UNIVERSIDADE DE SÃO PAULO
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE
DEPARTAMENTO DE CONTABILIDADE E ATUÁRIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS CONTÁBEIS

PROBABILIDADE DE INADIMPLÊNCIA DE GRANDES EMPRESAS NO
SISTEMA FINANCEIRO NACIONAL

Simone Rumi Akiama

Orientador: Prof. Dr. Gilberto de Andrade Martins

SÃO PAULO

2008

Profa. Dra. Suely Vilela Ferraz
Reitora da Universidade de São Paulo

Prof. Dr. Carlos Roberto Azzoni
Diretor da Faculdade de Economia, Administração e Contabilidade

Prof. Dr. Fábio Frezatti
Chefe do Departamento de Contabilidade e Atuária

Prof. Dr. Gilberto de Andrade Martins
Coordenador do Programa de Pós-Graduação em Ciências Contábeis

SIMONE RUMI AKIAMA

**PROBABILIDADE DE INADIMPLÊNCIA DE GRANDES EMPRESAS NO
SISTEMA FINANCEIRO NACIONAL**

Dissertação apresentada ao Departamento de Contabilidade e Atuária da Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo como requisito para a obtenção do título de Mestre em Ciências Contábeis.

Orientador: Prof. Dr. Gilberto de Andrade Martins

**SÃO PAULO
2008**

Dissertação defendida e aprovada no Departamento de Contabilidade e Atuária da Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo - Programa de Pós-Graduação em Ciências Contábeis, pela seguinte banca examinadora:

FICHA CATALOGRÁFICA

Elaborada pela Seção de Processamento Técnico do SBD/FEA/USP

Akiama, Simone Rumi

Probabilidade de inadimplência de grandes empresas no Sistema Financeiro Nacional / Simone Rumi Akiama. -- São Paulo, 2008.
140 p.

Dissertação (Mestrado) – Universidade de São Paulo, 2008
Bibliografia

1. Contabilidade financeira 2. Risco 3. Crédito bancário
4. Sistema financeiro – Brasil I. Universidade de São Paulo. Faculdade de Economia, Administração e Contabilidade. II. Título.

CDD – 657.42

**Aos meus pais, Thereza e Catuki,
pela dedicação incondicional aos seus cinco filhos,
e à minha avó Yasa (*in memoriam*),
pelo orgulho que sempre demonstrou pelos seus netos.**

AGRADECIMENTOS

Ao professor Gilberto de Andrade Martins, por ser um orientador acadêmico atuante que sempre incentivou a produção deste trabalho, dando valiosos conselhos e sugestões, com paciência e dedicação.

Aos professores Ariovaldo dos Santos, Gilberto A. Martins, Luiz N. G. de Carvalho, Luiz J. Corrar, Eliseu Martins, José R. Securato, Silvia P. C. Casa Nova, Reinaldo Guerreiro, Geraldo Barbieri, Iran S. Lima, Fábio Frezatti e Luiz P. L. Fávero, pelos ensinamentos e apoio durante o curso.

À Nadia W. H. Vianna e ao Luiz J. S. de Araújo, membros da banca de qualificação, pelos comentários e sugestões que ensejaram o aprimoramento deste trabalho.

Ao Daniel Martins Sanchez, meu orientador técnico no Banco Central do Brasil, pela atenção despendida e pelos conselhos dados durante todo o processo do mestrado.

À FIPECAFI, na pessoa do professor Ariovaldo dos Santos, pelo fornecimento dos dados das 500 Maiores, base para a realização deste estudo.

Ao Banco Central do Brasil pela oportunidade e incentivos para a realização do mestrado, além da disponibilização dos dados do Sistema de Informações de Crédito, fundamentais para o desenvolvimento deste trabalho.

Aos colegas que muito me auxiliaram, através de sugestões, críticas e revisões: Clodoaldo Annibal, Marco Verrone, Giovani Brito, Luis Stancato, Plínio Romanini, Jaime Gregório, Flávio Pereira, Caio Ferreira, Chen Ying Ling, Mitti e Mikio Koyama, Carlos e Janis Morimoto.

A todos os colegas do mestrado e do doutorado e aos funcionários da FEA, com os quais tive o prazer de conviver.

À minha família, por ter dado o apoio necessário durante todo esse período, compreendendo a importância dessa etapa na minha vida.

Aos meus eternos amigos, por entenderem a minha ausência e por estarem sempre prontos para me socorrerem.

RESUMO

O risco de crédito é uma das principais preocupações quando se trata de instituições financeiras. A probabilidade de inadimplência, conhecida também como probabilidade de *default*, tem papel importante na gestão de risco de crédito, auxiliando na constituição de provisões, na precificação das operações de crédito e no estabelecimento de limites de crédito. Com o Novo Acordo de Basiléia, surge a possibilidade de utilização de modelos internos para o cálculo dos componentes que determinam o requerimento mínimo de capital necessário para que a instituição financeira suporte o seu risco de crédito. Um desses componentes é a probabilidade de inadimplência, o que destaca ainda mais a importância de sua mensuração. Deste modo, este trabalho tem como objetivo a construção de um modelo com variáveis contábeis e cadastrais de grandes empresas, juntamente com variáveis macroeconômicas, para estimar a probabilidade de inadimplência dessas empresas no Sistema Financeiro Nacional. Saliente-se que, diferentemente da maioria dos estudos existentes, que utilizam falência e concordata como evento de *default*, a inadimplência no Sistema Financeiro Nacional é pouco abordada, em função da dificuldade na obtenção de dados desse tipo. As fontes utilizadas foram as bases de dados das Melhores e Maiores – As 500 maiores empresas do país (Revista Exame) e do Sistema de Informações de Crédito do Banco Central do Brasil. Como ponto de partida, define-se o evento de *default* como sendo o atraso superior a 90 dias de parcela material da dívida de uma empresa, em relação a uma instituição financeira, e determina-se o percentual que representa o termo parcela material. A técnica estatística utilizada para a modelagem é a regressão logística e adota-se um procedimento manual de seleção de variáveis, que apresentou melhores resultados em termos de qualidade de ajuste se comparado com o procedimento *stepwise*. As variáveis métricas que compõem o modelo final referem-se a indicadores contábeis, índice de inflação, variação do produto interno bruto, tempo de relacionamento com a instituição e tamanho da empresa; sendo que nem todas apresentaram a associação esperada com a probabilidade de inadimplência, possivelmente em função de limitações existentes na base de dados e na metodologia utilizada. As variáveis categóricas que compõem o modelo referem-se ao estado da sede da empresa, seu controle acionário e seu setor econômico. Através de testes estatísticos, verifica-se que o modelo construído apresenta boa qualidade de ajuste aos dados observados, que é importante dado o foco na previsão, e razoável poder discriminante, que representa um bom resultado tendo em vista o baixo percentual existente de eventos de *default* (2,6%). Essa restrição também impossibilitou a validação externa do modelo. Os principais diferenciais do estudo são a base de dados utilizada (dados de atraso de operações de crédito no Sistema Financeiro Nacional) e utilização de informações cadastrais categóricas, o que o torna uma referência potencial para outros estudos sobre previsão de inadimplência bancária, mostrando as diversas dificuldades enfrentadas e apresentando sugestões para aprimoramento.

ABSTRACT

Credit risk is one of the main concerns regarding financial institutions. The probability of default has an important role in credit risk management, constituting provisions, pricing loans, and establishing credit limits. The New Basel Accord allows for the use of internal models to calculate the components that determine the minimum capital requirement in order for the financial institution to support its credit risk. One such component is the probability of default, which further emphasizes the importance of its measurement. Therefore, the objective of this work is the construction of a model using large corporations' accounting and cadastre variables, along with macro economic variables, to estimate the probability of default of these companies in the Brazilian Financial System. Unlike the majority of the existing studies that use bankruptcy and forced agreement as the event of default, past due loans in the Brazilian Financial System are rarely addressed due to the difficulty collecting data of this type. The data used in this work comes from the "Melhores e Maiores – As 500 maiores empresas do país (Revista Exame)" and the "Sistema de Informações de Crédito do Banco Central do Brasil". As a starting point, the event of default is defined as past due more than 90 days on any material credit obligation to the financial institution, and the percentage that represents the term any material is determined. The statistical technique used for the modeling is the logistic regression, and a manual procedure of variable selection is adopted that presents better results in terms of model goodness of fit when compared with the stepwise procedure. The metric variables that compose the final model correspond to accounting ratios, inflation index, variation of the gross domestic product, duration of relationship with the financial institution, and company size. It was observed that some of these variables do not present the expected association with the probability of default, possibly due to limitations in the methodology and data. The categorical variables that compose the model correspond to the state of the company's headquarters, its shareholding control, and its economic sector. Through statistical tests, it was verified that the constructed model appropriately fits the observed data, which is important as forecasting is the focus. The model also presents reasonable discriminating power, which represents a good result, in view of the low percentage of events of default (2.6%). The low percentage of relevant data made it impossible to proceed to external validation. Some of the distinguishable contributions of this work are the use of a different database (past due loans in the Brazilian Financial System) and the use of several categorical cadastre information, that makes this work to be a potential reference for future studies of forecasting default in the financial system, as it describes several difficulties faced, and gives suggestions for further improvements.

SUMÁRIO

LISTA DE ABREVIATURAS E SIGLAS	3
LISTA DE GRÁFICOS	4
LISTA DE TABELAS	5
1 INTRODUÇÃO	7
1.1 Contextualização	7
1.2 Problema de Pesquisa.....	8
1.3 Objetivos da Pesquisa.....	8
1.4 Justificativas	9
1.5 Estrutura do Trabalho.....	10
2 REVISÃO BIBLIOGRÁFICA.....	11
2.1 Intermediação Financeira e Risco de Crédito	11
2.2 O Papel do Banco Central	16
2.3 Resolução 2.682/99, Classificação de Risco e SCR.....	19
2.4 Definição de <i>Default</i>	22
2.5 Probabilidade de <i>Default</i>	26
2.6 Escore de Crédito	30
2.7 Perda Esperada e Perda Não Esperada	32
2.8 Basileia II	36
2.9 Análise das Demonstrações Financeiras	40
2.10 Variáveis Macroeconômicas	45
2.11 Estudos Anteriores	45
2.11.1 Ohlson	46
2.11.2 Casey e Bartczak	48
2.11.3 Westgaard e Wijst	49
2.11.4 Brito.....	51
3 METODOLOGIA	53
3.1 Modelos e Técnicas Estatísticas	55
3.2 Análise Discriminante x Regressão Logística.....	56
3.3 Regressão Logística x Regressão <i>Probit</i>	58
3.4 Regressão Logística.....	59
3.4.1 Descrição do Modelo	61
3.4.2 Pressupostos do Modelo.....	64
3.4.3 Estimção do Modelo.....	65
3.4.4 Avaliação do Ajuste do Modelo.....	67
3.5 População, Amostras e Origem dos Dados	69
3.6 Descrição dos Métodos de Análise dos Dados.....	71
3.6.1 Determinação da Parcela Material da Dívida no Evento de <i>Default</i>	71
3.6.2 Seleção dos Eventos de <i>Default</i> e Não- <i>Default</i>	74
3.6.3 Variáveis Contábeis e Cadastrais	76
3.6.4 Variáveis do SCR	78
3.6.5 Variáveis Macroeconômicas	79
3.6.6 Seleção de Variáveis	79
3.6.7 Limitações das Bases de Dados	81
4 ANÁLISE DOS RESULTADOS.....	83
4.1 Estatísticas Descritivas da População e Amostras	83
4.1.1 Dados das 500 Maiores	83
4.1.2 Dados do SCR	86

4.1.3	Determinação da Parcela Material da Dívida no Evento de <i>Default</i>	87
4.1.4	Dados da Amostra	89
4.2	Construção do Modelo	92
4.3	Interpretação do Modelo	102
4.4	Limitações do Estudo	104
5	CONCLUSÕES.....	107
	REFERÊNCIAS	113
	APÊNDICES	121
	ANEXOS.....	131

LISTA DE ABREVIATURAS E SIGLAS

Basiléia I: Acordo de Basiléia de 1988
 Basiléia II: Novo Acordo de Basiléia
 BCB: Banco Central do Brasil
 BCBS: *Basel Committee on Banking Supervision* (Comitê da Basiléia sobre Supervisão Bancária)
 BIS: *Bank for International Settlements* (Banco de Compensações Internacionais)
 CMN: Conselho Monetário Nacional
 CNPJ: Cadastro Nacional da Pessoa Jurídica
 CRC: Central de Risco de Crédito
 CVM: Comissão de Valores Mobiliários
 DOAR: Demonstração das Origens e Aplicações de Recursos
 EAD: *Exposure at Default* (Valor da exposição no momento do evento de inadimplência)
 EL: *Expected Loss* (Perda esperada)
 FGV: Fundação Getúlio Vargas
 FIECAFI: Fundação Instituto de Pesquisas Contábeis, Atuariais e Financeiras – FEA/USP
 GEE: *Generalized Estimating Equations*
 IBGE: Instituto Brasileiro de Geografia e Estatística
 IF: Instituição Financeira
 IGP-M: Índice Geral de Preços – Mercado
 IRB: *Internal Ratings Based* (Baseadas em classificações internas)
 LGD: *Loss Given Default* (Perdas efetivas em função de um evento de inadimplência)
 PD: *Probability of Default* (Probabilidade de *default* ou probabilidade de inadimplência)
 PF: Pessoa Física
 PIB: Produto Interno Bruto
 PJ: Pessoa Jurídica
 PROER: Programa de Estímulo à Reestruturação e ao Fortalecimento do Sistema Financeiro Nacional
 ROC: *Receiver Operating Characteristic*
 SCR: Sistema de Informações de Crédito do Banco Central
 SELIC: Sistema Especial de Liquidação e de Custódia
 SFN: Sistema Financeiro Nacional
 SPSS: *Statistical Package for Social Science* (Pacote Estatístico para Ciências Sociais)
 UL: *Unexpected Loss* (Perda não Esperada)
 VaR: *Value-at-Risk* (Valor em Risco)
 500 Maiores: as 500 maiores empresas participantes das publicações das Melhores e Maiores
 – As 500 maiores empresas do país, da Revista Exame.

LISTA DE GRÁFICOS

Gráfico 1 – Comportamento das operações de crédito com relação ao PIB, no período de 07/1994 a 12/1995	15
Gráfico 2 - Evolução das operações de crédito no SFN – 12/1994 a 12/2006.....	15
Gráfico 3 - Gráficos representativos da taxa de perdas.....	33
Gráfico 4 – Exemplificativo de distribuição de perdas potenciais.....	34
Gráfico 5 - Logit x Probit.....	58
Gráfico 6 - Função logística	62
Gráfico 7 - Gráfico do Resíduo Padronizado x Registro	96
Gráfico 8 – Estatística de Cook x Registro	97
Gráfico 9 – Curva ROC.....	100

LISTA DE TABELAS

Tabela 1 – Participação das 500 Maiores nas edições de 2003 a 2006.....	83
Tabela 2 – Sede das 500 Maiores nas edições de 2003 a 2006.....	84
Tabela 3 – Setor econômico das 500 Maiores nas edições de 2003 a 2006.....	84
Tabela 4 – Controle acionário das 500 Maiores nas edições de 2003 a 2006.....	85
Tabela 5 – Situação de negociação em bolsa das 500 Maiores nas edições de 2003 a 2006...	85
Tabela 6 – Dados das 500 Maiores no SCR.....	86
Tabela 7 – Percentual do Primeiro Evento de Atraso no Período de 01/2003 a 09/2006	87
Tabela 8 – Matriz de Migração de Classes de Percentual de Atraso	88
Tabela 9 – Tabela Resumo da Migração das Classificações por Atraso.....	88
Tabela 10 – Distribuição dos registros da amostra pelo exercício das demonstrações financeiras	90
Tabela 11 – Distribuição dos registros da amostra em função da sede.....	90
Tabela 12 – Distribuição dos registros da amostra em função do setor econômico	91
Tabela 13 – Distribuição dos registros da amostra em função do controle acionário.....	92
Tabela 14 – Distribuição dos registros da amostra em função da negociação em bolsa.....	92
Tabela 15 – Estatísticas dos modelos testados	94
Tabela 16 – Descrição do modelo final obtido	95
Tabela 17 – Estatísticas resumo	97
Tabela 18 – Estatísticas do teste de Hosmer e Lemeshow	98
Tabela 19 – Tabela de contingência do teste de Hosmer e Lemeshow.....	98
Tabela 20 – Tabela de classificação com valor de corte de 0,5	99
Tabela 21 – Tabela de classificação com valor de corte de 0,028	101
Tabela 22 – Associações esperadas e obtidas das variáveis explicativas contínuas	102

1 INTRODUÇÃO

1.1 Contextualização

Dentre os diversos riscos aos quais uma IF (Instituição Financeira¹) está sujeita, o risco de crédito destaca-se em importância por ser inerente à sua principal função, a intermediação financeira, ou seja, trata-se de um risco diretamente associado ao próprio negócio das IFs.

O risco de crédito pode ser compreendido como sendo a ocorrência de perda após um evento de *default*², ou inadimplência, como denominamos neste trabalho, e pode ser considerado como uma combinação de três elementos: PD (*Probability of Default* - Probabilidade de *default* ou probabilidade de inadimplência), LGD (*Loss Given Default* - Perdas efetivas em função de um evento de inadimplência) e EAD (*Exposure at Default* - Valor da exposição no momento do evento de inadimplência). Deste modo, a PD é uma informação importante para a mensuração do risco de crédito incorrido pelas IFs. Adicionalmente, constitui-se também em informação fundamental para a definição do montante de capital próprio que as IFs devem manter para suportar os riscos a que estão expostas.

Na vigência do Basiléia I (Acordo de Basiléia de 1988), o requerimento de capital referente a operações de crédito foi estabelecido em um percentual fixo³, independentemente da reputação creditícia do cliente e da maturidade e existência de mitigadores de risco das operações de crédito. Apesar de o Basiléia I ser considerado como um grande avanço na direção de um sistema bancário internacional mais sólido e estável, devido à deterioração crescente de sua eficácia, houve a necessidade de reforma do acordo. Assim, foi elaborado o Basiléia II (Novo Acordo de Basiléia). Com ele, as abordagens IRB (*Internal Ratings Based* - Baseadas em classificações internas) possibilitam que as IFs utilizem os seus modelos

¹ A Lei 4.595/64 estabelece:

“Art.17 - Consideram-se instituições financeiras, para os efeitos da legislação em vigor, as pessoas jurídicas públicas ou privadas, que tenham como atividade principal ou acessória a coleta, intermediação ou aplicação de recursos financeiros próprios ou de terceiros, em moeda nacional ou estrangeira, e a custódia de valor de propriedade de terceiros.

