

Uma aplicação do modelo de regressão logística na previsão de falência empresarial no Brasil

Resumo: Este trabalho tem como objetivo principal aplicar regressão logística de modo a chegar a um modelo capaz de prever o estado de inadimplência de empresas não financeiras brasileiras com pelo menos um ano de antecedência. Como objetivo secundário, o trabalho busca identificar as variáveis explicativas mais relevantes para prever o evento de inadimplência. Assim, foi elaborado um breve histórico sobre os modelos mais reconhecidos na literatura sobre o assunto, de modo a definir a especificação de partida do modelo. Utilizando os métodos de seleção manual e *backward stepwise*, foi possível chegar a um modelo com bom ajuste aos dados e excelente capacidade de predição.

Palavras-chave: Insolvência; regressão logística; logit; *backward stepwise*; falência.

Abstract: *The main objective of this study is to apply logistic regression so to arrive at a model that can predict the default status of Brazilian non-financial companies at least one year in advance. As a secondary objective, this work tries to identify which explanatory variables are more important to predict a bankruptcy event. Thus, a brief history of the most recognized models in the literature on the subject was elaborated so as to define the departure specification of the model. Using the manual and backward stepwise selection methods, it was possible to arrive at a model with good data fit and excellent predictive capacity.*

Keywords: *Insolvency; logistic regression; logit; backward stepwise; bankruptcy.*

Classificação JEL: C35; C58; M21; G33.

Paulo Roberto Fonteles
Guimarães¹

Moisés de Andrade Resende
Filho²

¹ Bacharel em Ciências Econômicas pela Universidade de Brasília (UnB). Mestrando em Economia pela Universidade Católica de Brasília (UCB). Bolsista de mestrado da CAPES. Analista de Investimentos da Fundação dos Economistas Federais (FUNCEF). E-mail: guimaraes.paulo95@gmail.com.

² Professor Associado do Departamento de Economia da Universidade de Brasília. E-mail: moisesresende@unb.br

1. Introdução

Do ponto de vista da teoria microeconômica, a falência de uma empresa é uma indicação de sua ineficiência e, portanto, que os recursos alocados em sua atividade produtiva devem ser alocados de maneira diferente na economia. No entanto, o processo de realocação de recursos gera custos diretos aos *stakeholders* e indiretos à sociedade, o que leva governos nacionais a tomarem medidas para minimizar esses custos. Por exemplo, entrou em vigor no Brasil em 9 de junho de 2005 a Lei nº 11.101/05 (BRASIL, 2005), também conhecida como Nova Lei de Falências, que regulamenta a recuperação judicial, extrajudicial e a falência do empresário e da sociedade empresária. De acordo com Araújo e Funchal (2009), após essa lei, os requerimentos e falências decretadas no Brasil apresentaram tendências de queda em relação ao ano anterior, respectivamente, de 50% e 46%.

Como o processo de falência de uma empresa envolve diversas partes e elevado custo social, a previsão de insolvência é de grande relevância não apenas para agentes privados, mas também para agentes públicos (BALCAEN e OOHGHE, 2004). Aos primeiros, em especial aos *stakeholders*, a capacidade de previsão de insolvência da firma de interesse pode contribuir para evitar perdas individuais 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.

Portanto, a correta utilização de modelos de previsão de insolvência permite reduzir os custos associados ao processo de falência, uma vez que, identificada uma 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 mesma. Portanto, este tipo de modelagem permite que investidores precifiquem mais corretamente valores mobiliários e ações.

O presente estudo apresenta o processo de especificação de um modelo de classificação de risco de solvência aplicado ao caso brasileiro, a partir da estimação de um modelo de Regressão Logística e da aplicação de diversas técnicas e testes estatísticos e computacionais como, por exemplo, o uso do algoritmo *backward stepwise*. Além dos índices contábeis, já amplamente utilizados em estudos similares, o presente trabalho testa a inclusão de variáveis externas ao contexto financeiro-operacional da empresa. Para tanto, utiliza-se uma amostra de mais de 400 empresas dos mais diversos setores e, com dados respectivos ao final de 2014, busca-se prever o estado de solvência dessas empresas em 2015 e 2016.

Além dessa introdução, o presente artigo está estruturado tal que na seção 2 apresenta o referencial teórico, na seção 3, a estrutura conceitual, na seção 4, os modelos e procedimentos econométricos adotados, na seção 5, a fonte e análise preliminar dos dados, na seção 6, os resultados e discussão e, finalmente, na seção 7, as conclusões.

2. Histórico dos modelos e técnicas estatísticas de previsão de falência

De forma geral, seja no caso brasileiro ou não, a literatura sobre previsão de um evento de falência empresarial, o qual pode ser retratado de várias maneiras, envolve

a utilização de três técnicas estatísticas: a análise discriminante, a regressão logística e as redes neurais artificiais. A maioria dos estudos usa índices contábeis como variáveis explicativas do evento de falência, sendo mais rara a utilização de variáveis não presentes nos demonstrativos financeiros das empresas.

Os primeiros estudos que buscavam compreender o processo de falência corporativa datam de 1930 e usavam demonstrativos financeiros e índices contábeis para identificar padrões e tendências que levavam empresas saudáveis a se tornarem insolventes. Dentre esses estudos, Ramser e Foster (1931) apresentam um relatório analisando 24 índices financeiros de 29 firmas para tentar determinar características comuns às firmas em processo de falência e 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. Winakor e Smith (1935), Merwin (1942), Chudson (1945) e Jackendoff (1962) também tentaram detectar indicadores financeiros relevantes para caracterizar o processo de falência, comparando empresas solventes a empresas insolventes.

