p. 31-50, 1995.

WINAKOR, A.; SMITH, R. Changes in financial structure of unsuccessful industrial companies. **Bureau of Business Research**, v. 51. p. 19-45, 1935.

WOOLDRIDGE, J. M. **Introductory Econometrics: A Modern Approach**, 5. ed. Estados Unidos: Cengage, 2010.