Parágrafo único. Para os efeitos desta Lei e da legislação em vigor, equiparam-se às instituições financeiras as pessoas físicas que exerçam qualquer das atividades referidas neste artigo, de forma permanente ou eventual.”

² Segundo Ferreira (2004, p.609), *default* é um termo originalmente jurídico que significa “na falta de”, “na ausência de”; também representa “em dívida” e “inadimplente”.

³ Originariamente de 8%, mas modificado para 11% no Brasil, pela Resolução nº. 2.692 de 24.02.00 (BCB, 2007k).

internos para o cálculo das PDs, que são um dos parâmetros que compõem a fórmula de requerimento de capital para suportar o risco de crédito. Isso permite às IFs maior liberdade para avaliar os riscos de suas operações de crédito, podendo, em princípio, diminuir o seu requerimento de capital e, conseqüentemente, alavancar as suas operações com o capital liberado. Ressalte-se que as instituições deverão satisfazer a supervisão bancária quanto à comprovação da consistência do modelo adotado.

Segundo esse acordo, o evento de *default* ocorre quando a IF considera que o tomador não irá honrar totalmente a dívida e as garantias não são suficientes e/ou quando o tomador apresenta atraso superior a 90 dias, em parcela material de suas operações de crédito.

Deste modo, com base em indicadores contábeis e dados cadastrais das 500 Maiores (as 500 maiores empresas participantes das publicações das Melhores e Maiores – As 500 Maiores empresas do país, da Revista Exame) referentes aos exercícios de 2002 a 2005, e em dados de atraso em operações de crédito no SFN (Sistema Financeiro Nacional), do período de 12/2002 a 12/2006, tempo de relacionamento da empresa na IF, e variáveis macroeconômicas, constrói-se, neste trabalho, um modelo de score de crédito⁴, através da técnica estatística de regressão logística, para estimar a PD de cada empresa.

1.2 Problema de Pesquisa

O problema de pesquisa consiste em construir um modelo com variáveis contábeis e cadastrais de grandes empresas, juntamente com variáveis macroeconômicas, para estimar a PD dessas empresas no SFN.

1.3 Objetivos da Pesquisa

O objetivo geral desta pesquisa é construir um modelo que utilize dados contábeis e cadastrais de grandes empresas, juntamente com variáveis macroeconômicas, para estimar a PD dessas empresas no SFN.

⁴ “Credit scoring”.

Quanto aos objetivos específicos, estes são:

- apresentar a importância da informação da PD para a avaliação de risco de crédito de um cliente no SFN;
- estimar o percentual da dívida a ser considerado como parcela material na identificação do evento de *default*, através dos dados de atraso no SFN;
- selecionar dados cadastrais, indicadores contábeis e variáveis macroeconômicas que possam prever o evento de *default* de grandes empresas no SFN;
- avaliar o ajuste do modelo obtido; e
- interpretar os resultados.

1.4 Justificativas

Há uma grande diversidade de trabalhos que constroem e analisam modelos de previsão de *default*. Na grande maioria dos casos, o evento de *default* é caracterizado por falência ou concordata, que são os dados publicamente disponíveis. Assim, o principal diferencial desta pesquisa é a utilização de dados de atraso de operações de crédito no SFN, provenientes do SCR (Sistema de Informações de Crédito do Banco Central)⁵, para caracterização do evento de *default*. Isso só foi possível em função do compromisso de confidencialidade das informações, assumido com o BCB (Banco Central do Brasil), estabelecendo como finalidade para uso dos dados fornecidos a elaboração deste trabalho acadêmico, sem identificação de nenhuma empresa, nem IF. Salienta-se que compromisso semelhante foi assumido com a FIPECAFI (Fundação Instituto de Pesquisas Contábeis, Atuariais e Financeiras – FEA/USP), com relação à não apresentação de dados individualizados das 500 Maiores.

Deste modo, os elementos que compõem a pesquisa são as grandes empresas, também denominadas *corporate* pelo mercado financeiro, que possuem dívidas no SFN. Assim, a construção de um modelo que estime a PD, com base em dados cadastrais e contábeis de uma empresa contratante, além de variáveis macroeconômicas, pode ser utilizada para precificar

⁵ “O Sistema de Informações de Crédito do Banco Central – SCR é um instrumento de registro e consulta de informações sobre as operações de crédito, avais e fianças prestados e limites de crédito concedidos por instituições financeiras a pessoas físicas e jurídicas no país.” (BCB, 2007m).

suas operações de crédito, constituir provisão para créditos de liquidação duvidosa, além de determinar o capital requerido na IF em função do seu risco de crédito. Esse tipo de modelo tem utilidade tanto para as IFs que têm interesse em desenvolver modelos internos para uma melhor gestão de riscos e, também, para atender ao Basileia II; como para o BCB, órgão regulador e fiscalizador do SFN, para avaliar os modelos das IFs, já que a inadimplência dos tomadores de crédito põe em risco a solidez desse sistema.

1.5 Estrutura do Trabalho

Neste primeiro capítulo, são introduzidos os primeiros conceitos e os objetivos da pesquisa.

No segundo capítulo, as operações de crédito são contextualizadas na atividade de intermediação financeira e discute-se a sua importância na avaliação do risco de uma IF. Apresenta-se a importância da PD na mensuração do risco de crédito e a necessidade da definição do evento de *default*.

No terceiro capítulo, são apresentadas algumas técnicas estatísticas que foram e/ou estão sendo aplicadas nos modelos de previsão de *default*. A regressão logística, a técnica utilizada neste trabalho, é detalhadamente exposta. A população e as amostras consideradas são caracterizadas, tal como as fontes dos dados disponibilizados para o estudo e as potenciais variáveis explicativas selecionadas.

No quarto capítulo, os resultados empíricos são expostos em termos de estatísticas descritivas, do modelo obtido e das variáveis significantes. A qualidade do ajuste do modelo é avaliada e as limitações do modelo são apresentadas.

No quinto e último capítulo, encontram-se as conclusões e considerações finais sobre o trabalho.

2 REVISÃO BIBLIOGRÁFICA

2.1 Intermediação Financeira e Risco de Crédito

Para que se entenda a importância da intermediação financeira, citemos Servigny e Renault (2004, p. 12):

A Teoria Bancária baseia-se na intermediação. A instituição recebe depósitos, principalmente de curto prazo, e empresta tipicamente a longo prazo (empréstimos ilíquidos). Essa atividade de transformação de maturidade é uma função necessária para que se atinja o ótimo econômico global. Os bancos atuam como um intermediário e reduzem as deficiências dos mercados em três áreas: liquidez, risco, e informação.⁶

Assim, percebe-se que a IF está intimamente relacionada com o desenvolvimento econômico de uma sociedade, ao mesmo tempo em que assume um risco típico dessa atividade quando toma recursos, que devem estar disponíveis aos depositantes no vencimento ou quando da necessidade destes (no caso do depósito à vista), e os empresta a clientes que deverão honrar as suas dívidas nos prazos e nos montantes contratados.

Com relação à segunda parte da intermediação financeira, Silva (2000, p. 63) conceitua: “Crédito consiste na entrega de um valor presente mediante uma promessa de pagamento.”

Vicente (2001, p. 39) fala sobre o crédito nas empresas não financeiras, fazendo a distinção em função do tipo de atividade que ele representa na empresa:

O crédito pode ser conceituado nas empresas não financeiras, como uma postergação do recebimento do valor dos serviços prestados e/ou dos produtos vendidos. Nas instituições financeiras, os empréstimos, os financiamentos e todas as extensões a essas operações fazem parte das suas atividades operacionais.

Na mesma linha, Silva (2000, p. 65) comenta:

[...] num banco, crédito é o elemento tradicional na relação cliente-banco, isto é, é o próprio negócio. Numa empresa comercial ou industrial, por exemplo, é possível vender a vista ou a prazo.

⁶ “Banking theory is based on intermediation. The institution receives deposits, mainly short-term ones, and lends money for typically fixed long-term maturities (illiquid loans). This maturity transformation activity is a necessary function for the achievement of global economic optimum. Banks act as an intermediary and reduce the deficiencies of markets in three areas: liquidity, risk, and information.”

Num banco, não há como fazer um empréstimo ou financiamento a vista. A principal fonte de receita de um banco deve ser proveniente de sua atividade de intermediação.

Assim, dada a sua relevância, deve-se avaliar o risco que as operações de crédito representam para as IFs. Mas, o que significa risco? Securato (1996, p. 28) define risco como sendo “a probabilidade de ocorrência do evento gerador da perda ou da incerteza”. Quanto à sua origem, Jorion (2003, p. 8) expõe:

Os riscos originam-se de várias fontes. Podem ser criados pelos seres humanos, como por exemplo os ciclos de negócios, a inflação, as mudanças das políticas de governo e as guerras. O risco também provém de fenômenos naturais imprevisíveis, tais como o clima e os terremotos, ou resulta das principais fontes de crescimento econômico de longo prazo. É o caso das inovações tecnológicas que podem tornar a tecnologia existente obsoleta e criar deslocamentos de emprego. Portanto, o risco e a vontade de assumi-lo são essenciais para o crescimento da economia.

Com a combinação dos conceitos de risco e de crédito, temos o risco de crédito. Saunders (2000, p. 102), no âmbito das IFs, explica: “[...] há risco de crédito porque os fluxos de caixa prometidos pelos títulos primários possuídos por IFs podem não ser pagos integralmente.” De forma similar, Assaf Neto (2005, p. 97) conceitua risco de crédito:

[...] é a possibilidade de uma instituição financeira não receber os valores (principal e rendimentos de juros) prometidos pelos títulos que mantém em sua carteira de recebíveis. Como exemplos desses ativos apontam-se principalmente os créditos concedidos pelos bancos e os títulos de renda fixa emitidos pelos devedores.

Jorion (2003, p. 15) acrescenta o aspecto da intenção do devedor: “O *risco de crédito* surge quando as contrapartes não desejam ou não são capazes de cumprir suas obrigações contratuais.” Bessis (1998, p. 81), de uma forma mais detalhada, explica:

Risco de crédito é definido pelas perdas no evento de *default* do tomador, ou no evento da deterioração da qualidade do crédito do tomador. Essa definição simples esconde vários riscos subjacentes. A “quantidade” de risco resulta tanto das chances de o *default* ocorrer como das garantias que reduzem a perda no evento de *default*. O montante em risco, o saldo da dívida na data de *default*, difere da perda no evento de *default* por causa das recuperações potenciais. Essas dependem de qualquer mitigador de risco de crédito, como as garantias, tanto garantias reais como garantidores, a capacidade de negociar com o tomador, e os fundos disponíveis, se algum, para pagar a dívida após pagar outros credores.

Default é um evento incerto. Além disso, a exposição futura na época do *default* não é conhecida antecipadamente na maioria dos casos. Isso ocorre porque o fluxo de pagamentos dos empréstimos é contratual em apenas um número limitado de casos. Finalmente, recuperações potenciais a partir do *default* não podem ser previstas antecipadamente. Consequentemente, o risco de crédito pode ser dividido em três riscos: risco de *default*, risco de exposição e risco de recuperação.⁷

⁷ “Credit risk is defined by the losses in the event of default of the borrower, or in the event of a deterioration of the borrower’s credit quality. This simple definition hides several underlying risks. The ‘quantity’ of risk results

Note-se que podemos relacionar os três riscos definidos pelo autor com as siglas conhecidas: risco de *default* (PD), risco de exposição (EAD) e risco de recuperação (1-LGD).

De forma mais concisa e já expressando a dificuldade na sua mensuração: “[...] o risco de crédito, o mais antigo dos riscos para os bancos, é realmente o resultado final de riscos multidimensionais. Soa como um paradoxo que o mais familiar de todos os riscos permaneça tão difícil de quantificar.”⁸ (BESSIS, 1998, p. 6).

Seguindo a mesma linha, Cossin e Pirotte (2000, p. 1) acrescentam:

Risco de crédito, ou risco de *default*, tem sido sempre o principal objeto de preocupação para bancos e outros intermediários financeiros, e qualquer agente comprometido com um contrato financeiro no que diz respeito ao assunto. Enquanto a preocupação com o possível *default* de uma contraparte de um contrato financeiro pactuado é centenário, técnicas modernas e modelos que surgiram nos últimos anos ajudam a aumentar o problema.⁹

Deste modo, pode-se perceber que o risco de crédito é uma das grandes preocupações quando se pensa na saúde de uma IF ou do próprio SFN. Nesta linha, Saunders (2000, p. 195) afirma que:

[...] os problemas de qualidade de crédito, na pior das hipóteses, podem levar um IF à insolvência. Ou podem resultar em perdas significativas de capital e patrimônio líquido a ponto de prejudicar as perspectivas de crescimento e competição de um IF, tanto em nível doméstico quanto em nível internacional.

No que se refere à saúde do SFN, tratando especificamente do Brasil, merecem destaque os efeitos do Plano Real, que, modificando o ambiente econômico, acabou por alterar a forma de atuar das IFs, inclusive na área de crédito. Tais mudanças geraram dificuldades para algumas

both from the chances that the default occurs and from the guarantees that reduce the loss in the event of default. The amount at risk, the outstanding balance at the date of default, differs from loss in the event of default because of potential recoveries. Those depend upon any credit mitigators, such as guarantees, either collateral or third-party guarantees, the capability of negotiating with the borrower, and the funds available, if any, to repay the debt after repayment of other lenders.

Default is an uncertain event. In addition, the future exposures at the time of default are not known in advance in many cases. This is because the repayment schedule of loans is contractual only in a limited number of cases. Finally, potential recoveries from default cannot be predicted in advance. Hence, credit risk can be divided into three risks: default risk, exposure risk and recovery risk.”

⁸ *“In short, credit risk, the oldest of all risks for banks, is actually the end result of multidimensional risks. It sounds like a paradox that the most familiar of all risks remains so difficult to quantify.”*

⁹ *“Credit risk, or the risk of default, has always been a major topic of concern for banks and other financial intermediaries, and any agent committed to a financial contract for that matter. While the concern for the possible default of a counterparty on a agreed-upon financial contract is centuries old, modern techniques and models have arisen in the last few years that help to master the problem.”*

IFs, levando o governo a adotar medidas com o fim de prevenir uma crise de maiores proporções.

No *site* do BCB (2007g), no documento que contextualiza uma das mais importantes dessas medidas, o PROER (Programa de Estímulo à Reestruturação e ao Fortalecimento do Sistema Financeiro Nacional), temos uma boa descrição do ocorrido:

O longo período de convivência com o processo inflacionário permitiu que ganhos proporcionados pelos passivos não remunerados, como os depósitos à vista e os recursos em trânsito, compensassem ineficiências administrativas e, até mesmo, concessões de crédito de liquidação duvidosa.

Diante do novo quadro de estabilidade de preços, desejado pela sociedade brasileira após várias tentativas frustradas, verificou-se uma total incapacidade de nossas instituições financeiras em promover espontaneamente os ajustes necessários para sua sobrevivência nesse novo ambiente econômico.

Santos (2000, p. 21) comenta o mesmo fato:

Após a implantação do Plano Real, em Junho de 1994, a tarifa bancária de intermediação financeira teve que se ajustar a uma nova realidade econômica. A significativa queda inflacionária impactou na necessidade dos bancos de realocar seus ativos para investimentos alternativos geradores de receitas operacionais, haja vista a redução de receita com investimentos de curto prazo criados para eliminar ou reduzir as perdas causadas pela inflação. A consequência imediata foi o aumento da concessão de empréstimos, em parte impulsionado pela possibilidade de obtenção de taxas de retorno atrativas (*spreads*).

Isso pode ser constatado no Gráfico 1, que apresenta o comportamento da relação Crédito/PIB (Produto Interno Bruto), logo após o evento do Plano Real:

Gráfico 1 – Comportamento das operações de crédito com relação ao PIB, no período de 07/1994 a 12/1995

FONTE: Séries Temporais do Departamento Econômico do BCB (BCB, 2007I)

Mas, no final de 1994, já se percebe a mudança de comportamento de crescimento para queda da representatividade das operações de crédito em relação ao PIB. Em um horizonte mais longo, verifica-se que há a retomada do crescimento do indicador a partir do início de 2003, conforme mostra o Gráfico 2, que também apresenta as participações das carteiras de pessoas físicas (PF) e pessoas jurídicas (PJ), separadamente:

Gráfico 2 - Evolução das operações de crédito no SFN – 12/1994 a 12/2006

FONTE: Séries Temporais do Departamento Econômico do BCB (BCB, 2007I)

Note-se que, enquanto a carteira de pessoas físicas aumenta a sua participação¹⁰ nas operações de crédito, a carteira de pessoas jurídicas sofre uma diminuição. Uma das possíveis explicações para essa diminuição é a de que, com relação a grandes empresas, ocorre em todo o mundo um processo de desintermediação financeira¹¹, ou seja, devido ao custo inferior, as empresas de grande porte estão deixando de captar recursos no SFN para captá-los no mercado de capitais¹². Com isso, as IFs tendem a ter as suas carteiras de crédito de pessoas jurídicas diminuídas e/ou deterioradas, pois passam a lidar com empresas menores e de maior risco. Além disso, algumas das grandes empresas que continuam contratando operações de crédito com IFs podem ser aquelas que não conseguem colocar os seus papéis no mercado de capitais por não possuírem boa reputação no mercado e, conseqüentemente, representam riscos de grande impacto devido aos altos montantes de suas dívidas.

Assim, verificamos que a correta mensuração do risco de crédito, especialmente das grandes empresas, torna-se de fundamental importância para a manutenção da solidez do SFN.

Ainda, conforme exposto por Carling *et al.* (2007, p. 865), os avanços na tecnologia de sistemas, as conseqüências das crises asiáticas nos anos 90, e os subseqüentes trabalhos de revisão do Basileia I renovaram o interesse pela pesquisa sobre *default* e risco de crédito.

2.2 O Papel do Banco Central

Para entender as razões pelas quais as atividades das IFs devem ser reguladas e acompanhadas por uma autoridade superior, analisemos a colocação de Bessis (1998, p. 39):

¹⁰ Parte desse aumento pode ser atribuída à regulamentação do crédito consignado para trabalhadores do setor privado e para aposentados e pensionistas do INSS que entrou em vigor em 17/09/2003, com a Medida Provisória 130, posteriormente promulgada pela Lei 10.820, de 17/12/2003, segundo o Relatório de Economia Bancária e Crédito de 2005 do BCB (BCB, 2008b, p. 97).