A partir de 1966, há uma revolução nos estudos de previsão de falência corporativa com os trabalhos de Beaver (1966) e Altman (1968) que passam a empregar a análise discriminante. Beaver (1966) revolucionou o campo da previsão de insolvência corporativa, com a introdução de um modelo de análise discriminante univariada para uma amostra de 158 empresas (79 insolventes e 79 solventes) de 1954 e 1964, cujos ativos variavam de pouco mais de meio milhão a 45 milhões de dólares. Altman (1968) tornou-se uma das maiores referências na literatura sobre previsão de insolvência corporativa ao utilizar pela primeira vez a técnica da análise discriminante multivariada. A amostra utilizada no estudo era 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. Para selecionar os indicadores financeiros, Altman (1968) se baseou primeiramente na popularidade do indicador na literatura e na potencial relevância do mesmo para o estudo. Em seguida, refinou a seleção dos indicadores com base em testes de significância estatística, na correlação deles (a fim de evitar multicolinearidade), na precisão preditiva de cada indicador, medida pelo erro-padrão e, finalmente, no juízo pessoal. O modelo final foi batizado de *Z-Score*.

A análise discriminante é aplicável aos casos em que a amostra pode ser dividida em grupos baseados em uma variável dependente qualitativa de modo a 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 se resumem a dois grupos: o grupo das empresas solventes e o grupo das empresas insolventes. A discriminação é realizada por meio da maximização da variância entre os grupos e, conseqüentemente, minimização da variância intragrupo (HAIR et al., 2005).

Contudo, a análise discriminante exige suposições fortes como a normalidade das variáveis independentes e a igualdade nas matrizes de dispersão e covariâncias dos grupos solventes e insolventes. Com isso, estudos utilizando o método da regressão logística ou logit (OHLSON, 1980; DAMBOLENA e SHULMAN, 1988; AZIZ et al., 1989; TINOCO e WILSON, 2013) se tornaram mais populares em detrimento dos que usavam a análise discriminante (BELLOVARY et al., 2007). Ohlson (1980) foi um dos primeiros e mais importantes estudos a utilizar a análise logit na previsão de falência corporativa. Para tanto, utilizou dados de empresas da segunda metade da década de 1970. O presente estudo utiliza a análise logit.

2.1. Histórico dos modelos e técnicas estatísticas de previsão de falência no Brasil

No Brasil, os modelos de previsão de insolvência se tornaram conhecidos devido ao modelo conhecido como Termômetro de Kanitz (1974). Kanitz (1974) analisou, aproximadamente, 5.000 demonstrações contábeis de empresas brasileiras e aplicou análise discriminante a uma amostra de 21 empresas falidas e 21 empresas solventes selecionou aleatoriamente, entre 1972 e 1974. 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 utilizada uma amostra de 373 empresas, das quais 274 eram classificadas como solventes e 99 como insolventes. 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, aplicando-se a análise discriminante múltipla.

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, Guimarães e Moreira (2008) 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. Assaf Neto e Brito (2008) construíram um modelo baseado em Regressão Logística a partir de uma amostra de 60 empresas, 30 insolventes que se tornaram concordatárias ou falidas entre 1994 e 2004 e 30 solventes. Em 2009, Guimarães e Alves (2009) aplicaram o modelo Z de Altman (1968) e construíram um modelo próprio de regressão logística para uma amostra de 706 empresas, todas do setor de planos de saúde.

Segundo levantamento de Pereira e Martins (2015) que compila 48 estudos brasileiros sobre o tema previsão de insolvência, de 1974 a 2015, menos de 25% dos trabalhos aplicaram o modelo de regressão logística para prever o estado de insolvência de empresas. Além disso, a maioria dos estudos que usaram esta técnica surgiu após os anos 2000.

Portanto, apesar dos recentes esforços para aplicação de modelos *logit* de previsão de insolvência ao caso brasileiro, ainda há espaço para o desenvolvimento de novos modelos com esta técnica que utilizem técnicas estatísticas mais recentes, amostras com grande número de empresas e que incluam novos índices contábeis e variáveis exógenas ao contexto financeiro operacional da empresa. Por exemplo, Pereira e Martins (2015) sugerem que novos trabalhos no tema incluam informações não contábeis nos modelos, tendo em vista que a utilização desse tipo de variáveis tem sido pouco explorada na literatura. Além disso, também recomendam que se incluam variáveis de fluxo de caixa no modelo. Desta forma, o presente trabalho testará a inclusão de ambos os tipos de variáveis.

No momento da estimação, para se evitar o problema de *overfitting* do modelo aos dados, adotou-se o procedimento de validação cruzada, dividindo-se a amostra, aleatoriamente, em duas subamostras: uma para estimação dos modelos e outra para testar suas classificações (JAMES et al., 2000). Neste sentido, a amostra utilizada para estimação do modelo ficou composta por 308 empresas solventes e 32 empresas insolventes, enquanto que a amostra utilizada para teste ficou composta por 49 empresas solventes e 11 insolventes. Ademais, o modelo escolhido também foi utilizado para prever o estado de solvência no terceiro trimestre de 2016, gerando uma defasagem de, aproximadamente, dois anos para as variáveis explicativas. Decidiu-se pela utilização deste período por ser o mais recente possível no momento da realização deste estudo.