¹¹ Segundo Lima *et al.* (2006, p. 4): “Em virtude da mera interveniência da instituição financeira, costuma-se identificar o mercado de capitais com o conceito de desintermediação financeira.”

¹² “As operações são normalmente efetuadas diretamente entre poupadores e tomadores, de modo que a instituição financeira não atua, em regra, como parte na operação, mas como interveniente, e cobra uma comissão por facilitar a realização dos negócios.” (LIMA *et al.*, 2006, p. 4).

A origem da regulação recai sobre as diferenças entre os objetivos dos bancos e aqueles das autoridades reguladoras. Existem inúmeros incentivos para os bancos assumirem riscos. Mas, a assunção de riscos cria o “risco sistêmico”, o risco de todo o sistema bancário quebrar. Isso ocorre porque os bancos são altamente inter-relacionados através de comprometimentos mutuamente imbricados. Conseqüentemente, a quebra de uma instituição gera um risco de quebra para aqueles outros bancos que possuem fundos comprometidos com o banco que quebrou. Risco sistêmico é o principal desafio para o regulador. Instituições individuais estão mais preocupadas com o seu próprio risco. O regulador está mais preocupado com o risco de todo o sistema.¹³

Como exemplo desse tão temido risco sistêmico, podemos citar o caso da Venezuela¹⁴, quando houve uma grande corrida bancária e a necessidade de aporte de recursos por parte do governo.

Com relação ao arcabouço regulatório das finanças, Canuto e Lima (2002, p. 221) fazem a distinção de dois itens básicos: as redes de segurança financeira¹⁵ e os sistemas de supervisão e regulação. Segundo eles:

[...] a supervisão e regulação têm como principal objetivo contrarrestar as assimetrias de informação intrínsecas às atividades financeiras e os decorrentes problemas associados à seleção adversa e ao risco moral (*moral hazard*), inclusive a potencialização destes pela própria presença das redes de segurança financeira.

Para situar os sistemas de regulação e supervisão no SFN, tomemos a definição de Silva (2000, p. 42): “O Sistema Financeiro Nacional consiste em um conjunto de instrumentos e instituições que funcionam como meio para realização da intermediação financeira.” No Quadro 1, temos a composição do SFN:

¹³ “The source of regulation lies in the differences between the objectives of banks and those of the regulatory authorities. Numerous incentives for taking risks exist for banks. But taking risk creates ‘systemic risk’, the risk that the whole banking system fails. This is because banks are highly interrelates with mutual imbricated commitments. Hence, the failure of one institution generates a risk of failure for those other banks which have committed funds with the defaulting bank. Systemic risk is a major challenge for the regulator. Individual institutions are more concerned with their own risk. The regulator is more concerned by the risk of the whole system.”

¹⁴ Segundo Capelletto (2006, p. 86), “A crise financeira foi deflagrada em janeiro de 1994, quando o presidente recém-empossado, em face da decretação do *impeachment* do presidente anterior, determinou a intervenção no Banco Latino, o segundo maior banco comercial venezuelano, sendo o primeiro em número de correntistas. [...] O fechamento disseminou o medo de perder os valores depositados pelos correntistas e depositantes, provocando a maior corrida bancária noticiada na América Latina. Em apenas quinze dias, o equivalente a 5% do PIB venezuelano foi sacado dos bancos. A consequência imediata foi o fechamento de oito bancos por falta de fundos. [...] o governo foi obrigado a assumir o controle dos bancos quebrados e realizar o aporte de seis bilhões de dólares.”

¹⁵ Seguros de depósitos ou outras aplicações, empréstimos em última instância.

Quadro 1 - Composição do SFN

Órgãos normativos	Entidades supervisoras	Operadores		
<u>Conselho Monetário Nacional - CMN</u>	<u>Banco Central do Brasil - Bacen</u>	<u>Instituições financeiras captadoras de depósitos à vista</u>	<u>Demais instituições financeiras</u>	<u>Outros intermediários financeiros e administradores de recursos de terceiros</u>
	<u>Comissão de Valores Mobiliários - CVM</u>	<u>Bolsas de mercadorias e futuros</u>	<u>Bolsas de valores</u>	
<u>Conselho Nacional de Seguros Privados - CNSP</u>	<u>Superintendência de Seguros Privados - Susep</u>	<u>Sociedades seguradoras</u>	<u>Sociedades de capitalização</u>	<u>Entidades abertas de previdência complementar</u>
	<u>IRB-Brasil Resseguros</u>			
<u>Conselho de Gestão da Previdência Complementar - CGPC</u>	<u>Secretaria de Previdência Complementar - SPC</u>	<u>Entidades fechadas de previdência complementar (fundos de pensão)</u>		

FONTE: BCB, 2007c

Deste modo, com relação ao mercado de crédito¹⁶, que seria o foco de atuação das IFs, o órgão normativo é o CMN (Conselho Monetário Nacional), assim descrito no *site* do BCB (2007d, grifos nossos):

O Conselho Monetário Nacional (CMN), que foi instituído pela Lei 4.595, de 31 de dezembro de 1964, é o órgão responsável por expedir diretrizes gerais para o bom funcionamento do SFN. Integram o CMN o Ministro da Fazenda (Presidente), o Ministro do Planejamento, Orçamento e Gestão e o Presidente do Banco Central do Brasil. Dentre suas funções estão: adaptar o volume dos meios de pagamento às reais necessidades da economia; regular o valor interno e externo da moeda e o equilíbrio do balanço de pagamentos; orientar a aplicação dos recursos das instituições financeiras; propiciar o aperfeiçoamento das instituições e dos instrumentos financeiros; zelar pela liquidez e solvência das instituições financeiras; coordenar as políticas monetária, creditícia, orçamentária e da dívida pública interna e externa.

¹⁶ “Neste mercado, [...] realiza-se a atividade bancária por excelência – a **intermediação financeira**, e as instituições financeiras podem cumprir a importante função social e econômica de otimizar a utilização de recursos financeiros, alocando a poupança popular de forma eficiente, de modo a gerar desenvolvimento.” (LIMA *et al.*, 2006, p. 3, grifos do original).

E, como executor das orientações do CMN, temos o BCB, assim descrito pela mesma fonte (BCB, 2007a, grifos nossos):

O Banco Central do Brasil (Bacen) é uma autarquia vinculada ao Ministério da Fazenda, que também foi criada pela Lei 4.595, de 31 de dezembro de 1964. É o principal executor das orientações do Conselho Monetário Nacional e responsável por garantir o poder de compra da moeda nacional, tendo por objetivos: zelar pela adequada liquidez da economia; manter as reservas internacionais em nível adequado; estimular a formação de poupança; zelar pela estabilidade e promover o permanente aperfeiçoamento do sistema financeiro.

Segundo o Regimento Interno do BCB (2007h, grifos nossos):

Art. 2º. O Banco Central tem por finalidade a formulação, a execução, o acompanhamento e o controle das políticas monetária, cambial, de crédito e de relações financeiras com o exterior; a organização, disciplina e fiscalização do Sistema Financeiro Nacional; a gestão do Sistema de Pagamentos Brasileiro e dos serviços do meio circulante.

Ainda, podemos citar o Planejamento Estratégico do BCB (BCB, 2007f) que apresenta, como uma das orientações estratégicas vigentes¹⁷, a sua missão: “Assegurar a estabilidade do poder de compra da moeda e um sistema financeiro sólido e eficiente.”

2.3 Resolução 2.682/99, Classificação de Risco e SCR

Na função de regular o SFN, o CMN, através da Resolução 2.682/99¹⁸, determinou que as IFs passassem a classificar suas operações de crédito em níveis de risco (AA, B, C, D, E, F, G e H – em ordem crescente de risco), com critério de classificação mínimo para o atraso das operações, e a constituir um percentual de provisão mínima para cada nível de risco. Essa foi uma grande mudança em relação à Resolução 1.748 de 30.08.1990 (BCB, 2007i), que regulava o tratamento das provisões para créditos de liquidação duvidosa¹⁹ até então.

As regras de provisionamento da Resolução 1.748 baseavam-se essencialmente na ocorrência de atrasos e na suficiência das garantias para a cobertura das operações, desconsiderando o risco de crédito potencial do tomador de recursos.

¹⁷ Definidas em fevereiro de 2007.

¹⁸ Ver o Anexo A.

¹⁹ Iudícibus *et al.* (2003, p. 97) fazem a seguinte observação com relação a um outro termo já utilizado: “O nome Provisão para Devedores Duvidosos não é o mais adequado, uma vez que a dúvida não recai sobre quem é o devedor, mas se este honrará sua dívida, portanto, entendemos que a rubrica mais pertinente seja Provisão para Créditos de Liquidação Duvidosa.”

Segundo Iudícibus *et al.* (2003, p. 98), quando expõem sobre a forma de apuração da provisão para créditos de liquidação duvidosa:

Como se verifica, temos como prática comum e adequada:

- a) determinar o valor das perdas já conhecidas com base nos clientes atrasados, em concordata, falência ou com dificuldades financeiras;
- b) estabelecer um valor adicional de provisão para cobrir perdas prováveis, mesmo que ainda não conhecidas por se referirem a contas a vencer, mas comuns de ocorrer, com base na experiência da empresa, tipo de clientes, etc.

A Resolução 1.748 atendia apenas ao item “a”, não considerando o aspecto estimativo da provisão (item “b”). Nessa linha, Parente (2000, p. 3) afirma:

[...] como a determinação de provisionar estava vinculada a um atraso nas operações, o ambiente regulatório anterior acabava por não abordar o caráter estimativo previsto na definição. Guardar relação direta com o atraso dos pagamentos acabou por não abranger a totalidade das possíveis origens do risco nas atividades de crédito. De fato, o atraso é mais que um indicativo de risco elevado; ele é também o sintoma que antecede uma perda efetiva. A vinculação da constituição de provisões apenas a partir do não pagamento, desobrigou o caráter prospectivo que as provisões por definição devem ter.

Iudícibus *et al.* (2003, p. 98) comentam sobre as IFs e a Resolução 2.682/99 (grifos nossos):

As instituições financeiras são as entidades que possuem maior exposição ao risco de crédito por causa de suas atividades operacionais. A Resolução no. 2.682/99 do Banco Central do Brasil (BACEN), que dispõe sobre critérios de classificação das operações de crédito e regras para constituição de provisão para créditos de liquidação duvidosa, é uma boa fonte de informações de como se analisar o risco de crédito.

Esta resolução, quando se refere à classificação da operação no nível de risco correspondente (art. 2º.), exige que a IF contemple, em seu modelo, pelo menos alguns aspectos relativos ao devedor e seus garantidores, como também em relação à operação de crédito, conforme enumerado:

I - em relação ao devedor e seus garantidores:

- a) situação econômico-financeira;
- b) grau de endividamento;
- c) capacidade de geração de resultados;
- d) fluxo de caixa;
- e) administração e qualidade de controles;
- f) pontualidade e atrasos nos pagamentos;
- g) contingências;
- h) setor de atividade econômica;
- i) limite de crédito.

II - em relação à operação:

- a) natureza e finalidade da transação;
- b) características das garantias, particularmente quanto à suficiência e liquidez;
- c) valor.

Bessis (1998, p. 83) apresenta a possibilidade de se separar esses riscos: “Os sistemas de classificação internos do banco podem ser utilizados para classificar separadamente o cliente e os vários produtos mantidos pelo cliente. Isso permite que se separe o risco de *default* do tomador, do risco de recuperação associado com cada produto.”²⁰

De acordo com a tipologia apresentada por Silva (2000, p. 75):

Os riscos de crédito de um banco (*bank credit risk*) podem ser classificados em quatro grupos: (i) risco do cliente ou risco intrínseco (*intrinsic risk*); (ii) risco da operação (*transaction risk*); (iii) risco de concentração (*concentration risk*); e (iv) risco da administração do crédito (*credit management risk*).

Essa tipologia também separa o risco do cliente do risco da operação, tal como Bessis e a própria Resolução 2.682/99 o fazem. Além disso, a Resolução 2.682/99 estipula um critério mínimo de classificação de risco das operações de crédito por atraso:

- a) atraso entre 15 e 30 dias: risco nível B, no mínimo;
- b) atraso entre 31 e 60 dias: risco nível C, no mínimo;
- c) atraso entre 61 e 90 dias: risco nível D, no mínimo;
- d) atraso entre 91 e 120 dias: risco nível E, no mínimo;
- e) atraso entre 121 e 150 dias: risco nível F, no mínimo;
- f) atraso entre 151 e 180 dias: risco nível G, no mínimo;
- g) atraso superior a 180 dias: risco nível H.

Tais informações compõem a base de dados de operações de crédito do SFN, administrada pelo BCB. Para se avaliar como está estruturado esse sistema de informações de operações de crédito, o SCR, que é uma das fontes de dados para a nossa pesquisa, apresentemos um breve relato sobre ele.

²⁰ “In-house bank rating systems can be used to rate separately the customer and the various facilities existing with the customer. This allow one to separate the default risk of the borrower from the recovery risk associated with each facility.”

O sistema CRC (Central de Risco de Crédito) foi implementado a partir da Resolução 2.390/97, posteriormente substituída pela Resolução 2.724/00, através da determinação do envio, pelas IFs, de informações sobre o montante dos débitos e responsabilidades por garantias de seus clientes, ao BCB. Em função da necessidade de evolução do sistema, iniciaram-se os estudos que levaram à construção do SCR, sucessor do CRC, em 2000.

No novo sistema, ampliou-se o escopo das informações existentes, com o objetivo de atender, além das necessidades da área de supervisão bancária, as necessidades de outras áreas de atuação do BCB, como também fornecer subsídios às IFs para o aperfeiçoamento da gestão de suas carteiras de crédito.

Analizando a Circular 3.098 de 20.03.2002 (BCB, 2007b), que deu início à coleta das informações do SCR, e o Leiaute do Documento 3020 (BCB, 2007e), que trata dos dados individualizados de risco de crédito, verifica-se a obrigatoriedade da informação da classificação de risco do devedor para os clientes relevantes²¹; para os demais casos, somente as operações de crédito precisam ser classificadas, levando em consideração a avaliação conjunta do risco do cliente e do risco da operação e também os prazos de atraso anteriormente mencionados.

2.4 Definição de *Default*

Retomando o assunto risco de crédito, nota-se a importância da definição do evento de *default* para que se possa mensurar este tipo de risco, como se verifica na afirmação de Westgaard e Wijst (2001, p. 339): “[...] risco de crédito é o risco de um tomador/contraparte entrar em *default*”²². No caso deles, define-se que: “[...] entrar em *default* é fracassar em pagar uma quantia devida a um banco.”²³

²¹ Clientes que possuem operação de crédito com valor acima de R\$ 5 milhões.

²² “[...] *credit risk is the risk that the borrower/counterparty will default.*”

²³ “[...] *default is fail to repay an amount owed to the bank.*”

Segundo Fitch (2000, p. 134), *default* significa “[...] fracasso em cumprir uma obrigação contratual, como o pagamento de um empréstimo pelo devedor, ou pagamento de juros aos detentores de títulos.”²⁴

Altman (1993, p. 3) levanta a complexidade do assunto, apresentando quatro termos que podem caracterizar uma empresa “com problemas”²⁵:

O insucesso de uma empresa tem sido definido de muitas maneiras na tentativa de descrever o processo formal enfrentado pela firma e/ou para caracterizar os problemas econômicos envolvidos. Quatro termos genéricos que são comumente encontrados na literatura são fracasso, insolvência, *default* e falência. Embora esses termos sejam algumas vezes utilizados indistintamente, eles são diferentes na sua utilização formal.²⁶

Com isso, é possível entender porque, em vários trabalhos pesquisados, utilizam-se dos mesmos termos com diferentes significados e vice-versa. Nessa obra, o autor descreve o *default* técnico como ocorrido quando o devedor viola uma condição do acordo assumido com o credor, podendo estar sujeito a uma ação legal. O *default* formal ocorre quando o devedor não efetua um pagamento contratual (juros e/ou principal).

Servigny e Renault (2004, p. 119) dizem que antes de uma empresa estar formalmente em *default*, se estiver enfrentando dificuldades com suas dívidas, ela será considerada “com problemas”. No entanto, o banco pode não ser capaz de identificar esse estado, e o *default* pode ocorrer diretamente, sem essa fase intermediária.

Note-se que há diferenças no entendimento sobre o estado “com problemas” nos dois casos citados: para Altman, seria uma definição geral, que incluiria o *default*; para Servigny e Renault, seria uma fase intermediária, que antecederia o *default*.

Bessis (1998, p. 82) apresenta várias definições possíveis de *default*: “[...] deixar de pagar uma obrigação, quebrar um acordo, entrar em um procedimento legal, ou *default*

²⁴ “[...] failure to meet a contractual obligation, such as repayment of a loan by a borrower or payment of interest to bond holders.”

²⁵ O termo utilizado é *distress*.

²⁶ “The unsuccessful business enterprise has been defined in numerous ways in attempts to depict the formal process confronting the firm and/or to categorize the economic problems involved. Four generic terms that are commonly found in the literature are failure, insolvency, default and bankruptcy. Although these terms are sometimes used interchangeably, they are distinctly different in their formal usage.”

econômico.”²⁷ Segundo ele, as agências classificadoras de risco consideram ocorrência de *default* quando algum pagamento contratual não foi cumprido por pelo menos três meses, e salienta que o evento do *default* não provoca perdas necessariamente, mas aumenta a chance do *default* final, que é a falência. Ainda, comenta: “A definição de *default* é importante para a estimativa das chances de *default*, por exemplo, através de registros históricos.”²⁸

Segundo o BCBS (*Basel Committee on Banking Supervision* - Comitê da Basileia sobre Supervisão Bancária) (2006, p. 100):

Considera-se que o *default* tenha ocorrido com relação a um devedor específico quando um ou ambos os eventos seguintes tiverem acontecido:

- O banco considera improvável que o devedor pague as suas obrigações ao conglomerado bancário na totalidade, sem que o banco tenha que recorrer a ações, tais como realizar garantias (caso possua).
- O devedor está atrasado em mais de 90 dias em alguma obrigação material com o conglomerado bancário. Saques a descoberto serão considerados como operações em atraso quando o cliente infringir um limite recomendado ou tenha lhe sido recomendado um limite menor do que a dívida atual.²⁹

Com relação ao atraso acima citado, é apresentada a seguinte observação:

No caso de obrigações do varejo e das entidades não governamentais do setor público, com relação ao número de 90 dias, um supervisor pode substituir este número para até 180 dias para produtos diferentes, se ele considerar apropriado para as condições locais. Em um país membro, condições locais tornam apropriado utilizar um número de até 180 dias também para empréstimos concedidos pelos seus bancos para empresas; isso é aplicável no período de transição de 5 anos.³⁰ (BCBS, 2006, p. 100)

Em um levantamento feito por Altman e Narayanan (1997) sobre os estudos de identificação e previsão de fracasso de empresas, eles verificaram que a maioria dos modelos utiliza dois grupos de amostras: empresas fracassadas e empresas “mais saudáveis”. Segundo os mesmos autores, a definição de fracasso pode variar dependendo da inclinação do pesquisador ou das

²⁷ “[...] missing a payment obligation, breaking a covenant, entering a legal procedure, or economic default.”