3. Metodologia

Seguindo a tradição recente dos modelos de previsão de insolvência, o presente estudo utiliza o método de regressão logística. No modelo de Regressão Logística, assume-se que há uma variável latente y_i^* (não observável) que representa a propensão de que a variável resposta y seja um sucesso, ou seja, assumo o valor 1, tal que:

$$y_i^* = x_i' \beta + \varepsilon_i \quad (1)$$

onde:

x_i é o vetor das K variáveis explicativas do modelo;

β 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 $\pi^2/3$ (GREENE, 2012: p. 686).

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

$$y_i = 1(y_i^* > 0) \equiv 1(x_i' \beta + \varepsilon_i > 0) \quad (2)$$

em que $1(\cdot)$ é uma função que assume o valor um se a condição entre parenteses ocorre e zero, caso contrário.

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

$$\ln L(\tilde{\beta}) = \sum_{i \in S} \ln \Lambda(x_i' \tilde{\beta}) + \sum_{i \notin S} \ln(1 - \Lambda(x_i' \tilde{\beta})) \quad (3)$$

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 $y_i=1$) e $i \notin S$ denota uma observação não pertencente ao conjunto S (ou seja, o conjunto em que $y_i=0$).

Em suma, o estimador de máxima verossimilhança $\hat{\beta}$ de β maximiza a verossimilhança de se obter a amostra de dados observada (HOSMER e LEMESHOW, 2000) e $\Lambda(x_i' \hat{\beta})$ são estimativas da probabilidade de sucesso e, portanto, estão necessariamente no intervalo de 0 a 1, o que é muito conveniente em termos de interpretação. Vários aspectos da Regressão Logística são similares à Regressão Linear Múltipla, como por exemplo, a mensuração da precisão das estimativas por meio do erro-padrão, a utilização de testes estatísticos para inferência, entre outros (JAMES et al., 2000). Por isso, o modelo logit é utilizado no presente estudo.

3.1. Proxy de insolvência

Para fins de classificar empresas como insolventes ou não na amostra foram considerados dois critérios concorrentes apresentados a seguir.

O primeiro critério de classificação considerado foi o mesmo estabelecido pela Comissão de Valores Mobiliários (CVM). No entanto, considerando as empresas classificadas como “Em Liquidação Judicial”, “Em Liquidação Extrajudicial”, “Liquidação Extrajudicial”, “Em Recuperação Judicial ou Equivalente”, “Falida” e “Paralisada”, o total de empresas classificadas como insolventes foi muito baixo para fins de análise, apenas 19 em um total de 401 empresas na amostra final.

O segundo critério de classificação leva em conta o Passivo a Descoberto da empresa. Assim, uma empresa é classificada como insolvente se apresentar Patrimônio Líquido Negativo ao final do exercício social de 2015 e em dois ou mais trimestres antecedentes. Embora seja possível que uma empresa consiga se manter operacional

mesmo se suas exigibilidades superam seus ativos, vários autores (RUBENS REQUIÃO, 1998; GUIMARÃES e ALVES, 2009) 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 insolvência.

Decidiu-se por utilizar o segundo critério de classificação, pois o total de empresas classificadas como insolventes foi de 43 em um total de 401 empresas na amostra final. Ademais, a utilização de um maior grupo de empresas insolventes pode evitar problemas na estimação dos modelos, como o de se encontrar alguma relação válida entre a variável resposta e alguma variável explicativa exclusivamente para o grupo de empresas insolventes na amostra e não para a população de empresas como um todo (HOSMER e LEMESHOW, 2000).

3.2. A Amostra

Para realização do estudo, construiu-se uma amostra inicial de 615 empresas, todas retiradas do *software* Economática. Elegeu-se, como momento do evento de falência modelado pela variável resposta, o fim do exercício social de 2015, que era o último ano com exercício social fechado no momento de elaboração do estudo. Eliminaram-se da amostra inicial as empresas que, segundo classificação da Economática, são instituições financeiras ou prestam algum tipo de serviço financeiro como, por exemplo, seguradoras. Adotou-se tal procedimento com base no fato de que tais empresas apresentam demonstrações muito diferentes das empresas não financeiras e na constatação de que os indicadores contábeis normalmente utilizados na análise de insolvência de empresas financeiras diferirem daqueles utilizados em estudos com empresas não financeiras. Além disso, eliminaram-se da amostra também as 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. Assim, a amostra final é constituída por 401 companhias não financeiras, tendo sido 43 destas classificadas como insolventes e 358, classificadas como solventes.

3.3. Seleção de Variáveis Explicativas

A seleção das variáveis a serem incluídas no vetor x_i das K variáveis explicativas do modelo de previsão de insolvência foi realizada com base, principalmente, em Sanvicente e Minardi (1998), Pereira e Martins (2015) e Bellovary (2007), conforme está descrito na Tabela 1.

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

Variável	Abreviação	Estudo de referência
Log da Receita Total	LGRT	Pereira e Martins (2015)
Lucro Líquido/Ativo Total	LL.AT	Bellovary (2007)
Liquidez Corrente	AC.PC	Bellovary (2007)
Capital de Giro/Ativo Total	CG.AT	Sanvicenti e Minardi (1998)
Lucros Acumulados/Ativo Total	LACUM.AT	Bellovary (2007)
EBIT/Ativo Total	EBIT.AT	Bellovary (2007)
Liquidez Seca	AC.EST.PC	Pereira e Martins (2015)
Dívida Total/Ativo Total	DT.AT	Pereira e Martins (2015)
Ativo Circulante/Ativo Total	AC.AT	Bellovary (2007)
Caixa/Ativo Total	CX.AT	Bellovary (2007)
Passivo Exigível/Ativo Total	PE.AT	Bellovary (2007)
Fluxo de Caixa Operacional/Ativo Total	FCO.AT	Sanvicenti e Minardi (1998)
Passivo Circulante/Ativo Total	PC.AT	Pereira e Martins (2015)
Caixa Operacional/Passivo Exigível	FCO.PE	Sanvicenti e Minardi (1998)
Dívida Total/Patrimônio Líquido	DT.PL	Pereira e Martins (2015)
Fluxo de Caixa Operacional/Dívida Total	FCO.DT	Sanvicenti e Minardi (1998)
EBIT/Despesas Financeiras	EBIT.DFIN	Standard & Poor's (1997)
Passivo Exigível/Patrimônio Líquido	PE.PL	Pereira e Martins (2015)
(Ativo Circulante - Estoques)/Ativo Total	AC.EST.AT	Pereira e Martins (2015)
Patrimônio Líquido/Ativo Total	PL.AT	Bellovary (2007)
Empréstimos de Longo Prazo/Ativo Total	EMLP.AT	Bellovary (2007)
Caixa/Passivo Circulante	CX.PC	Pereira e Martins (2015)
Caixa Operacional/Passivo Circulante	FCO.PC	Sanvicenti e Minardi (1998)
Capital de Giro/Patrimônio Líquido	CG.PL	Bellovary (2007)
Log do Ativo Total	LGAT	Pereira e Martins (2015)
Grande Porte	Grande.Porte	-
Médio-Grande Porte	MedioGrandePorte	-
Médio Porte	MedioPorte	-

Fonte: Elaboração própria.

Os dados dos indicadores financeiros correspondem à mediana dos valores observados 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, com defasagem de, aproximadamente, um ano em relação ao período do evento de falência. (OLHSON, 1980).

As variáveis “GrandePorte”, “Médio-Grande Porte” e “Médio Porte” são variáveis dummy, que são iguais a 1 caso o nível de faturamento da empresa corresponda ao porte indicado pela variável e zero, caso contrário. Para construção das variáveis, foi utilizada a classificação de porte empresarial realizada pelo Banco Nacional do Desenvolvimento Econômico e Social (BNDES)³.

Além das variáveis explicativas proveniente de demonstrativos financeiros, seguindo o que sugerem Pereira e Martins (2015), decidiu-se pré-selecionar cinco variáveis externas ao contexto contábil, quais sejam: número de vezes que as ações da empresa foram negociadas no 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

³ Disponível em <http://www.bndes.gov.br/wps/portal/site/home/financiamento/guia/quem-pode-ser-cliente/>

das ações da empresa em relação ao Ibovespa e, finalmente, uma variável *dummy* que recebe 1 se a empresa tem ações listadas na Bovespa e zero, caso contrário. Tais variáveis são apresentadas na Tabela 2.

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 apresentadas nas Tabelas 1 e 2, têm-se ainda a variável "Insolvente" que recebe o valor 1 se a empresa é classificada como insolvente no quarto trimestre de 2015 e 0, caso contrário e a variável "Insolvente.2016" que recebe 1 se a empresa está insolvente no terceiro trimestre de 2016 e 0, caso contrário. A variável "Insolvente" foi utilizada como variável dependente na construção do modelo, ao passo que a variável "Insolvente.2016" foi utilizada para testar o modelo.

As estatísticas descritivas das variáveis explicativas consideradas nos modelos no presente estudo estão apresentadas na Tabela 3.

Tabela 3. Estatísticas Descritivas das Variáveis Explicativas.

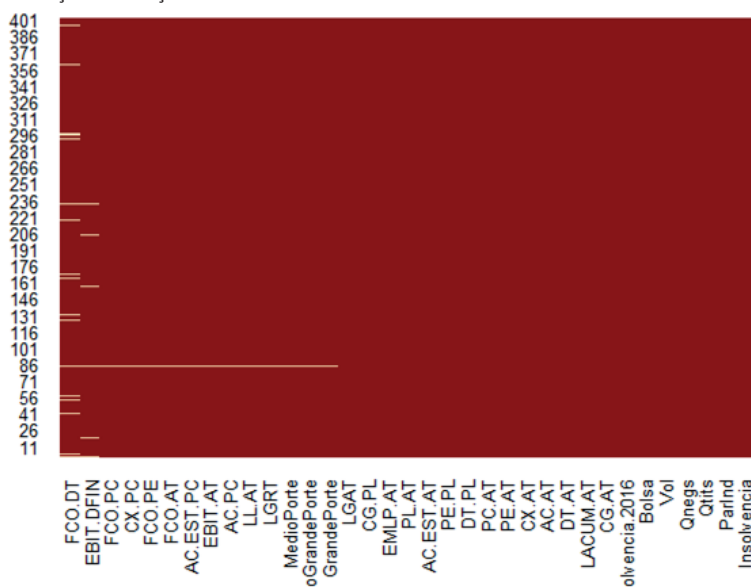
Variável	Média	Mínimo	1º Quartil	Mediana	3º Quartil	Máximo
ParInd	0,14	0,00	0,00	0,00	0,00	7,16
Qtits	86.288,00	0,00	0,00	126,00	21.865	7.767.888,00
Qnegs	192.651,00	0,00	0,00	262,00	69.969	7.223.334,00
Vol	1.250.949	0,00	0,00	1.062,00	275.362	96.291.487
Bolsa	0,74	0,00	0,00	1,00	1,00	1,00
GrandePorte	0,73	0,00	0,00	1,00	1,00	1,00
MedioGrandePort	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	-2.414,2	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 Tabela 3, algumas variáveis apresentam distribuição assimétrica, tendo em vista o intervalo interquartil entre a mediana e o terceiro quartil e entre este e o valor máximo. Este é o caso, por exemplo, das variáveis Qtits, Qnegs, LGRT, PE.PL, AC.PC, entre outras. Com isso, pode-se perceber que há empresas que destoam muito 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.