²⁸ “The definition of default is important in estimating the chances of default, for instance from historical records.”

²⁹ “A default is considered to have occurred with regard to a particular obligor when either or both of the following events have taken place:

- The bank considers that the obligor is unlikely to pay its credit obligations to the banking group in full, without recourse by the bank to actions such as realising security (if held).
- The obligor is past due more than 90 days on any material credit obligation to the banking group. Overdrafts will be considered as being past due once the customer has breached an advised limit or been advised of a limit smaller than current outstanding.”

³⁰ “In the case of retail and PSE obligations, for the 90 days figure, a supervisor may substitute a figure up to 180 days for different products, as it considers appropriate to local conditions. In one member country, local conditions make it appropriate to use a figure of up to 180 days also for lending by its banks to corporates; this applies for a transitional period of 5 years.”

condições locais: “Fracasso poderia significar falência sofrida por uma companhia, *default* de título, *default* de empréstimo bancário, remoção de uma empresa da lista oficial da bolsa, intervenção governamental através de financiamento especial, e liquidação.”³¹ (1997, p. 2).

Servigny e Renault (2004, p. 119) expõem, quando se referem a empréstimos bancários, que ainda há muita incerteza quanto à definição de *default* no mercado financeiro. Eles apresentam as seguintes considerações:

- A definição do mercado para *default* está relacionada aos instrumentos financeiros. Corresponde ao atraso do principal ou dos juros.
- A definição do Basileia II considera um evento de *default* com base em diferentes opções alternativas, como atraso de 90 dias em instrumentos financeiros ou provisionamento. Pode também ser baseado em uma avaliação julgamental de uma firma pelo banco.
- A definição legal está relacionada com a falência da firma. Isso dependerá tipicamente da legislação nos diferentes países.³²

Eles observam que a dificuldade de se entender o *default* provém do fato de não ser uma consequência lógica de um único e bem definido processo. Um outro aspecto é a não convergência de interesses dos envolvidos sobre o momento em que se deve declarar o *default*. Por exemplo, os empresários tentarão comandar a empresa o máximo possível, mas, para os bancos, o ponto ótimo do estabelecimento do *default* poderia ser a data em que o valor da garantia fosse equivalente ao valor do empréstimo acrescido dos custos de realização da garantia.

Ainda, Sicsú (2003, p. 330) relata:

A dificuldade surge na definição do conceito de inadimplência para um determinado produto. Alcançar o consenso entre os analistas de crédito de uma instituição tem-se mostrado uma árdua tarefa na prática. Além da dificuldade natural ao definir inadimplência, os objetivos dos analistas envolvidos na definição operacional de inadimplência podem ser conflitantes. Alguns analistas adotarão definições extremamente rigorosas, objetivando que o modelo de escoragem aprove o crédito de forma parcimoniosa. Outros analistas, preocupados com a possibilidade de gerar um sistema muito conservador, que limite os negócios da instituição financeira, procurarão uma conceituação menos restritiva de inadimplência.

³¹ “Failure could mean bankruptcy filing by a company, bond default, bank loan default, delisting of a company, government intervention via special financing, and liquidation.”

³² “• The market definition for default is related to financial instruments. It corresponds to principal or interest past due.
 • The Basel II definition considers a default event based on various alternative options such as past due 90 days on financial instruments or provisioning. It can also be based on a judgmental assessment of a firm by the bank.
 • The legal definition is linked with the bankruptcy of the firm. It will typically depend on the legislation in various countries.”

Na tese defendida por Veiga (2006, p. 33), o termo *default* foi utilizado como sinônimo de inadimplência³³.

Verrone (2007, p. 119) apresenta a preocupação de algumas IFs quanto à definição do prazo a ser considerado nos seus modelos internos para a caracterização da inadimplência, que poderia ser de 90 dias, conforme consta em Basiléia II, ou de 60 dias, conforme a prática mais comum de mercado.

Neste trabalho, considera-se o termo *default* como sinônimo de inadimplência, e adota-se, como definição do evento de *default*, que caracteriza a variável dependente da regressão logística, o atraso superior a 90 dias de parcela material da dívida de uma empresa, em relação a uma IF.

2.5 Probabilidade de *Default*

Com o evento de *default* definido, é possível “medir” a variável dependente da regressão, que é a PD.

Segundo Bessis (1998, p. 83):

A probabilidade de *default* não pode ser medida diretamente. Estatísticas históricas de *default* podem ser utilizadas. Tais dados podem ser coletados internamente ou obtidos de agências de *rating* ou de autoridades centrais. A partir das estatísticas dos *defaults* observados, a proporção de *defaults* em um dado período sobre a amostra total de tomadores pode ser obtida. É a taxa de *default*, que freqüentemente serve como uma representação da probabilidade de *default*. Tais proporções ou freqüências de *default* são disponibilizadas por indústria ou por classe de *rating*. A deficiência de tais dados históricos é que eles não capturam as probabilidades esperadas de *default*.³⁴

³³ “Conquanto possa haver alguma diferença entre esses conceitos, este trabalho utilizará o termo consagrado *default* como inadimplência.”

³⁴ “*Default probability cannot be measured directly. Historical statistics of defaults can be used. Such data can be collected internally or obtained from rating agencies or from central authorities. From statistics of observed defaults, the ratio of defaults in a given period over the total sample of borrowers can be derived. It is a default rate, which often serve as an historical proxy for default probability. Such ratios or default frequencies are available by industry or by rating class. The shortcoming of such historical data is that they do not capture expected default probabilities.*”

Nesse sentido, Schrickel (2000, p. 35) expõe:

[...] embora a análise de crédito deva lidar com eventos passados do tomador de empréstimos (a análise histórica), as decisões de crédito devem considerar primordialmente o futuro desse mesmo tomador. “O risco situa-se no futuro; no passado, encontra-se apenas a história.” História relevante ao extremo, é bem verdade, mas apenas história...

Sendo assim, este trabalho propõe ir além das simples frequências de *default*, buscando a capacidade de previsão desse evento através de um modelo de escore de crédito que utiliza dados históricos.

Westgaard e Wijst (2001, p. 339) apresentam a dificuldade que existe com relação aos dados que devem ser utilizados:

Risco de crédito é o risco do tomador/contraparte entrar em *default*, ou seja, não pagar o montante devido ao banco. Essa definição de *default* pode ser discriminada utilizando o montante (acumulado) em atraso ou o período de tempo em que o pagamento está em atraso (normalmente, o número de meses). Tais medidas detalhadas são tipicamente utilizadas dentro de um banco. Na maioria das informações públicas disponíveis, entretanto, essas medidas detalhadas não existem e definições mais gerais têm que ser utilizadas, tais como falências decretadas.³⁵

Isso explica porque muitos estudos relacionados com previsão de *default* de empresas utilizam dados de falência ou concordata, dados publicamente disponíveis, uma vez que dados de atraso são de difícil obtenção. Nesses casos, parte-se do pressuposto de que as empresas em processo de concordata/falência tornar-se-ão inadimplentes, tal como salienta Silva (2000, p. 315): “[...] cabe enfatizar que o método utilizado nesse trabalho apresenta sua validade a partir do ponto em que admitimos que as empresas insolventes (falidas ou concordatárias) serão inadimplentes.”, mesmo que o caminho mais natural seja o da inadimplência para a concordata/falência³⁶.

Sob esse aspecto, podemos citar o trabalho de Lawrence e Smith (1992), que é um dos poucos estudos em que se teve acesso a dados detalhados de operações de crédito. Eles analisaram o

³⁵ “Credit risk is the risk that a borrower/counterparty will default, i.e. fail to repay an amount owed to the bank. This definition of default can be differentiated using the (cumulative) amount overdue or the time that payment is behind (usually the number of months). Such detailed measures are typically used within a bank. In most publicly available information, however, these detailed measures are missing and more general definitions have to be used, such as bankruptcies pronounced by a court of law.”

³⁶ Servigny e Renault (2004, p. 120) mostram esse caminho: “Os passos do *default* para a falência e liquidação também dependem muito da legislação de insolvência vigente em cada país” (“The steps from default to bankruptcy and liquidation also depend very much on the insolvency legislation enacted in each country.”).

risco de *default* em empréstimos para compra de *mobile home*³⁷, e as informações foram fornecidas por uma companhia financeira nacional americana, que disponibilizou dados sobre mais de 170 mil empréstimos, inclusive sobre atraso. Um outro caso a citar é de Gonçalves (2005), que obteve informações, inclusive de atraso, de 20 mil contratos de crédito pessoal de um grande banco de varejo que atua no Brasil, realizados no período de agosto de 2002 a fevereiro de 2003. As variáveis explicativas foram divididas em variáveis cadastrais (relacionadas ao cliente) e variáveis de utilização e restrição (relativas às restrições de crédito e apontamentos sobre outras operações de crédito do cliente existentes no mercado). Com relação a esses estudos, temos o diferencial de trabalhar com dados de grandes empresas, e de as informações serem provenientes de todo o SFN e não apenas de uma IF.

Com relação à mensuração da PD para grandes e pequenos tomadores de crédito, Saunders (2000, p. 207) explica:

A disponibilidade de mais informações e o custo médio mais reduzido de coleta de tal informação permitem que os IFs utilizem métodos mais sofisticados e, geralmente, mais quantitativos de mensuração de probabilidades de inadimplência de grandes tomadores, em comparação com pequenos tomadores. Entretanto, progressos tecnológicos e de coleta de informações estão tornando até mesmo a avaliação quantitativa de pequenos tomadores cada vez mais viável e menos dispendiosa.

Assim, graças ao avanço da tecnologia de informações, os métodos utilizados passam a não ser tão diversos quando se avaliam grandes e pequenos clientes.

Com relação aos empréstimos voltados aos estabelecimentos industriais e comerciais, Saunders (2000, p. 196) observa que esses tipos de operações estão perdendo importância nas carteiras de crédito das IFs. O principal motivo apontado é o crescimento de sucedâneos não bancários, especialmente o *commercial paper*³⁸. Mas, como somente as empresas maiores têm acesso a esse mercado, as IFs acabam tendo que lidar com empresas cada vez menores e de maior risco, o que faz com que a avaliação do risco de crédito torne-se mais importante. Deve-se atentar, no entanto, que o autor faz referência ao mercado americano. No Brasil, verifica-se que as debêntures têm assumido o papel que os *commercial papers* têm nos

³⁷ Segundo o Oxford – *Advanced Learner's Dictionary* (HORNBY, 2000) significa uma construção que pode ser movida, às vezes sobre rodas, que é geralmente estacionada em um local e usada para moradia (“*a building that can be moved, sometimes with wheels, that is usually parked in one place for living in*”).

³⁸ “*Commercial paper* é um instrumento de dívida a curto prazo emitido por empresas diretamente ou por meio de banco de investimento a investidores nos mercados financeiros, tais como fundos mútuos de investimento a curto prazo.” (SAUNDERS, 2000, p. 197).

Estados Unidos, sendo que a principal diferença entre estes papéis é o prazo mais longo das debêntures (LIMA *et al.*, 2006, p.136). Canuto e Lima (2002, p. 222) comentam esse assunto:

A substituição relativa do crédito bancário por emissão e comercialização em mercados secundários de títulos de dívida se deu principalmente no financiamento empresarial de prazos maiores, com a atividade bancária deslocando-se deste em direção ao suprimento de liquidez à intermediação financeira não-bancária.

Comparando as decisões de crédito no varejo e no atacado³⁹, Saunders (2000, p. 205, grifos do original) esclarece:

Geralmente, no nível do varejo, um IF controla seus riscos de crédito por meio do **racionamento de crédito**, e não com o uso de taxas de juros ou preços diferentes. No nível do atacado, os IFs utilizam tanto taxas de juros quanto quantidade de crédito para controlar o risco de crédito. Portanto, quando os bancos quotam uma taxa básica de juros (*L*) para certas empresas, os clientes de risco menor acabam pagando uma taxa inferior à taxa básica. Os tomadores mais arriscados pagam uma margem sobre a taxa básica, ou um prêmio por risco de inadimplência (*m*), para remunerar o IF pelo risco adicional existente.

Ou seja, para as pessoas jurídicas, além do racionamento de crédito, o risco de crédito é controlado através de precificação diferenciada de suas operações.

Bessis (1998, p. 298) apresenta a importância da avaliação do risco individual e do risco de carteira:

O risco individual é relevante quando se toma decisão com relação a novas operações e/ou novas contrapartes. O nível de carteira torna-se importante quando o impacto da operação na carteira tem que ser considerado ou quando o risco de solvência e o capital têm que ser estimados.⁴⁰

Apenas com relação ao risco individual, mas de uma maneira mais ampla, Saunders (2000, p. 194) expõe:

A mensuração do risco de crédito de empréstimos ou obrigações individuais é crucial para que um IF possa (1) precificar um empréstimo ou avaliar uma obrigação corretamente e (2) fixar limites apropriados ao volume de crédito a ser concedido a qualquer tomador, ou a exposição aceitável a perdas com qualquer tomador.

Mas, o risco individual deve ser periodicamente avaliado mesmo após a contratação. É necessário que se proceda ao acompanhamento das operações, especialmente quando se refere

³⁹ Varejo refere-se a pessoas físicas, e atacado, a pessoas jurídicas.

⁴⁰ “*The standalone risk is relevant when making decisions with respect to new transactions and/or new counterparties. The portfolio level becomes important when the impact of a transaction on the portfolio has to be considered or when solvency risk and capital have to be estimated.*”

a uma grande empresa, pois, apesar de, geralmente, representar baixo risco, quando ocorre o *default*, as perdas podem ser enormes.

Com relação ao horizonte de tempo a ser utilizado na construção do modelo, para a estimativa da PD, adota-se o horizonte de um ano. Ou seja, pretende-se estimar a probabilidade de uma grande empresa ficar inadimplente (entrar em *default*) em uma IF no período de um ano após a data-base dos dados utilizados para a análise.

2.6 Escore de Crédito

Para se aferir o risco de inadimplência, Saunders (2000, p. 210) discorre sobre os modelos de escore de crédito⁴¹:

Os modelos de escore de crédito utilizam dados relativos a características observadas do tomador, seja para calcular a probabilidade de inadimplência, seja para colocar os tomadores em classes de risco de inadimplência. Mediante a seleção e combinação de diversas características econômicas e financeiras do tomador, o administrador de um IF pode:

1. Estabelecer numericamente os fatores que são explicações importantes do risco de inadimplência.
2. Avaliar o grau relativo de inadimplência desses fatores.
3. Melhorar a precificação do risco de inadimplência.
4. Ser mais capaz de fazer a triagem de maus devedores.
5. Colocar-se em posição melhor para calcular as reservas necessárias para cobrir perdas futuras esperadas com empréstimos.

O autor salienta a necessidade de se identificar medidas econômicas e financeiras para cada classe específica de tomador (pessoas físicas diferentemente do segmento *corporate*, por exemplo). Após a identificação desses dados, deve-se escolher uma técnica estatística que quantifique ou atribua escores à probabilidade ou classificação em termos de risco. Ele relaciona quatro técnicas aos modelos de escore de crédito: (1) modelos lineares de probabilidade, (2) modelos *logit*, (3) modelos *probit* e (4) análise discriminante linear.

Segundo Thomas *et al.* (2002, p. 5):

⁴¹ Cabe ressaltar a diferenciação que Sicsú (2003) faz com relação ao conceito de *credit scoring* “*stricto sensu*”, que se refere à avaliação do risco de uma operação de crédito para a decisão de aprová-la ou não, do termo *credit scoring* utilizado para outras aplicações das técnicas estatísticas da análise discriminante, ainda que não se destinem à aprovação ou não de operações de crédito.

Há um desenvolvimento paralelo ao escore de crédito que utiliza abordagens de escore para prever o risco de empresas falirem (Altman, 1968). Embora tenha gerado alguns resultados interessantes ligando indicadores contábeis à subsequente falência, por causa das amostras serem tão menores que no crédito ao consumidor, e pela informação contábil ser suscetível à manipulação dos administradores, os preditores são menos precisos do que no caso do crédito ao consumidor.⁴²

No entanto, Servigny e Renault (2004, p. 63) têm uma visão diferente sobre o assunto (grifos nossos):

Escore de crédito é freqüentemente considerado como não muito sofisticado. À primeira vista, parece que a indústria bancária considera que nada de significativo foi descoberto desde a função Z de Altman (1968). Juntam-se informações em um pequeno conjunto de variáveis financeiras-chaves e colocam-nas em um modelo simples, que separa as firmas boas das firmas ruins. A realidade é totalmente diferente: o escore de crédito é um pedaço bem pequeno de um enorme quebra-cabeça de mineração de dados [...] Escore de crédito aplica-se a qualquer tipo de devedor. Para as maiores corporações (sociedades por ações com capital aberto), os modelos estruturais podem ser um atrativo.⁴³

De onde se conclui que não há, na opinião deles, nenhuma restrição quanto à utilização do escore de crédito para as grandes empresas. Eles apontam uma opção para as grandes empresas: modelos estruturais⁴⁴, ou também conhecidos como modelos baseados no valor da firma. Esses modelos não serão abordados neste trabalho, já que nossa amostra vai além de companhias de capital aberto.

Servigny e Renault (2004, p. 73) relatam o desenvolvimento dos modelos desde Fitzpatrick, que estabeleceu a dependência entre a probabilidade de *default* e as características individuais dos créditos *corporate*, em 1932; passando pelo grande marco para as técnicas quantitativas que foi o cartão de crédito, na década de 60, fazendo com que as decisões de aprovação de crédito fossem automatizadas, dado o tamanho da população a utilizar esse produto; até chegar ao escore de crédito, na década de 70, quando foi completamente reconhecido. Uma observação importante dos autores é a seguinte:

⁴² “There is a parallel development to credit scoring in using scoring approaches to predict the risk of companies going bankrupt (Altman 1968). Although this has provided some interesting results connecting accounting ratios to subsequent bankruptcy, because samples are so much smaller than in consumer credit and because accounting information is open to manipulation by managers, the predictions are less accurate than for the consumer credit case.”