4. Construção e Seleção dos Modelos

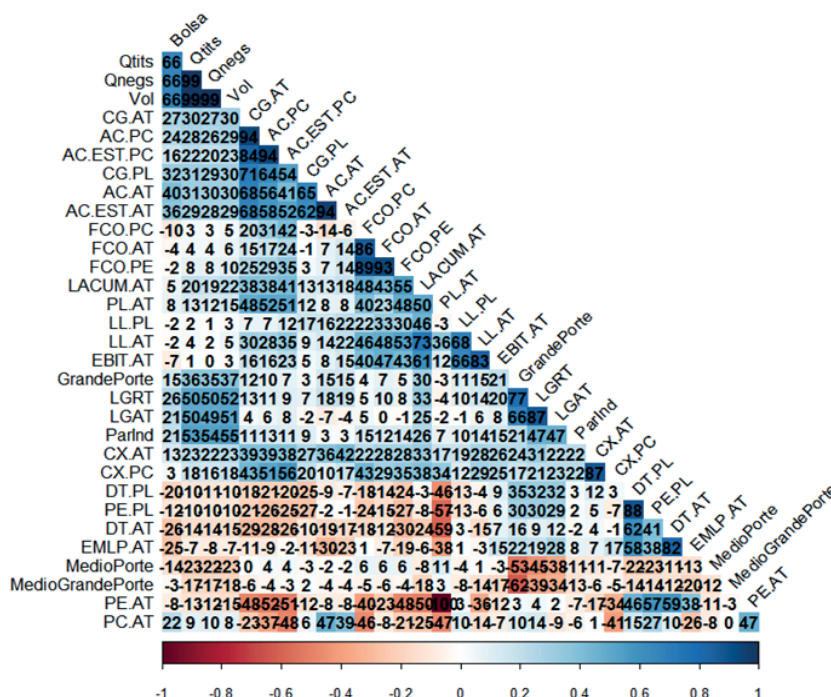


Fonte: Elaboração própria.

A interpretação da Figura 1 é bastante simples, os segmentos em branco denotam as variáveis (eixo x) faltantes para cada observação no eixo y. Por exemplo, a observação 86 não tem dados para as variáveis FCO.CT, EBIT.DFIN e, assim, sucessivamente até a variável LG.AT. Conforme se pode notar, as variáveis FCO.DT e EBIT.DFIN foram as que apresentaram maior número de observações com dados faltantes e, por isso, foram retiradas das análises. Por sua vez, a observação de número 86 não apresentou dados para uma grande quantidade de variáveis explicativas e, portanto, foi eliminada da amostra, reduzindo-a para 400 observações.

A matriz de correlação de posto de Spearman das variáveis explicativas é apresentada com a Figura 2.

Figura 2. Representação da matriz de correlação das variáveis explicativas.



Fonte: Elaboração própria.

A Figura 2 é intuitiva no sentido de que cada quadrado apresenta o coeficiente de correlação de Spearman do par de variáveis em porcentagem, tal que quanto mais vermelho, mais próximo de -100% e, quanto mais azul, mais próximo de 100%.

A visualização da correlação entre duas variáveis ajuda evitar problemas de multicolinearidade no modelo, isto é, forte correlação entre variáveis explicativas. Assim, é interessante manter no modelo final apenas uma variável explicativa dos pares de variáveis que apresentem uma correlação muito elevada, desde que não exista nenhuma razão teórica para a manutenção das duas variáveis.

Para ajudar na seleção das variáveis explicativas a serem, com base em Scalzer et al. (2015), 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 (2015), o resultado do teste de Mann-Whitney-Wilcoxon permite detectar quais variáveis são diferentes entre os grupos de empresas solventes e insolventes e, portanto, seriam, possivelmente, boas classificadoras. Embora, para algumas variáveis, não tenha sido possível rejeitar a hipótese nula de igualdade das funções de distribuição, optou-se por não retirá-las da lista de candidatas aos modelos. As variáveis explicativas candidatas são apresentadas na Tabela 4 na ordem crescente dos p-valores do teste de Mann-Whitney-Wilcoxon

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.

Pela Tabela 4, 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 com 340 observações, enquanto a de teste, com 60 observações, isto é, 15% da amostra total.