⁴³ “Credit scoring is often perceived as not being highly sophisticated. At first sight it seems that the banking industry considers that nothing significant has been discovered since Altman’s (1968) Z-score. One gathers information on a small set of key financial variables and inputs them in a simple model that separates the good firms from the bad firms. The reality is totally different: Credit scoring is a very little piece of the large data-mining jigsaw [...] Credit scoring models apply to any type of borrower. For the largest corporates (those with listed equity), structural models may be an appeal.”

⁴⁴ Segundo Crook *et al.* (2007, p. 11), a idéia principal é que a empresa entrará em *default* se a dívida exceder o valor da empresa.

Todas essas abordagens focaram na previsão de fracasso e na classificação da qualidade de crédito. Essa distinção é muito importante, mesmo assim ainda não está claro para muitos dos usuários dos escores se o aspecto mais importante a ser focado é a classificação ou a previsão. Isso geralmente representará dificuldades quando da seleção dos critérios para as medidas de desempenho.⁴⁵

À parte desse aspecto, o escore de crédito tornou-se uma técnica bastante difundida nos bancos, pois possibilita o aumento da produtividade alcançada através de uma avaliação de crédito rápida, com redução dos custos, o que aumenta a sua competitividade. Além disso, pesa o foco dado às PDs pelo Basileia II.

Apesar de a escolha do modelo de escore de crédito ótimo continuar sendo um desafio, Galindo e Tamayo *apud* Servigny e Renault (2004, p. 75) definem cinco requisitos de qualidade para que se efetue essa escolha:

1. *Acurácia*. Ter baixas taxas de erro geradas das premissas do modelo
2. *Parcimônia*. Não utilizar um número grande de variáveis explicativas
3. *Significância*. Produzir resultados interessantes
4. *Viabilidade*. Rodar em um período de tempo razoável e utilizando recursos realistas
5. *Transparência e interpretabilidade*. Prover uma percepção de alto nível nas relações entre os dados e tendências, e entendimento sobre a origem dos resultados do modelo⁴⁶

Esses requisitos serão considerados na construção do modelo do tipo escore de crédito, que tem como foco a previsão de inadimplência de grandes empresas.

2.7 Perda Esperada e Perda Não Esperada

Segundo o BCBS (2005, p.1):

Na área de crédito, perdas de juros e principal ocorrem a todo tempo – há sempre alguns devedores que não cumprem suas obrigações. As perdas que realmente ocorrem em um determinado ano variam de um ano para outro, dependendo do número e da severidade dos eventos de *default*, mesmo se assumirmos que a qualidade da carteira é consistente no tempo.⁴⁷

⁴⁵ “All these approaches focused both on the prediction of failure and on the classification of credit quality. This distinction is very important, as it is still not clear in the minds of many users of scores whether classification or prediction is the most important aspect to focus on. This will typically translate into difficulties when selecting criteria for performance measures.”

⁴⁶ “1. *Accuracy*. Having low error rates arising from the assumptions in the model
2. *Parcimony*. Not using too large a number of explanatory variables
3. *Nontriviality*. Producing interesting results
4. *Feasibility*. Running in a reasonable amount of time and using realistic resources
5. *Transparency and interpretability*. Providing high-level insight into the data relationships and trends and understanding where the output of the model comes from”

⁴⁷ “In credit business, losses of interest and principal occur all the time – there are always some borrowers that default on their obligations. The losses that are actually experienced in a particular year vary from year to year,

O comitê apresenta dois gráficos esclarecedores com relação às perdas esperadas e às perdas não esperadas, reproduzidos no Gráfico 3:

Gráfico 3 - Gráficos representativos da taxa de perdas

FONTE: Adaptado do BCBS, 2005, p. 2

Como não se sabe, com antecedência, que perdas o banco sofrerá em um determinado ano, é possível prever a perda média, tal como no Gráfico 3, que é uma expectativa razoável da EL (*Expected Loss* - Perda Esperada). Essa perda é considerada como um componente do custo do negócio, sendo utilizada na precificação das exposições de crédito e, também, na constituição de provisões.

Quanto aos picos de perdas, estes não são frequentes, mas, quando ocorrem, podem representar perdas elevadas. Essa perda que excede a perda esperada é chamada de UL (*Unexpected Loss* - Perda não Esperada), que deve ser suportada pelo capital do banco.

O BCBS (2005, p. 3) apresenta o gráfico de distribuição de perdas potenciais com mais detalhe, como mostra o Gráfico 4:

depending on the number and severity of default events, even if we assume that the quality of the portfolio is consistent over time.”

Gráfico 4 – Exemplificativo de distribuição de perdas potenciais

FONTE: Adaptado do BCBS, 2005, p. 3

Explica-se que:

A curva mostra que pequenas perdas próximas ou um pouco abaixo da EL ocorrem com maior frequência do que grandes perdas. A probabilidade de que as perdas excedam a soma da EL e UL - isto é, a probabilidade de que um banco não seja capaz de honrar suas próprias obrigações creditícias com seu lucro e capital - equivale à área hachurada abaixo do lado direito da curva. 100% menos essa probabilidade é chamada de **nível de confiança** e o correspondente limite é chamado de **VaR (Valor em Risco)**.⁴⁸ (BCBS, 2005, p. 3, grifos do original).

Jorion (2003, p. 19) define de uma forma intuitiva: “O VAR sintetiza a maior (ou pior) perda esperada dentro de determinados período de tempo e intervalo de confiança [...]”, e ainda faz uma introdução à mensuração do risco integrado:

O propósito original dos sistemas de VAR era quantificar o risco de mercado. [...] os bancos que estavam sujeitos a restrição de capital por causa do risco de crédito logo aprenderam a transformar esse risco em risco de mercado, sujeito a menores exigências de capital [...] A metodologia de VAR de mercado está sendo estendida para agregar o risco de mercado ao risco de crédito.

Voltando ao Gráfico 4, reforça-se o comentado anteriormente: a EL é coberta por provisões e o capital é determinado pela diferença entre o VaR e a EL, ou seja, a UL.

⁴⁸ “The curve shows that small losses around or slightly below the Expected Loss occur more frequently than the large losses. The likelihood that losses will exceed the sum of Expected Loss (EL) and Unexpected Loss (UL) – i.e. the likelihood that a bank will not be able to meet its own credit obligations by its profits and capital – equals the hatched area under the right hand side of the curve. 100% minus the likelihood is called **confidence level** and the corresponding threshold is called **Value-at-Risk (VaR)** at this confidence level.”

Cabe salientar que o assunto foi exposto do ponto de vista de carteira, ou seja, a perda esperada de uma carteira é função da proporção de devedores que irão entrar em *default* em um certo período de tempo, multiplicada pela exposição no momento de *default* e multiplicada pela taxa de perda dado o *default*. Vale lembrar que todas essas variáveis são aleatórias, mas as IFs podem estimá-las.

Assim, temos a seguinte equação para o cálculo da perda esperada:

$$EL = PD \times EAD \times LGD$$

onde:

PD : *Probability of Default* (Probabilidade de inadimplência)

EAD : *Exposure at Default* (Valor da exposição no momento do evento de inadimplência)

LGD : *Loss Given Default* (Perdas efetivas em função de um evento de inadimplência).

Segundo Niyama e Gomes (2005, p. 54):

A constituição de provisão para créditos de liquidação duvidosa representa, em qualquer empresa, uma estimativa de perda provável dos créditos, em atendimento aos Princípios Fundamentais de Contabilidade, em especial ao da Realização da Receita e Confrontação com a Despesa e ao da Prudência ou Conservadorismo.

Ou seja, mais uma confirmação de que a EL deverá ser suportada pela provisão para créditos de liquidação duvidosa.

Do mesmo modo, a Resolução n. 2.682/99 estabelece que (grifos nossos):

Art. 6º A provisão para fazer face aos créditos de liquidação duvidosa deve ser constituída mensalmente, não podendo ser inferior ao somatório decorrente da aplicação dos percentuais a seguir mencionados, sem prejuízo da responsabilidade dos administradores das instituições pela constituição de provisão em montantes suficientes para fazer face a perdas prováveis na realização dos créditos:

- I - 0,5% (meio por cento) sobre o valor das operações classificadas como de risco nível A;
- II - 1% (um por cento) sobre o valor das operações classificadas como de risco nível B;
- III - 3% (três por cento) sobre o valor das operações classificadas como de risco nível C;
- IV - 10% (dez por cento) sobre o valor das operações classificadas como de risco nível D;
- V - 30% (trinta por cento) sobre o valor das operações classificadas como de risco nível E;
- VI - 50% (cinquenta por cento) sobre o valor das operações classificadas como de risco nível F;

- VII - 70% (setenta por cento) sobre o valor das operações classificadas como de risco nível G;
- VIII- 100% (cem por cento) sobre o valor das operações classificadas como de risco nível H.

Assim, temos que a classificação de risco de uma operação de crédito está relacionada com a EL da operação, segundo o estabelecido na resolução. Este trabalho visa criar um modelo para estimar a PD de uma grande empresa, que é parâmetro fundamental para a determinação da EL⁴⁹ e, conseqüentemente, da classificação de risco da empresa e/ou operação de crédito.

Além disso, deve-se destacar a importância que a PD assume no contexto do Basileia II, que impõe uma nova estrutura de adequação de capital, com aumento da sensibilidade ao risco do requerimento mínimo de capital. Dada a relevância desse assunto, seguem-se alguns aspectos desse acordo.

2.8 Basileia II

Nos anos 70, o ambiente bancário encontrava-se estabilizado devido a vários fatores, dentre os quais podemos citar: a indústria estava pesadamente regulada; as operações dos bancos comerciais consistiam basicamente em tomar recursos e emprestá-los; havia pouca competição, facilitando uma lucratividade justa e estável; os reguladores estavam preocupados com a segurança da indústria e o controle do seu poder de criar moeda; as regras limitaram o escopo das operações de vários tipos de instituições, limitando seus riscos; e existiam poucos incentivos para mudanças e competição (BESSIS, 1998, p. 3). Mas, ainda nos anos 70 e durante os anos 80, ocorreram as primeiras “ondas drásticas” de mudanças na indústria bancária, aumentando os riscos envolvidos. Os motivos das mudanças foram: a nova competição, inovações dos produtos, mudança da fonte de financiamento dos bancos comerciais para o mercado de capitais, aumento da volatilidade, e desaparecimento das barreiras que limitavam o escopo das operações para vários tipos de IFs, além da globalização, que exigia menor heterogeneidade nos marcos regulatórios⁵⁰.

Nesse contexto, o Basileia I foi aprovado pelo BCBS, sob o patrocínio do BIS (*Bank for*

⁴⁹ Ver Equação 1.

⁵⁰ Segundo Canuto e Lima (2002, p. 226): “[...] riscos sistêmicos podem ser criados mediante contágios de desordem financeira originados em centros financeiros com regulação mais frouxa, além da exposição diante de riscos jurídicos externos sobre os quais o sistema regulatório doméstico não pode monitorar.”

International Settlement - Banco de Compensações Internacionais). Canuto e Lima (2002, p. 231) explicam que:

O Acordo de Adequação de Capital da Basiléia, em 1988, estabeleceu um nível mínimo de 8% para a proporção entre o capital e o valor dos ativos dos bancos, como forma de garantir solidez e segurança. Foi motivado pela percepção de que a intensa concorrência estava induzindo os bancos internacionalmente ativos a assumir patamares baixos em tais reservas de capital, na busca de ampliação de suas parcelas de mercado.

Salientam, também, a busca pela harmonização regulatória e a distinção entre os tipos de ativos por risco. Basicamente, os ativos deveriam ser enquadrados em quatro grupos de risco, aplicando os percentuais de 0%, 20%, 50% e 100% (dependendo do grupo⁵¹) sobre o seu valor, obtendo-se assim o ativo ponderado pelo risco, sobre o qual deveria ser aplicado o percentual de 8% (11%⁵², no Brasil), resultando no capital mínimo exigível. O foco era o risco de crédito.

Posteriormente, em 1996, foi estabelecido o requisito de capital para cobertura dos “riscos de mercado” (riscos de perdas originados de movimentos nos preços de mercado) [...] a grande novidade foi a permissão de que os bancos usassem modelos internos próprios para determinar a carga de capital para cobrir o risco de mercado, como alternativa ao “modelo padronizado”. (CANUTO; LIMA, 2002, p. 232).

Com isso, um conceito relacionado a esses modelos internos tornou-se muito disseminado no mercado: o VaR⁵³, que se constituía na base para o cálculo do requerimento de capital correspondente ao risco de mercado.

Cossin e Pirotte (2000, p. 2) expõem uma das consequências dessa mudança:

Como os seus sistemas de gestão tornaram-se mais sofisticados com respeito ao risco de mercado, instituições financeiras ficaram mais atentas à fraqueza dos seus cálculos de exposição ao risco de crédito e à necessidade de valorar essa exposição, mais do que racioná-la através de linhas de crédito.⁵⁴

Canuto e Lima (2002, p. 233) salientam que o Basiléia I, apesar das emendas efetuadas, continuava em processo de deterioração da sua eficácia devido, principalmente, à prática de

⁵¹ No Brasil, conforme disposto na Tabela anexa ao Anexo IV da Resolução nº. 2.099, de 17.08.94 (BCB,2007j). No caso de empresas privadas, o fator de risco é normal, com ponderação de 100%.

⁵² Conforme modificado pela Resolução nº. 2.692 de 24.02.00 (BCB, 2007k).

⁵³ Segundo Jorion (2003, p. 20), a criação do termo pode ser atribuída a Till Guldemann, quando ele trabalhava para a J.P.Morgan, no final dos anos 80.

⁵⁴ “As their risk management systems become more sophisticated with respect to market risk, financial institutions become more aware of the weakness of their credit risk exposure calculation and of the need to value this exposure, rather than ration it through credit lines.”

“arbitragem de capital regulatório”⁵⁵. Eles acrescentam:

A arbitragem de capital regulatório seguiu-se como estratégia mais viável e relevante, particularmente à medida em que as novas tecnologias de mensuração de riscos, pelos bancos, permitiram maiores diferenças entre os requisitos por eles estimados como necessários e aqueles estabelecidos para ativos específicos pelos reguladores.

Esse tipo de prática, aliada às contínuas inovações financeiras, resultaram na necessidade de reforma nesse acordo. Carneiro *et al.* (2005, p. 29) explicam:

Como resultado de intensos e continuados estudos liderados pelo BCBS sobre supervisão bancária, em conjunto com representantes de bancos centrais e órgãos de fiscalização, pesquisadores, estudiosos, acadêmicos e profissionais do mercado financeiro, o Acordo de 1988 foi totalmente revisado, culminando na publicação, aberta a comentários públicos, em junho de 1999, da primeira versão do documento Convergência Internacional de Mensuração e Padrões de Capital: Uma Estrutura Revisada, conhecida como Novo Acordo de Capital ou ainda como Basiléia II. Esse documento foi objeto de vários aprimoramentos, tendo crescido em sofisticação e complexidade.

Servigny e Renault (2004, p. 395) apresentam os objetivos finais do BCBS relativos ao Basiléia II:

- Prover segurança e estabilidade para o sistema financeiro internacional através da manutenção de um nível apropriado de capital nos bancos
- Prover incentivos para competição justa entre bancos
- Desenvolver uma abordagem mais ampla para mensurar riscos
- Focar em bancos internacionais⁵⁶

E, para se ter uma idéia das grandes mudanças impostas por esse novo acordo, eles apresentam os três pilares do Basiléia II (2004, p. 396):

- Pilar 1 – requerimentos de capital mínimo. O objetivo é determinar o montante de capital requerido, dado o nível de risco de crédito na carteira do banco.
- Pilar 2 – revisão pela supervisão. A revisão pela supervisão permite ação antecipada pelos reguladores e impede os bancos de usar dados não confiáveis. Os reguladores devem lidar com assuntos tais como o efeito pró-cíclico, que pode surgir como uma consequência de uma maior sensibilidade de risco da mensuração do capital.

⁵⁵ “Essas são estratégias de redução de requisitos regulatórios de capital sem uma concomitante diminuição no grau de exposição a riscos. Por exemplo, a possibilidade de venda – ou outro deslocamento para fora do balanço – de ativos para os quais a avaliação, pelo banco, do capital adequado é menor do que a estabelecida pela regulação, deixando em carteira os ativos em situação oposta. A consequência tende a ser registros de capital/ativos regulatórios abaixo dos que seriam economicamente adequados. Mesmo quando os primeiros se mostram crescentes, a segurança pode ser ilusória.” (CANUTO; LIMA, 2002, p. 233).

⁵⁶ “• *Providing security and stability to the international financial system by keeping an appropriate level of capital within banks*

- *Providing incentive for fair competition among banks*
- *Developing a wider approach to measure risks*
- *Focusing on international banks*”

- Pilar 3 – disciplina do mercado. A divulgação de informações de um banco perante seus competidores e mercados financeiros visa permitir a monitoração externa e uma melhor identificação do seu perfil de risco pela comunidade financeira.⁵⁷

Os autores esclarecem que focam exclusivamente o risco de crédito no Pilar 1, mas salientam que, além do risco de crédito, esse pilar incorpora o risco de mercado (já existente no acordo anterior) e o risco operacional (incluído nesse novo acordo). Como o interesse deste estudo está no risco de crédito e na sua mensuração, o Pilar 1, tal e qual descrito por eles, representa a nossa “área de estudo”.

Apesar da novidade da inclusão do tratamento dos riscos operacionais dos bancos no Basileia II, o maior impacto foi gerado pelo risco de crédito, conforme expõem Canuto e Lima (2002, p. 234):

A mudança maior, porém, está na inclusão de opções para o tratamento dos riscos de crédito. Para bancos menos habilitados a manejar sistemas sofisticados de avaliação e gestão de riscos, prevê a manutenção de pesos padronizados pelos supervisores para as classes dos ativos, com o recurso, contudo a *ratings* efetuados por instituições externas aos bancos no caso de cada ativo. Alternativamente, dependendo da aprovação pelos supervisores, faculta a permissão do manejo, pelos bancos, de seus sistemas próprios de avaliação de riscos (*Internal Risk Based Approaches – IRB*), em dois níveis: *foundation* e *advanced*. A busca de alinhamento entre requisitos de capital regulatório e aqueles considerados adequados pelos próprios bancos seria reforçada pela transição do enfoque padronizado para os IRB.

Carneiro *et al.* (2005, p. 29) explicam com mais detalhes a abordagem IRB e os seus dois tipos:

A abordagem IRB é baseada na estimativa das perdas esperadas (*expected losses – EL*), que serão confrontadas com as provisões constituídas; e das perdas não-esperadas (*unexpected losses – UL*), que serão base para a exigência de capital. Esses valores são calculados a partir de parâmetros fundamentais, definidos como componentes de risco: probabilidade de inadimplência (*probability of default – PD*); perdas efetivas em função de um evento de inadimplência (*loss given default – LGD*); valor da exposição no momento do evento de inadimplência (*exposure at default – EAD*); prazo até o vencimento (*maturity – M*).

A abordagem IRB pode ser implementada de duas formas alternativas definidas como: abordagem IRB *foundation*, na qual a PD é calculada pelas próprias entidades bancárias, enquanto outros são definidos pelo órgão supervisor; e abordagem IRB *Advanced*, na qual o órgão de supervisão admite que todos os parâmetros fundamentais para a definição do risco de crédito sejam calculados

⁵⁷ “• *Pillar 1 – minimum capital requirements. The objective is to determine the amount of capital required, given the level of credit risk in the bank portfolio.*

• *Pillar 2 – supervisory review. The supervisory review enables early action from regulators and deters banks from using unreliable data. Regulators should also deal with issues such as procyclicality that may arise as a consequence of a higher risk sensitivity of capital measurement.*

• *Pillar 3 – market discipline. The disclosure of a bank vis-à-vis its competitors and financial markets is devised to enable external monitoring and a better identification of its risk profile by the financial community.”*

e definidos pelas entidades supervisionadas, a partir de seus próprios modelos internamente desenvolvidos.

Esse redirecionamento do foco da supervisão bancária é também analisado por Canuto e Lima (2002, p. 234), que justificam a mudança de abordagem em relação ao controle do modelo de avaliação de riscos e, conseqüentemente, do requerimento de capital das IFs:

A supervisão focaria, cada vez mais, a qualidade da gestão de riscos e a adequação de sua medida, além de alguns parâmetros mais gerais a serem observados. Além disso, a intenção manifesta na atual proposta de Basileia 2 é a de que seja estabelecida uma estrutura de incentivos favorável à adoção dos IRB e a seu aprimoramento. O peso médio efetivo dos riscos para o banco como um todo deveria cair com os enfoques mais sofisticados, conforme extensão da arbitragem de capital através deles alcançadas. Portanto, tais bancos poderiam atingir encargos mais baixos de capital regulatório total e, assim, razões de capital com ponderação de riscos registradas como mais altas.

Deste modo, haveria, em princípio, vantagens competitivas para o banco que conseguisse “liberar” capital para outras aplicações, ou mesmo, conseguir melhor precificar as operações, obtendo melhores resultados em função de uma gestão de riscos mais eficaz. Crook *et al.* (2007, p. 1461) comentam a adoção da abordagem IRB:

A vantagem da abordagem IRB sobre a abordagem padronizada é que a primeira pode proporcionar ao banco uma diminuição de capital e, portanto, um maior retorno sobre o patrimônio. Portanto, espera-se que os grandes bancos adotarão a abordagem IRB.⁵⁸

Ou seja, para usufruir dessa vantagem, as IFs devem calcular a PD, já que ela é requerida nas duas abordagens IRB. Assim, torna-se necessário às IFs construir modelos que calculem a PD de seus clientes. Com relação a grandes empresas, o modelo pode ter, como base, dados contábeis e cadastrais, considerando ou não variáveis macroeconômicas.

2.9 Análise das Demonstrações Financeiras

Matarazzo (2003, p. 15) expõe que:

As demonstrações financeiras fornecem uma série de *dados* sobre a empresa, de acordo com regras contábeis. A Análise de Balanços transforma esses *dados* em *informações* e será tanto eficiente quanto melhores informações produzir.

⁵⁸ “The attraction of the IRB approach over the Standardised Approach is that the former may enable a bank to have less capital and so earn higher returns on equity than the latter. Therefore it is likely that most large banks will adopt the IRB approach.”

A Lei das Sociedades por Ações (Lei n. 6.404/76) determina a elaboração de quatro demonstrações financeiras: balanço patrimonial, demonstração dos lucros ou prejuízos acumulados⁵⁹, demonstração do resultado do exercício e demonstração das origens e aplicações de recursos. Apesar de ser direcionada somente às sociedades por ações, a legislação fiscal tornou essa determinação obrigatória também para os demais tipos de sociedades (MATARAZZO, 2003, p. 41).

Uma breve descrição de cada uma dessas demonstrações é apresentada, com base em Matarazzo:

- Balanço Patrimonial (2003, p. 41):

É a demonstração que apresenta todos os bens e direitos da empresa – Ativo –, assim como as obrigações – Passivo Exigível – em determinada data. A diferença entre Ativo e Passivo é chamada Patrimônio Líquido e representa o capital investido pelos proprietários da empresa, quer através de recursos trazidos de fora da empresa, quer gerados por esta em suas operações e retidos internamente.

- Demonstração do Resultado do Exercício (2003, p. 45, grifos do original): “[...] é uma demonstração dos aumentos e reduções causados no Patrimônio Líquido pelas operações da empresa. [...] retrata apenas o **fluxo econômico** e não o **fluxo monetário** (fluxo de dinheiro).”

- Demonstração das Mutações do Patrimônio Líquido (2003, p. 46):

[...] apresenta as variações de todas as contas do Patrimônio Líquido ocorridas entre os dois balanços, independentemente da origem da variação, seja ela proveniente da correção monetária⁶⁰, de aumento de capital, de reavaliação de elementos do ativo, de lucro ou simples transferência entre contas, dentro do próprio Patrimônio Líquido.

- Demonstração das Origens e Aplicações de Recursos - DOAR (2003, p. 48):

[...] mostra as novas origens e aplicações verificadas durante o exercício. [...] mas apenas aquelas ocorridas nos itens não Circulantes do Balanço, ou seja, no Exigível a Longo Prazo, Patrimônio

⁵⁹ Segundo Iudícibus *et al.* (2003, p. 31), a Lei das Sociedades por Ações aceita as Demonstrações das Mutações do Patrimônio Líquido e de Lucros ou Prejuízos Acumulados, sendo que a primeira é mais completa e uma das suas colunas é a de lucros e prejuízos acumulados.

⁶⁰ Segundo Iudícibus *et al.* (2003, p. 501), a Lei n.º 9.249/95 revogou a correção monetária das demonstrações contábeis, ao mesmo tempo em que criou a figura dos Juros Sobre o Capital Próprio.

Líquido, Ativo Permanente e Realizável a Longo Prazo. A diferença entre as novas origens não circulantes e as novas aplicações não circulantes será igual ao Capital Circulante Líquido.

Isto posto, para que as demonstrações contábeis reflitam o valor econômico de uma empresa, Eliseu Martins (2001, p. 265) apresenta algumas dificuldades existentes:

1. os relatórios contábeis normalmente se baseiam em custos históricos afastados dos valores correntes;
2. alguns itens, principalmente as contas a receber, estão por seu valor futuro, quando o adequado seria o valor presente;
3. o conservadorismo tende a subestimar os valores dos ativos; e
4. existem várias operações relevantes que normalmente não são registradas (*off-balance sheet*), tais como: arrendamento mercantil, posições em derivativos, garantias, *goodwill*, etc.

Para que se amenizem os efeitos dessas limitações, são necessários alguns ajustes. Nesse sentido, Matarazzo (2003, p.76) expõe:

As modificações de preços e serviços representam o fato mais perturbador para a contabilidade tradicional, e a solução desse problema, o maior desafio à contabilidade moderna. [...] O processo de correção monetária porém, convencional, em que se corrige Ativo Permanente e Passivo Permanente, está correto, embora incompleto.

Com relação a esse assunto⁶¹, complementa (MATARAZZO, 2003, p. 106):

Daí surgiu a chamada Correção Monetária Integral, que nada mais é do que a denominação que se deu ao processo de correção monetária em que todas as contas de Balanço e Demonstração do Resultado são convertidas numa moeda de capacidade aquisitiva constante (que poderá ser real ou fictícia).

Mas, o autor ainda discorre sobre a inadequação dos índices de inflação (MATARAZZO, 2003, p. 117):

Além de provocar a inflação com a má administração, déficit público e tentativas de enganar o público com alquimias que pretensamente acabariam com a inflação, os sucessivos governos federais conseguiram deturpar a série histórica de quase todos os índices que tentam mensurar a inflação. [...] Com isso qualquer tentativa de fazer correção monetária baseada em índices oficiais fica totalmente prejudicada.

Assim, ele apresenta a opção da dolarização das demonstrações financeiras, ainda que com ressalvas (MATARAZZO, 2003, p. 119):

⁶¹ Embora a Instrução Normativa da CVM n°. 248/96 tenha tornado facultativa a elaboração e divulgação das demonstrações em moeda de capacidade aquisitiva constante, a Orientação CVM n°. 29/96 estabeleceu os requisitos a serem observados para as empresas que optassem por divulgar voluntariamente informações complementares (IUDÍCIBUS *et al.*, 2003, p. 501).

Diante de inúmeros inconvenientes no emprego de índices de inflação, a melhor solução parece mesmo o uso de moeda forte como por exemplo o Dólar. Aliás, não se trata de eleger a melhor solução, mas da (sic) menos ruim, pois o próprio Dólar, além da inflação americana e de sua desvalorização internacional a longo prazo, historicamente tem sofrido sobressaltos na paridade cambial com a moeda nacional. Contudo, é ainda de longe menos ruim que (sic) o uso dos mencionados índices. [...] Entretanto, como é proibido o uso da moeda estrangeira em demonstrações financeiras oficiais, a dolarização somente se aplicaria para uso interno e para análises das demonstrações financeiras.

Além do ajuste dos efeitos da inflação, pelo menos uma conta deve ser reclassificada (MATARAZZO, 2003, p. 136):

[...] Duplicatas descontadas; do ponto de vista contábil, é uma dedução de Duplicata a Receber; do ponto de vista de financiamentos, porém, é um recurso tomado pela empresa junto aos bancos, devido à insuficiência de recursos próprios. Em nada se distingue de empréstimos bancários, do ponto de vista financeiro. Por isso, as Duplicatas Descontadas devem figurar no Passivo Circulante.

Quando o autor mostra o panorama das técnicas de análise de balanços (MATARAZZO, 2003, p. 20), observa que ela surgiu e se desenvolveu dentro do sistema bancário, que continua sendo o seu principal usuário. Dentre as técnicas expostas, apresenta a análise através de índices, com destaque para a utilização na previsão de insolvência. Os índices de balanço têm como principal característica: “[...] fornecer avaliações genéricas sobre diferentes aspectos da empresa em análise, sem descer a um nível maior de profundidade.” (MATARAZZO, 2003, p. 24).

Gitman (2004, p. 42) discorre sobre a análise de índices (grifos nossos):

As informações contidas nas quatro demonstrações financeiras básicas são extremamente importantes para diversos grupos que necessitam regularmente construir medidas relativas de eficiência operacional da empresa. A palavra-chave é *relativas*, neste caso, porque a análise de demonstrações financeiras baseia-se no uso de *valores relativos* ou *índices*. **A análise de índices** envolve métodos de cálculo e interpretação de índices financeiros visando analisar e acompanhar o desempenho da empresa. Os elementos básicos dessa análise são a demonstração de resultados e o balanço patrimonial da empresa.

Ainda:

A análise de índices a partir das demonstrações financeiras é importante para os acionistas, os credores e os administradores da própria empresa. Tanto os acionistas atuais como os possíveis acionistas futuros estão interessados no nível corrente e no nível futuro de risco e retorno da empresa, os quais afetam diretamente o preço da ação. Os credores preocupam-se principalmente com a liquidez de curto prazo da empresa e com a sua capacidade de fazer pagamentos de juros e amortização. Uma de suas preocupações secundárias tem a ver com a rentabilidade da empresa: os

credores querem certificar-se de que ela é saudável. Os administradores, como os acionistas, preocupam-se com todos os aspectos da situação financeira da empresa e procuram construir índices financeiros que sejam considerados favoráveis tanto pelos proprietários como pelos credores. Além disso, os administradores utilizam índices que acompanham o desempenho da empresa de período em período. (GITMAN, 2004, p. 42, grifos nossos).

Com relação aos credores das empresas, quando se tratam de IFs, dependendo do tipo de operação e do relacionamento com a empresa, a visão pode ser tanto de curto como de longo prazos. Nesse sentido, Matarazzo (2003, p. 32) comenta sobre bancos com carteira comercial:

Os bancos concedem crédito a curto prazo, ou seja, o empréstimo deve ser pago dentro de dois ou três meses. Em virtude, porém, da renovação dos créditos que concede e da permanência ou manutenção dos contatos ou do relacionamento com os clientes, há algo de longo prazo no relacionamento banco-cliente.

Sendo assim, verificar se a empresa é “saudável” pode não ser apenas uma preocupação secundária para as IFs em geral.

Considerando que os dados das demonstrações financeiras estão adequados para que as empresas sejam analisadas através de índices, parte-se para a seleção desses índices. Nesse sentido, Matarazzo (2003, p. 150) explica que: “Cada autor apresenta um conjunto de índices que, de alguma forma, difere dos demais. Mesmo em relação aos índices que constam de praticamente todas as obras, pode haver algumas pequenas diferenças de fórmula.” No Anexo C, apresenta-se o quadro-resumo dos índices do autor, dividido em: estrutura de capitais, liquidez e rentabilidade.

Com relação às grandes empresas, Bessis (1998, p. 88) expõe: “Para clientes *corporate*, os critérios para avaliar riscos são bem conhecidos: lucratividade, crescimento, perspectiva da indústria, vantagens competitivas, administradores e acionistas, além do conjunto de indicadores padrão.”⁶²

Assim, consideramos os índices propostos por Matarazzo (2003, p. 152) e complementamos com alguns índices utilizados em diversos estudos sobre previsão de *default*.

⁶² “For corporate borrowers, the criteria for assessing risk are well known: profitability, growth, industry outlook, competitive advantages, management and shareholders, in addition to the standard set of ratios.”

2.10 Variáveis Macroeconômicas

Segundo Carling *et al.* (2007, p. 863), há indícios de que, enquanto os modelos baseados em informações específicas podem ordenar as empresas (ou operações) em função dos seus riscos, a inclusão de variáveis macroeconômicas é essencial para se obter o nível absoluto do risco de *default* corretamente estimado. No seu modelo, a macroeconomia foi representada por variáveis simbólicas de condições de demanda, atividade econômica futura real e expectativa das famílias sobre a atividade econômica futura.

2.11 Estudos Anteriores

Para se ter uma idéia das técnicas estatísticas mais utilizadas em modelos de previsão de *default*, podemos tomar como base, em relação aos trabalhos realizados no Brasil, o Quadro 2:

Quadro 2 - Estudos de previsão de falência no Brasil - 1976 a 2001

Ano	Autor(es)	Técnica Estatística
1976	Elizabetsky	análise discriminante
1978	Kanitz	análise discriminante
1978	Matias	análise discriminante
1979	Altman, Baydia e Dias	análise discriminante
1980	Marques	análise discriminante
1985	Bragança e Bragança	análise discriminante
1986	Kasznar	análise discriminante
1987	Carmo	análise fatorial
1988	Nunes	análise discriminante
1996	Santos	análise discriminante
1996	Matias e Siqueira	regressão logística
s.d.	Matias e Almeida	redes neurais
1997	Silva	análise discriminante
1997	Securato e Famá	cálculo do coeficiente de variação
1999	Rocha	modelo de risco proporcional de Cox
1999	Janot	modelo de risco proporcional de Cox e regressão logística
2001	Lachtermacher e Espenchitt	redes neurais
2001	Minussi	regressão logística

FONTE: MARTINS, Márcio, 2003, p. 22

Um levantamento mais abrangente, de caráter internacional, elaborado por Aziz e Dar (2004, p. 32-34) pode ser visto no Anexo B. Nos dois levantamentos, verifica-se a grande

predominância de modelos de análise discriminante nos anos 70 até os anos 90, com o “aparecimento” dos modelos de regressão logística nos anos 80⁶³.

Alguns dos trabalhos que motivaram a escolha da técnica estatística e das variáveis explicativas utilizadas neste estudo são: Ohlson (1980), Casey e Bartczak (1985), Wetsgaard e Wijst (2001) e Brito (2005).

2.11.1 Ohlson

Ohlson (1980) apresentou os resultados de um estudo de previsão de quebra de empresas, evidenciada pelo evento de falência, comparando com alguns estudos anteriores ao dele. Alguns diferenciais citados foram: o maior tamanho da amostra, a utilização da análise logística condicional em substituição à análise discriminante multivariada (a técnica mais popular) e, principalmente, a preocupação com a informação da época em que os relatórios foram divulgados. A grande ênfase dada a esse último item deve-se à possibilidade de a falência ocorrer após o final do exercício fiscal, mas antes da publicação das demonstrações financeiras. Se o propósito é investigar as relações previstas, a utilização de informações sem o devido cuidado prejudica a avaliação do modelo.

A população constituía-se de empresas industriais com ações negociadas em bolsa ou balcão, no período de 1970 a 1976. A amostra era composta por 105 empresas que faliram e 2.058 que não faliram no período (para estas, foi determinado, aleatoriamente, um único ano para obter os dados das demonstrações contábeis).

Na ausência de uma teoria positiva sobre falência, pela simplicidade computacional e interpretativa, a função escolhida foi a logística. Também, pelo critério da simplicidade, foram escolhidas as seguintes variáveis explicativas:

- SIZE = $\log$ (total dos ativos / índice de preços);
- TLTA = exigibilidades / total dos ativos;

⁶³ Segundo Thomas *et al.* (2002, p. 4), quando apresentam o histórico do escore de crédito, expõem: “Nos anos 80, a regressão logística e a programação linear, as duas principais técnicas dos criadores de escore de crédito de hoje em dia, foram introduzidos.” (“*In the 1980’s, logistic regression and linear programming, the two main stalwarts of today’s card builders, were introduced.*”).