A seleção das variáveis explicativas do modelo final foi efetuada segundo duas metodologias: manual e automática, conforme apresentado nas seções a seguir.

4.1. Seleção do Modelo pelo Método Manual

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

Tabela 5. Resultados do Modelo Nulo.

Coeficientes	Estimativa	Erro-Padrão	Wald p-valor
(Intercepto)	-2,2644***	0,1857	<2e-16
AIC:	214,13	McFadden:	0
Número de Iterações de Fisher:	5		

Nota: ***, **, * denotam significante ao nível de 1%, 5% e 10%.

Fonte: Elaboração própria.

Conforme se pode notar pela Tabela 5, 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 R^2 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). Para testar a inclusão de cada variável ao modelo, foi elaborada a Tabela 6.

Tabela 6. Resultados para Inclusão da Primeira Variável.

Variável	p-valor de Wald	AIC	R^2 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 Tabela 6, as variáveis PE.AT e PL.AT foram as que apresentaram menor AIC e maior R^2 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. Tal resultado condiz com o esperado, pois empresas mais alavancadas devem apresentar maior risco de insolvência. De maneira geral, o aumento de capital de terceiros na estrutura de capital de uma companhia, embora constitua uma forma de financiamento que, *ceteris paribus*, reduz o custo ponderado médio de capital, tendo em vista, por exemplo, o benefício fiscal associado a tal tipo de capital, representa uma elevação do risco de solvência. Cada empresa deve, portanto, escolher uma estrutura de capital que otimize a relação risco-retorno pretendida.

Além disso, o evento de falência modelado é a condição de uma estrutura de capital cujo volume de capital de terceiros é superior aos ativos da empresa em 2015, de forma que, uma vez exigidas as obrigações devidas a terceiros, a companhia não tem recursos para honrar todas as dívidas. Faz sentido, portanto, que o volume de recursos devido a terceiros em 2014, isto é, a alavancagem em 2014, ajude a explicar a solvência um ano depois.

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.

A terceira etapa envolveu testar a possibilidade de incluir uma segunda variável explicativa no modelo. Nessa etapa, 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 estatísticas da Tabela 6. Desta vez, o coeficiente de correlação entre a variável PE.AT e cada uma das variáveis também foi considerado, com o intuito de se avaliar a possibilidade de existência de multicolinearidade e também a taxa de acerto de cada um dos modelos.

O modelo com melhor desempenho foi aquele com a inclusão da variável LACUM.AT, o que gerou um resultado condizente com a teoria. Ao fim de cada exercício social, a empresa pode auferir lucros ou prejuízos. De maneira simplificada, uma vez tendo apresentado lucros, a empresa pode optar por distribuir dividendos ou por reter lucros para, por exemplo, realizar futuros investimentos. Portanto, de maneira geral, uma empresa com elevado volume de lucros acumulados, é uma empresa que, recorrentemente, auferiu lucros em seus exercícios sociais passados.

Finalmente, na quarta etapa, testou-se a inclusão de uma terceira variável explicativa ao modelo. 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 melhorou as medidas de AIC e R^2 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 contribuiria para melhorar o ajuste do modelo aos dados de maneira significativa. Entretanto, para fins comparativos, optou-se pela inclusão de mais uma variável explicativa ao modelo.

Desta maneira, observou-se que as variáveis cuja inclusão, efetivamente, reduzem o AIC, aumentam o R^2 de McFadden e melhoram a previsibilidade do modelo, 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. Como a inclusão de tais variáveis poderia trazer estimativas infladas e como não há nenhuma fundamentação teórica para inclusão conjunta destas, optou-se por não considerá-las nessa etapa.

As Tabelas 7 e 8 apresentam os resultados para os modelos que acrescentam, respectivamente, as variáveis PC.AT e MedioGrandePorte. Além disso, na Tabela 8, estimaram-se os modelos também com 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 7. Resultados do Modelo com PE.AT, LACUM.AT e PC.AT.

Coeficientes	Estimativa	Erro-Padrão	Wald p-valor
(Intercepto)	-24,5577***	6,3399	0,0001
PE.AT	23,1752***	6,5336	0,0004
LACUM.AT	-1,1793*	0,4757	0,0132
PC.AT	3,9184	2,441	0,1084
AIC:	49,66	McFadden:	80,36%
Número de Iterações de Fisher:	11		

Nota: ***, **, * denotam significante ao nível de 1%, 5% e 10%.

Fonte: Elaboração própria.

Tabela 8. Resultados do Modelo com PE.AT, LACUM.AT, MedioGrandePorte, PE.AT*MedioGrandePorte e LACUM.AT*MedioGrandePorte.

Coeficientes	Estimativa	Erro-Padrão	Wald p-valor
(Intercepto)	-18,8579***	5,0604	0,0002
PE.AT	18,3316***	5,5283	0,0009
LACUM.AT	-1,1888**	0,4615	0,0100
MedioGrandePorte	-301,7378	43.209,7133	0,9944
PE.AT*MedioGrandePorte	314,5209	42.426,4468	0,9945
LACUM.AT*MedioGrandePorte	-96,9357	17.089,1619	0,9955
AIC:	50,34	McFadden:	81,92%
Número de Iterações de Fisher:	23		

Nota: ***, **, * denotam significante ao nível de 1%, 5% e 10%.