- $WCTA = \text{capital circulante líquido} / \text{total dos ativos}$;
- $CLCA = \text{passivo circulante} / \text{ativo circulante}$;
- $OENEG = 1$ se o total dos passivos exceder o total dos ativos, e 0, caso contrário;
- $NITA = \text{lucro líquido} / \text{total dos ativos}$;
- $FUTL = \text{fluxo de caixa operacional} / \text{exigibilidades}$;
- $INTWO = 1$ se o lucro líquido foi negativo nos últimos dois anos, e 0, caso contrário;
- $CHIN = (NT_t - NT_{t-1}) / (|NT_t| + |NT_{t-1}|)$, onde NT_t é o lucro líquido do período mais recente (pretendeu-se medir alterações no lucro líquido).

Os sinais esperados dos coeficientes dessas variáveis, na função logística, são os seguintes:

- Positivo: TLTA, CLCA e INTWO;
- Negativo: SIZE, WCTA, NITA, FUTL e CHIN;
- Indeterminado: OENEG.

Foram estimados três modelos com essas variáveis independentes:

- Modelo 1: prevê falência no período de 1 ano;
- Modelo 2: prevê falência no período de 2 anos, se não faliu no primeiro ano;
- Modelo 3: prevê falência no período de 1 ou 2 anos.

Como já era esperado, por causa do perfil do tempo até a falência (média aproximada de 13 meses), o Modelo 1 apresentou os melhores resultados. Os coeficientes de todas as variáveis apresentaram os sinais previstos. Apenas as variáveis WCTA, CLCA e INTWO não possuíam coeficientes estatisticamente significativos em um nível relevante. O percentual de casos corretamente previstos foi de 96,1% (valor de corte de 0,5). Verificou-se que os coeficientes das variáveis de situação financeira não são correlacionados com os das variáveis de desempenho, sugerindo que ambos os conjuntos de variáveis são importantes no modelo.

Analisando os resultados obtidos, quatro fatores obtidos das demonstrações financeiras são estatisticamente significativos para avaliar a probabilidade de falência:

- tamanho (SIZE);

- estrutura financeira representada pela medida de alavancagem (TLTA);
- alguma medida de desempenho ou combinação de medidas de desempenho (NITA e/ou FUTL);
- alguma(s) medida(s) de liquidez corrente (WCTA ou WCTA e CLCA juntas).

Foi implementado um quarto modelo, tendo como base o Modelo 1, com acréscimo de medidas de margem de lucro (fluxo de caixa das operações / vendas; ativos diferidos e intangíveis / ativo total). Intencionava-se melhorar os resultados do Modelo 1, mas isso não ocorreu.

Ohlson explicou que seria possível dividir a amostra em duas para proceder ao tipo usual de validação do modelo, mas, como o principal propósito do estudo não era o de obter uma avaliação precisa de um modelo preditivo, decidiu-se pela utilização da amostra inteira para produzir os menores erros na estimação dos coeficientes.

2.11.2 Casey e Bartczak

Casey e Bartczak (1985) conduziram o estudo para avaliar se os dados de fluxo de caixa operacional e indicadores relacionados tornariam as previsões de insolvência mais precisas. Em um estudo anterior dos próprios autores, concluiu-se que modelos discriminantes multivariados com base no regime de competência tinham maior poder de previsão de insolvência do que qualquer modelo univariado utilizando indicadores de fluxo de caixa operacional. Assim, partindo-se de um modelo com os mesmos indicadores do regime de competência utilizados, avaliou-se o efeito preditivo marginal de um modelo com um ou mais indicadores de fluxo de caixa operacional, além desses indicadores do regime de competência. Foram utilizadas a análise discriminante linear múltipla e a análise logística, e, em ambos os casos, utilizou-se o procedimento de dividir a amostra em duas partes de igual tamanho para validação dos modelos.

A amostra era composta por 60 empresas que tiveram o pedido de falência decretado entre 1971 e 1982, e 230 empresas não falidas, que possuíam a mesma classificação industrial das empresas falidas. Não houve preocupação com o tamanho das empresas, com o intuito de não limitar a generalização dos resultados do estudo.

As variáveis do regime de competência foram:

- caixa / ativo total;
- ativo circulante / ativo total;
- ativo circulante / passivo circulante;
- vendas / ativo circulante;
- lucro líquido / ativo total;
- exigibilidades / patrimônio líquido.

As variáveis do fluxo de caixa operacional foram:

- fluxo de caixa operacional;
- fluxo de caixa operacional / passivo circulante;
- fluxo de caixa operacional / exigibilidades.

Os resultados sugerem que dados de fluxo de caixa operacional não incrementam o poder preditivo dos indicadores do regime de competência. Além disso, devido aos níveis significantes de multicolinearidade, não se tentou avaliar a contribuição individual das variáveis independentes nos modelos multivariados. Salientam que é questionável ignorar a multicolinearidade e fazer essa avaliação, podendo ser até enganadora.

2.11.3 Westgaard e Wijst

Westgaard e Wijst (2001) focaram seu estudo na estimação da frequência esperada de *default*⁶⁴, ao nível de cliente, de uma carteira de clientes *corporate*. Eles utilizaram dados cadastrais, contábeis e de falência de empresas norueguesas no período de 1995 a 1999, para construir *proxies* de variáveis da teoria de finanças que compuseram o modelo em questão.

As variáveis explicativas utilizadas foram:

- CASHDEBT: (resultado operacional + depreciação)/endividamento total, como *proxy* de entradas de caixa;

⁶⁴ “*Expected default frequency*”, que seria equivalente à PD.

- FINANCOV: resultado líquido antes dos custos financeiros/custos financeiros, como *proxy* de folga em serviços da dívida;
- LIQUIDIT: ativo circulante/passivo circulante, como *proxy* de valor da firma no curto prazo;
- SOLIDITY: patrimônio líquido/capital, como *proxy* de valor da firma no longo prazo;
- AGE: idade da empresa em anos, como relacionada com a distribuição de fluxo de caixa;
- SIZE: tamanho da empresa, utilizando logaritmo do ativo total da empresa, como relacionada com a distribuição do fluxo de caixa.

Além dessas variáveis, foram utilizadas variáveis dicotômicas para atividade econômica (REALSERV e HOTRESTA) e localização geográfica (MNORWAY e NNORWAY).

A hipótese testada foi a de que as variáveis explicativas (exceto as dicotômicas) têm uma influência negativa na probabilidade de falência, o que é esperado em função da teoria de finanças.

A técnica estatística utilizada foi a regressão logística. Essa técnica foi escolhida em detrimento da regressão *probit*, pois, segundo os autores, os seus coeficientes são mais fáceis de interpretar. A técnica de análise discriminante foi descartada por não resultar em probabilidades e pela dificuldade de testar a hipótese, pois os coeficientes das funções discriminantes não são únicos.

Após analisarem diversas combinações, aquela com os dados contábeis de 1996 e os dados de falência de 1998 foi a que apresentou os melhores resultados. De uma base de dados inicial de 70.574 empresas em 1996, das quais 1.989 registravam falência em 1998, metade foi selecionada para compor a amostra de estimação (35.287 empresas, das quais 954 faliram nesse período), e a outra metade compôs a amostra de verificação.

Todas as variáveis mostraram-se significativas ao nível de 5%, exceto a variável dicotômica HOTRESTA, e a hipótese nula de que as variáveis explicativas (exceto as dicotômicas) têm uma influência negativa na probabilidade de falência não pôde ser rejeitada para os dados utilizados.

O modelo proporcionou alguma capacidade na discriminação entre empresas falidas e não falidas e os resultados apresentaram uma forte relação entre falência e as variáveis explicativas utilizadas no modelo, mostrando que o tipo de modelo *logit* pode ser utilizado para o cálculo direto de PDs nos modelos de risco de crédito.

As sugestões para próximos estudos foram: incluir fatores de risco macroeconômicos, trabalhar ao nível de carteira de clientes e desenvolver modelos para pequenas empresas e clientes *private*.

2.11.4 Brito

Brito (2005), como parte de sua pesquisa, desenvolveu um modelo que quantifica o risco de *default* da empresa. Para ele, o conceito de evento de *default* adotado foi o início de um procedimento legal de concordata ou falência, pois era um tipo de informação disponível publicamente e suficiente para atender aos propósitos do estudo. O horizonte de tempo considerado foi de um ano, ou seja, o modelo quantifica a probabilidade de a empresa falir ou se tornar concordatária no período de um ano.

As variáveis consideradas no modelo eram de natureza quantitativa, associadas à situação econômico-financeira das empresas. Partiu-se de 25 índices freqüentemente utilizados em estudos dessa natureza, sem buscar identificar novos indicadores.

A população definida para o estudo era composta por empresas de capital aberto não financeiras, delimitando o período de pesquisa de 1994 a 2004. A amostra utilizada compreendeu 60 empresas, sendo 30 insolventes e 30 solventes, estas últimas emparelhadas com as primeiras pelo critério de mesmo setor econômico e tamanho equivalente.

Após a realização de teste-t (teste de igualdade de médias), 17 índices econômico-financeiros apresentaram diferenças estatisticamente significativas. Dentre os índices excluídos, ou seja, aqueles que não são relevantes para o modelo, estão os três índices relacionados ao fluxo de caixa: fluxo de caixa operacional sobre ativo, fluxo de caixa operacional sobre exigível total e fluxo de caixa sobre endividamento financeiro.

Utilizando a regressão logística como técnica de análise de dados, obteve-se o seguinte modelo final:

$$\ln\left(\frac{p}{1-p}\right) = -4,535 - 5,152 * X_{12} + 19,069 * X_{16} - 13,364 * X_{19} - 17,488 * X_{22}$$

onde:

- p : probabilidade de a empresa se tornar insolvente;
- X_{12} : lucros retidos sobre ativo, ou seja, (patrimônio líquido – capital social) / ativo total;
- X_{16} : endividamento financeiro, ou seja, (passivo circulante financeiro + exigível a longo prazo financeiro) / ativo total;
- X_{19} : capital de giro líquido, ou seja, (ativo circulante – passivo circulante) / ativo total;
- X_{22} : saldo de tesouraria sobre vendas, ou seja, (ativo circulante financeiro – passivo circulante financeiro) / ativo total.

A matriz de classificação do modelo apresentou taxa de acerto global de 91,7%, as variáveis explicativas não eram altamente correlacionadas e, ao nível de significância de 5%, não se rejeitou a hipótese nula de que não há diferenças significativas entre os valores preditos e observados (teste de Hosmer e Lemeshow). Justificou-se a não divisão da amostra em amostra de desenvolvimento e amostra de controle para validação do modelo devido ao pequeno tamanho da amostra (deveria ter pelo menos cem observações para justificar essa subdivisão). Como alternativa, utilizou-se o método *jackknife*⁶⁵ e a curva ROC (*Receiver Operating Characteristic*)⁶⁶, confirmando a capacidade preditiva do modelo desenvolvido.

⁶⁵ Segundo Servigny e Renault (2004, p. 112): “Esse procedimento é utilizado para reduzir o viés da taxa de erro aparente.” (“*This procedure is used to reduce the bias of the apparent error rate.*”), e “A taxa de erro aparente é a taxa de erro obtida através da utilização da amostra de desenvolvimento.” (2000, p. 425) (“*The apparent error rate is the error rate obtained by using the design set.*”).

⁶⁶ Segundo Servigny e Renault (2004, p. 94), é uma das abordagens mais utilizadas para medir o desempenho de modelos de escore de crédito.

3 METODOLOGIA

Como ponto de partida, é interessante buscar o significado da palavra método. Assim, apresenta-se o conceito dado por Marconi e Lakatos (2005, p. 83): “[...] é o conjunto das atividades sistemáticas e racionais que, com maior segurança e economia, permite alcançar o objetivo -conhecimentos válidos e verdadeiros -, traçando o caminho a ser seguido, detectando erros e auxiliando as decisões do cientista.”

Com base na tipologia apresentada por Gilberto Martins (2002, p. 34), este estudo pode ser classificado como sendo empírico-analítico, ou seja, caracterizado pela:

[...] utilização de técnicas de coleta, tratamento e análise dos dados marcadamente quantitativos. Privilegiam estudos práticos. Suas propostas têm caráter técnico, restaurador e incrementalista. Têm forte preocupação com relação causal entre variáveis. A validação da prova científica é buscada através de testes dos instrumentos, graus de significância e sistematização das definições operacionais.

Agora, para que se tenha uma visão geral de todo o processo de pesquisa, tomemos como base os passos da prática da Econometria, enumerados por Hill *et al.* (2000, p. 9):

1. Tudo começa com um problema – a falta de informação ou uma incerteza sobre um resultado, ou uma questão que envolve “o que ocorrerá se...”.
2. A teoria econômica nos dá uma forma de pensar no problema: que variáveis econômicas estão em jogo e qual é a direção possível do(s) relacionamento(s)? Como utilizaríamos uma nova informação se viéssemos a obtê-la – por exemplo, como utilizaríamos uma função produção para determinar a melhor combinação de fatores, o nível ótimo de produção ou a demanda por um insumo?
3. Essa informação é então disposta em termos de um modelo econômico eficiente que leve em conta nossas hipóteses fundamentais. Especificam-se as hipóteses de interesse.
4. O modelo econômico ativo conduz-nos ao modelo estatístico que descreve o processo pelo qual foram geradas as observações amostrais, à classificação de variáveis e à forma funcional do relacionamento.
5. Geram-se ou, mais comumente, coletam-se observações amostrais coerentes com o modelo econômico.
6. Dados o modelo estatístico e as observações amostrais, escolhe-se ou estabelece-se uma regra de estimação que tenha boas propriedades estatísticas.
7. Obtêm-se estimativas dos parâmetros desconhecidos com o auxílio de um pacote estatístico ou de um computador; fazem-se testes de hipóteses apropriados.
8. Analisam-se e avaliam-se as conseqüências estatísticas e econômicas e as implicações dos resultados empíricos. Por exemplo, as variáveis explicativas eram todas importantes? Utilizou-se a forma funcional correta? Que alocação de recursos econômicos e que resultados de distribuição estão implícitos, e quais suas implicações numa escolha política?
9. Se não se obteve convergência entre modelos econômico e estatístico e os dados amostrais, quais são os pontos dúbios em potencial, e quais são as sugestões para análise e avaliação futuras? Por exemplo, os dados terão sido inadequados para fundamentar as questões formuladas? As variáveis no modelo econômico foram classificadas corretamente e foram

apresentadas com os sinais e defasagens corretos? O modelo estatístico deveria ter envolvido não-linearidades tanto nas variáveis como nos parâmetros?

Os mesmos autores explicam a utilidade do modelo estatístico (2000, p. 4):

Num modelo estatístico, a primeira coisa a considerar é que as relações econômicas não são exatas. A teoria econômica não pretende ser capaz de prever o comportamento específico de um indivíduo ou firma; ao contrário, descreve o comportamento médio ou sistemático de muitos indivíduos ou firmas. Num estudo do consumo, devemos reconhecer que o consumo efetivo é a soma dessa parcela sistemática e de um componente aleatório e imprevisível, e que chamaremos erro aleatório.

Eles acrescentam que: “A inclusão dos erros aleatórios transforma nosso modelo econômico em um modelo estatístico que constitui uma fonte para a inferência estatística; isto é, dá base para estimarmos parâmetros desconhecidos e testarmos hipóteses sobre eles.” (HILL *et al.*, 2000, p. 5).

O SPSSTRaining (2003, p. 3-3) apresenta um breve relato sobre os procedimentos da construção de um modelo estatístico (no caso, uma regressão logística, também conhecida como *logit*):

Deve-se, primeiramente, selecionar um conjunto razoável de preditores, e examinar os dados atentamente antes da sua utilização, procurando por padrões não usuais, *outliers*⁶⁷, problemas com dados inexistentes, e assim por diante. Depois de estimar a equação e examinar os efeitos das variáveis individualmente, deve-se verificar se os dados atendem os pressupostos do modelo logístico e procurar por casos que tenham influência excessiva nos resultados. Se houver interesse em prever a classificação de um elemento no futuro, então é muito importante que o modelo seja validado. Isso significa construir um modelo com base num subconjunto dos dados e, então, testá-lo numa amostra de verificação⁶⁸.

Com relação aos dados, Hair Jr. *et al.* (2005, p. 3) comentam que as informações disponíveis para tomada de decisão explodiram nos últimos anos e deverão aumentar ainda mais no futuro. Muitas delas, no entanto, foram perdidas, por não terem sido coletadas, ou por terem

⁶⁷ Segundo Pestana e Gageiro (2003, p. 84): “Os *outliers* são observações aberrantes que podem existir numa distribuição de frequências e classificam-se como severos ou moderados consoante o seu afastamento em relação às outras observações seja mais ou menos pronunciado. Os *outliers* moderados encontram-se entre 1,5 e 3 amplitudes inter-quartis para baixo do primeiro quartil ou para cima do terceiro quartil. Os *outliers* são severos quando se encontram para valores superiores ou iguais a 3 amplitudes inter-quartis para baixo do primeiro quartil ou para cima do terceiro quartil.”

⁶⁸ “You must first select a reasonable set of predictors, and you must examine the data closely beforehand to look for unusual patterns, outliers, missing data problems, and so forth. After estimating the equation and examining the effect of individual variables, you should do a few checks to see whether the data meet the assumptions of the logistic model and look for cases that have undue influence on the results. If you are interested in predicting the category membership on the future data then it is very important that the model be validated. This means deriving model on a subset of the data and then testing it on the holdout sample.”

sido descartadas. Atualmente, essas informações estão sendo coletadas e armazenadas em grandes bancos de dados, e estão disponíveis para serem trabalhadas, a fim de aprimorar as tomadas de decisão.