Fonte: Elaboração própria.

Conforme podemos observar pelas Tabelas 7 e 8, alguns dos coeficientes estimados apresentaram elevada magnitude, assim como elevados erros-padrão. Pode-se notar ainda 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 8 apresentaram erros-padrões elevados, resultando em grandes p-valores. Além disso, o AIC de nenhum dos modelos das Tabelas 7 e 8 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 na Tabela 8, esperava-se que, para uma empresa

de porte maior, um acréscimo aos valores de PE.AT e LACUM.AT contribuísssem 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.

Para fins de escolher entre os dois modelos, decidiu-se por aquele com maior taxa de acerto na amostra de teste, na amostra de maneira geral e para 2016, tendo em vista que o principal objetivo da modelagem é a correta classificação das empresas, especialmente das insolventes.

Tendo em vista os resultados encontrados, decidiu-se pelo modelo com PE.AT, LACUM.AT e PC.AT como modelo final do método manual.

Embora o objetivo principal do modelo seja a correta previsão das empresas nos grupos, seguem abaixo, os valores estimados dos efeitos marginais das variáveis do modelo escolhido:

Tabela 9. Efeitos Marginais do Modelo Selecionado pelo Método Manual.

Coefficientes	Efeito Marginal	Erro-Padrão	P > z
PE.AT	4,95943**	1,76247	0,004894
LACUM.AT	-0,25237	0,10977	0,021501
PC.AT	0,83853	0,55812	0,132991

Nota: ***, **, * denotam significativo ao nível de 1%, 5% e 10%.

Fonte: Elaboração própria.

Conforme se pode perceber pela Tabela 9, há um efeito marginal, estatisticamente significativo, a um nível de significância de 5% da variável PE.AT. Portanto, cada aumento de 0,1 em PE.AT, ou seja, 10% na relação Passivo Exigível/Ativo Total, aumenta a probabilidade de falência em 0,05%. Todas as outras variáveis não apresentaram efeitos marginais estatisticamente significantes, a um nível de significância de 5%.

4.2. Seleção do Modelo pelo Método Automático Backward Stepwise

O método selecionado para o procedimento automático foi o *Backward Stepwise* que consiste em um algoritmo que elimina, a cada passo do processo, variáveis explicativas do Modelo Completo na Tabela 10 que não foram individualmente significativas a 10%. 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 10.

Tabela 10. Resultados do Modelo Completo.

Coeficientes	Estimativa	Erro-Padrão	Wald p-valor
(Intercepto)	-2,76E+01	3,13E+01	0,3795
PE.AT	1,99E+01**	6,35E+00	0,0017
LACUM.AT	-1,71E+00*	9,89E-01	0,0835
PC.AT	2,71E+00	3,35E+00	0,4181
CG.AT	2,91E-01	3,33E+00	0,9306
DT.AT	1,42E+00	2,42E+00	0,5581
CX.PC	6,11E-01	1,59E+00	0,7011
FCO.PE	-2,42E+00	7,88E+00	0,7591
LGAT	-4,03E-01	6,28E-01	0,5208
Vol	5,55E-07	6,93E-07	0,4234
ParInd	-6,76E+01	1,32E+04	0,9959
GrandePorte	1,11E+01	2,93E+01	0,7038
MedioGrandePorte	1,14E+01	2,92E+01	0,6959
PE.PL	2,27E-02	4,59E-02	0,6206
AIC:	61,576	McFadden:	84,17%
Número de Iterações de Fisher:	23		

Nota: ***, **, * denotam significante ao nível de 1%, 5% e 10%.

Fonte: Elaboração própria.

Conforme pode ser observado na Tabela 10, 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 11. Resultados do Modelo Selecionado por 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
AIC:	47,22	McFadden:	82,45%
Número de Iterações de Fisher:	12		

Nota: ***, **, * denotam significante ao nível de 1%, 5% e 10%.

Fonte: Elaboração própria.

Diferentemente do ocorrido com o modelo descrito na Tabela 8, as variáveis de porte apresentaram, embora estatisticamente insignificantes a um nível de significância de 5%, sinais contrários aos esperados para as estimativas dos coeficientes.

As estimativas dos efeitos marginais das variáveis do modelo na Tabela 11 estão apresentadas na Tabela 12.

Tabela 12. Efeitos Marginais das Variáveis do Modelo Automático.

Coeficientes	dF/dx	Erro-Padrão	P> z
PE.AT	2,04928	6,34858	0,7469
LACUM.AT	-0,18066	0,56047	0,7472
GrandePorte	0,61985	0,91703	0,4991
MedioGrandePorte	0,97373***	0,20372	0,0000

Nota: ***, **, * denotam significativo ao nível de 1%, 5% e 10%.

Fonte: Elaboração própria.

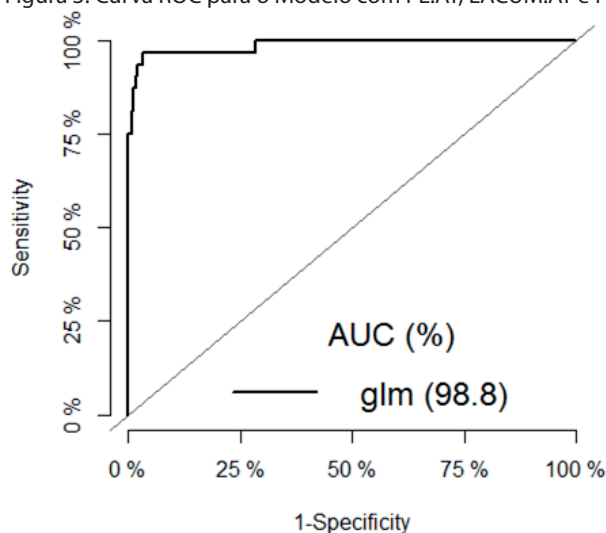
Desta vez, o efeito marginal da variável PE.AT não foi estatisticamente significativo, de forma que o único efeito marginal considerado estatisticamente significativo foi o da variável MedioGrandePorte, embora com um efeito diferente do esperado, isto é, positivo. Ou seja, uma empresa de médio ou grande porte tem 0,97% mais chance de falir que uma empresa pequena. Assim, para fins de teste, aplicou-se o Backward Stepwise a partir do modelo da Tabela 10, porém sem as variáveis de porte. O modelo encontrado foi o mesmo ao modelo final do método manual na Tabela 7.

5. Resultados

Após a etapa de seleção de modelos, os modelos selecionados foram utilizados para construir curvas ROC (Receiver Operating Characteristic), para que se fosse possível representar, graficamente, o desempenho preditivo dos modelos enquanto classificadores binários. Cabe frisar, novamente, que o principal objetivo da modelagem é encontrar modelos com elevada capacidade preditiva, uma vez esta característica ajuda a diminuir os custos associados à falência descritos na Introdução.

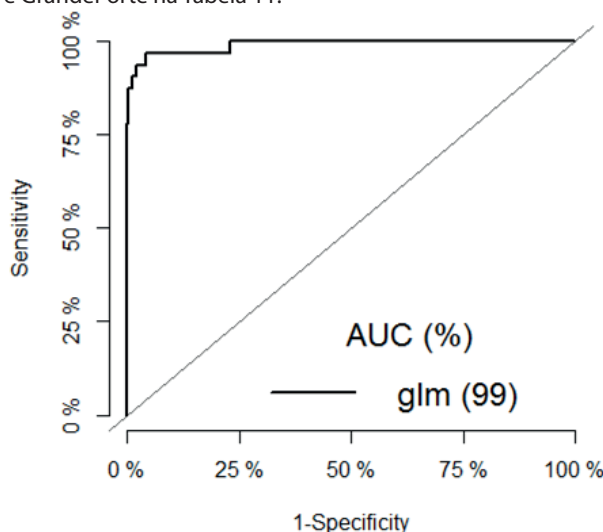
Construiu-se a curva ROC para os dois modelos encontrados ao final do método automático e para o modelo encontrado ao final do método manual. Como existe um modelo em comum, foram construídas duas curvas ROC, exibidas a seguir.

Figura 3. Curva ROC para o Modelo com PE.AT, LACUM.AT e PC.AT na Tabela 7.



Fonte: Elaboração própria.

Figura 4. Curva ROC para o Modelo Manual com PE.AT, LACUM.AT, MedioGrandePorte e GrandePorte na Tabela 11.



Fonte: Elaboração própria.

Conforme pode ser observado pela análise das curvas ROC nas Figuras 3 e 4, 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.

A Tabela 13 apresenta 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. No escopo desse estudo, classificar corretamente empresas insolventes representa o principal objetivo, tendo em vista o maior custo de se classificar incorretamente uma empresa insolvente como solvente. Por este motivo, adotou-se um ponto de corte conservador de 0,35.

Tabela 13. 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+MedioGrandePorte+GrandePorte	95,00%	97,75%	96,00%

Fonte: Elaboração própria.

Conforme se pode 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.

A Tabela 14 apresenta 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.

Tabela 14. 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+MedioGrandePorte+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

Considerando-se as estimativas dos modelos selecionados, 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 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 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. Além disso, no presente caso, a utilização de apenas três índices contábeis foi suficiente para prever a insolvência.

Referências

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. 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.

AZIZ, A.; EMANUEL, D.C.; LAWSON, G.H. Bankruptcy Prediction: An Investigation of Cash Flow Based Models. *Journal of Management Studies*, p. 419-437, 1988.

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.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.

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.

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.

CHUDSON, W. The Pattern of Corporate Financial Structure. New York: National Bureau of Economic Research, 1945.

DAMBOLENA, L.G.; SHULMAN, J.M. A Primary Rule for Detecting Bankruptcy: Watch the Cash. Financial Analysts Journal, p. 74-78, 1988.

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.

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

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.

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.

KANITZ, S. C. Como prever falências. São Paulo: Abril, 1974.

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.

MERWIN, C. L. Financing Small Corporations in Five Manufacturing Industries, 1926–1936, *National Bureau of Economic Research*, v. 22, 1942.

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

ONUSIC, L.M.; CASA NOVA, S.P.C. A Utilização Conjunta das Técnicas de Análise por Envolvória de Dados e Regressão Logística no Estudo de Insolvência de Empresas: Um Estudo Exploratório. 30º Encontro da ANPAD, 2006.

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.

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.

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

TINOCO, M. H.; WILSON, N. Financial Distress and Bankruptcy Prediction Among Listed Companies using Accounting, Market and Macroeconomic Variables. *International Review of Financial Analysis*, v. 30, p. 394-419, 2013.

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

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.