3.1 Modelos e Técnicas Estatísticas

Servigny e Renault (2004, p. 107) apresentam vários fatores que são importantes na escolha de uma classe particular de modelo de mensuração de risco de crédito, em detrimento a uma outra classe:

- *Desempenho* – o fator primário. Um modelo mais complexo deveria fornecer detecção ou classificação de empresas arriscadas significativamente melhor se comparado a uma regra simples. Definir classificação ou previsão como o objetivo terá um impacto na escolha.
- *Disponibilidade de dados e qualidade*. Muitos modelos têm bom desempenho em laboratório, mas não suportam dificuldades práticas, tais como valores faltantes ou *outliers*. Um modelo mais simples, mas mais robusto pode ser preferível a um sistema “estado da arte” em conjunto de dados imperfeitos.
- *Entendimento dos usuários*. Usuários dos modelos de escore deveriam entender perfeitamente como o modelo funciona e o que norteia os resultados. Do contrário, eles não serão capazes de reconhecer vies sistemático ou entender as limitações do seu modelo.
- *A robustez do modelo para dados novos*. Alguns modelos simulados em um dado conjunto de dados podem gerar diferentes resultados se o conjunto de dados for aumentado. Essa instabilidade deve ser evitada, pois pode indicar que o modelo está detectando o ótimo local em vez do ótimo global.
- *O tempo requerido para calibrar ou recalibrar o modelo*. Isso também importa, dependendo da frequência que o usuário quer usar para rodar o modelo.⁶⁹

Percorrendo a literatura existente sobre modelos de previsão de *default*, destacam-se, pela grande predominância, os modelos que utilizam a análise discriminante e a regressão logística. No entanto, novos modelos, mais sofisticados, têm surgido nesses últimos tempos, tais como aqueles que utilizam redes neurais, árvores de decisão, algoritmos genéticos etc.,

⁶⁹ “• *Performance – the primary factor. A more complex model should provide significantly improved detection or classification of risky firms over a naïve rule. Defining classification or prediction as the goal will have an impact on the choice.*

• *Data availability and quality. Many models perform well in the laboratory but do not cope with practical difficulties such as missing values or outliers. A simpler but more robust model may be preferable to a state-of-the-art system on patchy data sets.*

• *Understanding by users. Users of the scoring models should understand perfectly how the model works and what drives the results. Otherwise they will not be able to spot systematic bias or to understand the limits of their model.*

• *The robustness of the model to new data. Some models trained on a given data set will provide very different results if the data set is increased slightly. This instability should be avoided, as it may mean that the model is detecting local optima rather than global optima.*

• *The time required to calibrate or recalibrate the model. This also matters, depending on what frequency the user wants to use to run the model.”*

em função da grande importância que o assunto tem adquirido.

Neste trabalho, a técnica estatística utilizada é a regressão logística. E, quando se trata de regressão, dois aspectos devem ser esclarecidos. Um deles é apresentado por Gujarati (2000, p. 8):

Embora a análise de regressão lide com a dependência de uma variável em relação a outras variáveis, ela não implica necessariamente causalção. Nas palavras de Kendall e Stuart, "uma relação estatística, por mais forte e sugestiva que seja, jamais pode estabelecer uma relação causal: nossas idéias sobre causalção devem vir de fora da estatística, enfim, de outra teoria."

Essa diferença deve ser considerada para que não se tirem conclusões equivocadas dos resultados da regressão, ou seja, as variáveis explicativas não são necessariamente as causadoras dos efeitos na variável independente.

O outro aspecto refere-se à correlação: "Intimamente relacionada, porém conceitualmente muito diferente da análise de regressão, é a análise de correlação, cujo objetivo básico é medir a intensidade ou o grau de associação linear entre duas variáveis." (GUJARATI, 2000, p. 9). Ou seja, na regressão, não se está necessariamente interessado na medição do grau de associação entre as variáveis, mas na estimativa ou previsão do valor médio de uma variável (dependente) em função dos valores fixados de outras variáveis (independentes).

3.2 Análise Discriminante x Regressão Logística

Conforme Hair Jr. *et al.* (2005, p. 273), na tentativa de se escolher uma técnica analítica, pode-se deparar com um problema que envolva uma variável dependente categórica e diversas variáveis independentes métricas. Utilizando o exemplo de distinção entre risco de crédito bom ou ruim, se o risco de crédito for uma medida métrica, pode-se utilizar a regressão linear múltipla. Mas, se a medida não for métrica, ou seja, se somente for possível afirmar que o risco de crédito é bom ou ruim (se pertence a um determinado grupo), não é possível utilizar a regressão linear (pois a variável dependente deve ser métrica). Quando a variável dependente é categórica (nominal ou não métrica), a análise discriminante e a regressão logística são as técnicas estatísticas apropriadas. No entanto, a análise discriminante

é capaz de trabalhar com dois ou mais grupos, mas a regressão logística é limitada, na sua forma básica⁷⁰, a dois grupos.

Os autores (HAIR JR. *et al.*, p. 355) explicam que, mesmo quando a variável dependente possui apenas dois grupos, a regressão logística pode ser preferida à análise discriminante por dois motivos:

- A análise discriminante depende do atendimento rígido dos pressupostos de normalidade multivariada e igualdade das matrizes de variância-covariância dos grupos – pressupostos que não são atendidos em muitas situações. A regressão logística não possui essas restrições e é muito mais robusta quando esses pressupostos não são atendidos, tornando-a apropriada em muitas situações.
- Mesmo quando os pressupostos são atendidos, alguns pesquisadores preferem a regressão logística porque ela é similar à regressão linear múltipla. Ela tem testes estatísticos simples, abordagens similares para incorporar variáveis métricas e não métricas e efeitos não lineares, e uma ampla variedade de diagnósticos.⁷¹

Por essas razões técnicas, a regressão logística é equivalente à análise discriminante de dois grupos e pode ser mais adequada em várias situações.

Nesse sentido, no estudo de Westgaard e Wijst (2001, p. 346), afirma-se que não seria possível utilizar a análise discriminante, já que ela não gera probabilidade como resultado. Assim, a análise logística e a *probit* seriam as “candidatas mais óbvias” para o modelo.

Especificamente para o setor bancário, Bandyopadhyay (2006, p. 258) apresenta a regressão logística como solução para os bancos que pretendem estimar a PD diretamente, pois, diferentemente da análise discriminante, além de produzir probabilidade como resultado, tem a flexibilidade de incorporar variáveis financeiras e não financeiras.

⁷⁰ Segundo os autores, outras formulações de regressão logística podem trabalhar com mais de dois grupos (categorias). Hosmer e Lemeshow (2000, p. 260) apresentam o modelo de regressão linear multinomial (onde a variável dependente é nominal e possui mais de dois níveis).

⁷¹ “• *Discriminant analysis relies on strictly meeting the assumptions of multivariate normality and equal variance-covariance matrices across groups – assumptions that are not met in many situations. Logistic regression does not face these strict assumptions and is much more robust when these assumptions are not met, making its application appropriate in many situations.*

• *Even if the assumptions are met, many researchers prefer logistic regression because it is similar to multiple regression. It has straightforward statistical tests, similar approaches to incorporating metric and nonmetric variables and nonlinear effects, and a wide range of diagnostics.”*

3.3 Regressão Logística x Regressão Probit

Gujarati (2000, p. 573) compara as duas técnicas estatísticas:

[...] as formulações logística e *probit* são bem comparáveis, sendo que a principal diferença está no fato de a logística ter caudas ligeiramente mais achatadas, ou seja, a curva normal (ou *probit*) se aproxima do eixo mais rapidamente do que a curva logística. Portanto, a escolha entre as duas é uma questão de conveniência (matemática) e de pronta disponibilidade de programas de computador. Neste aspecto, o modelo *logit* é em geral preferido ao *probit*.

No Gráfico 5, é possível verificar essa característica graficamente:

Gráfico 5 - Logit x Probit
FONTE: GUJARATI, 2000, p. 574

O SPSS (*Statistical Package for Social Science* - Pacote Estatístico para Ciências Sociais)⁷², no mecanismo de Ajuda (*Help*) do sistema, de maneira similar, explica:

A análise *probit* está estreitamente relacionada à regressão logística; de fato, caso se escolha a transformação *logit*, esse procedimento essencialmente calculará a regressão logística. Em geral, a análise *probit* é apropriada para experimentos projetados, enquanto a regressão logística é mais apropriada para estudos observacionais. As diferenças nas saídas refletem essas diferentes ênfases. O procedimento da análise *probit* reporta estimativas de valores efetivos para várias taxas de respostas (inclusive mediana da quantidade efetiva), enquanto o procedimento da regressão logística reporta estimativas das chances das variáveis independentes⁷³.

Segundo Servigny e Renault (2004, p. 80), quando se referem às regressões *logit* e *probit*:

⁷² Versão 15.

⁷³ “*Probit analysis is closely related to logistic regression; in fact, if you choose the logit transformation, this procedure will essentially compute a logistic regression. In general, probit analysis is appropriate for designed experiments, whereas logistic regression is more appropriate for observational studies. The differences in output reflect these different emphases. The probit analysis procedure reports estimates of effective values for various rates of response (including median effective dose), while the logistic regression procedure reports estimates of odds ratios for independent variables.*”

Em ambos os casos, a estimativa dos parâmetros é realizada pela máxima verossimilhança. A principal diferença entre as duas abordagens é relacionada ao fato de a distribuição logística ter as caudas mais espessas do que a distribuição normal [...] Entretanto, isso não fará uma grande diferença em aplicações práticas desde que não existam muitas observações extremas na amostra.⁷⁴

Como visto no estudo de Westgaard e Wijst (2001, p. 346), a análise logística e a *probit* seriam as “candidatas mais óbvias” para a regressão, por produzirem probabilidades como resultado. Segundo eles, apesar de os modelos serem muito parecidos e dificilmente produzirem conclusões qualitativamente diferentes, a regressão logística foi escolhida por ser linear na conversão logarítmica, o que faz com que os seus coeficientes sejam, de certa maneira, mais fáceis de interpretar.

3.4 Regressão Logística

Segundo Hosmer e Lemeshow (2000, p. 1):

Os métodos de regressão tornaram-se um componente essencial de qualquer análise de dados que se preocupa em descrever a relação entre uma variável dependente e uma ou mais variáveis explicativas. Frequentemente, a variável dependente é discreta, assumindo dois ou mais possíveis valores. Na última década, o modelo de regressão logística tornou-se, em muitos campos, o método padrão de análise nessa situação.⁷⁵

Os autores enfatizam a importância de compreender que o objetivo de uma análise utilizando esse método é o mesmo de qualquer técnica de modelagem usada em estatística: encontrar a melhor aderência e o modelo mais parcimonioso que esteja relacionado com a teoria, para descrever a relação entre uma variável dependente e um conjunto de variáveis independentes. O exemplo mais comum de modelagem é o modelo de regressão linear, cuja variável dependente é considerada como contínua. Assim, o que distingue o modelo de regressão logística do modelo de regressão linear é que a variável dependente da regressão logística é binária ou dicotômica. Essa diferença é refletida tanto na escolha de um modelo paramétrico quanto em seus pressupostos. Desconsiderando essa diferença, os métodos aplicados em uma

⁷⁴ “In both cases the estimation of the parameters is performed by maximum likelihood. The main difference between the two approaches is linked to the fact that the logistic distribution has thicker tails than the normal distribution [...] However, this will not make a huge difference in practical applications as long as there are not too many extreme observations in the sample.”

⁷⁵ “Regression methods have become an integral component of any data analysis concerned with describing the relationship between a response variable and one or more explanatory variables. It is often the case that the outcome variable is discrete, taking on two or more possible values. Over the last decade the logistic regression model has become, in many fields, the standard method of analysis in this situation.”

análise usando a regressão logística seguem os mesmos princípios gerais utilizados em uma regressão linear. Deste modo, as técnicas usadas na análise com regressão linear motivam a abordagem da regressão logística.

Na mesma linha, Hair Jr. *et al.* (2005, p. 356) expõem que a regressão logística difere da regressão linear por prever diretamente a probabilidade de ocorrência de um evento (ou seja, a probabilidade de uma observação estar no grupo denominado “1”).

Com base em vários estudos sobre mensuração de riscos, Dias Filho (2003, p. 203) observou os principais fatores que guiaram a sua opção pela regressão logística:

- a) comparada a outras técnicas de dependência, a regressão logística acolhe com mais facilidade variáveis categóricas. Aliás, esta é uma das razões pelas quais ela se torna uma boa alternativa à análise discriminante, sobretudo quando o pesquisador se defronta com problemas relacionados à variância;
- b) mostra-se mais adequada à solução de problemas que envolvem estimação de probabilidades, pois trabalha com uma escala de resultados que vai de zero a um;
- c) requer um menor número de suposições iniciais, se comparada com outras técnicas utilizadas para discriminar grupos;
- d) admite variáveis independentes métricas e não-métricas, simultaneamente;
- e) facilita a construção de modelos destinados a previsão de riscos em diversas áreas do conhecimento. Os chamados *Credit Scoring* e tantos outros que são utilizados no contexto da análise de sobrevivência ilustram essa realidade;
- f) tendo em vista que o referido modelo é mais flexível quanto às suposições iniciais, tende a ser mais útil e a apresentar resultados mais confiáveis;
- g) os resultados da análise podem ser interpretados com relativa facilidade, já que a lógica do modelo se assemelha em muito à de outras técnicas bem conhecidas, como a regressão linear.

Saunders (2000, p. 211) aponta uma restrição: “Sua principal deficiência é a hipótese de que a probabilidade acumulada de inadimplência assume determinada forma funcional que corresponde à função logística.” Por outro lado, o SPSS TRAINING (2003, p. 3-2) expõe que, quando se prevê o valor de uma variável que varia de 0 a 1, faz sentido ajustar os dados a uma curva em formato de “S”, tal como é a curva logística, *probit*, entre outras da mesma família. Ainda, Johnston e Dinardo (1997, p. 430) explicam que, quando tratam de modelos com variáveis dependentes discretas e limitadas, não é menos “heróico” assumir que a função tem uma forma especial do que assumir que ela é linear.

O SPSS TRAINING (2003, p. 3-3) apresenta os dois objetivos gerais da regressão logística:

1. Determinar o efeito de um conjunto de variáveis independentes na probabilidade, mais o efeito

individual das variáveis;

2. Obter a máxima acurácia preditiva possível dado um conjunto de preditores.⁷⁶

Expõe, ainda, que uma análise tende a focar em um dos dois objetivos, apesar de eles não serem incompatíveis: “Aqueles interessados em teoria e efeitos causais tipicamente estão mais preocupados com o primeiro objetivo; aqueles interessados em prever se um evento futuro se enquadrará em uma ou outra categoria da variável dependente focarão no segundo objetivo.”⁷⁷ (SPSSTRaining, 2003, p. 3-3).

Neste estudo, o foco é a obtenção da máxima acurácia preditiva possível, dado um conjunto de variáveis explicativas.

3.4.1 Descrição do Modelo

Tomando-se a regressão linear múltipla com o intuito de criar um modelo cuja variável dependente seja uma variável categórica de dois estados (0, 1), ou uma variável métrica que represente probabilidade (varia de 0 a 1), depara-se com um problema: a regressão linear não permite restringir os valores da variável dependente.

Assim, com o intuito de melhor entender o modelo da regressão logística, parte-se de uma regressão linear múltipla:

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_n X_{ni} + \mu_i$$

onde:

Y_i é a variável dependente;

$X_{1i}, X_{2i}, \dots, X_{ni}$ são as variáveis explicativas, ou independentes;

α é o intercepto da reta;

$\beta_1, \beta_2, \dots, \beta_n$ são os parâmetros desconhecidos; e

μ_i é o erro aleatório.

⁷⁶ “1. Determine the effect of a set of variables on the profitability, plus the effect of individual variables
2. Attain the highest predictive accuracy possible with a given set of predictors.”

⁷⁷ “Those interested in theory and causal effects typically are more concerned with the first goal; those concerned with predicting whether a future event will fall in one or other category of the dependent variable focus on the second goal.”

Dado que $X_{1i} \dots X_{ni}$ variam de $-\infty$ a $+\infty$, temos que: $-\infty \leq Y_i \leq +\infty$. Assim, em princípio, a regressão linear múltipla não seria adequada, já que a probabilidade deve restringir-se ao intervalo entre 0 e 1.

Por outro lado, temos que a função logística atende à necessidade de restringir o resultado na faixa de valores de 0 a 1, como se pode ver no Gráfico 6:

Gráfico 6 - Função logística

FONTE: Adaptado de HAIR JR. *et al.*; 2005, p. 356

Deste modo, uma solução para esse problema pode ser obtida através das seguintes transformações:

$$Chance = \frac{P(evento)}{1 - P(evento)}$$

onde:

$P(evento)$ é a probabilidade de o evento ocorrer;

$Chance$ é a razão entre as probabilidades de o evento ocorrer e de não ocorrer.

Com isso, temos que: se $0 < P(evento) < 1$, então $0 < Chance < +\infty$.

Se aplicarmos o logaritmo natural da *Chance*, temos, então: $-\infty < \ln Chance < +\infty$.

Com essas transformações, é possível, a partir de uma variável, $P(evento)$, com valores restritos à faixa de 0 a 1, gerar uma outra variável, $\ln Chance$, sem restrição de valores, sendo possível utilizá-la como a variável dependente em uma regressão múltipla linear, da seguinte maneira:

$$\ln Chance_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_n X_{ni} + \mu_i$$

que, também, pode ser apresentada da seguinte forma:

$$\ln \left[\frac{P_i(evento)}{1 - P_i(evento)} \right] = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_n X_{ni} + \mu_i$$

onde o termo $\ln \left[\frac{P_i(evento)}{1 - P_i(evento)} \right]$ é chamado de *logit* (logaritmo da chance), origem da outra denominação da regressão logística: regressão *logit*.

Assim, o *logit* é não somente linear em X_i , mas também (do ponto de vista da estimativa) é linear nos parâmetros $\beta_1, \beta_2, \dots, \beta_n$.

Outra forma de escrever a mesma equação, mas em função da *Chance* seria:

$$Chance_i = \frac{P_i(evento)}{1 - P_i(evento)} = e^{\alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_n X_{ni} + \mu_i}$$

Ou ainda, em função da probabilidade de ocorrer o evento:

$$P_i(evento) = \frac{1}{1 + e^{-(\alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_n X_{ni} + \mu_i)}}$$

O modelo logístico pode ser interpretado da seguinte maneira:

- $\beta_1, \beta_2, \dots, \beta_n$ correspondem às inclinações, ou seja, medem a variação em *logit* para uma variação unitária em X_i ;
- α é o intercepto, ou seja, o valor da *Chance*, em logaritmo natural, de ocorrer o evento quando as variáveis explicativas forem zero. Segundo Gujarati (2000, p. 560): “Assim, como a maioria das interpretações dos interceptos, esta interpretação pode não ter significado físico.”

3.4.2 Pressupostos do Modelo

Segundo o SPSS TRAINING (2003, p. 3-4), a regressão logística requer os seguintes pressupostos:

1. As variáveis independentes são intervalares, razões ou dicotômicas;
2. Todos os preditores relevantes estão incluídos, nenhum preditor irrelevante está incluído, e a forma do relacionamento é linear;
3. O valor esperado do termo de erro é zero;
4. Não há autocorrelação;
5. Não há correlação entre o erro e as variáveis independentes;
6. Há ausência de multicolinearidade perfeita entre as variáveis independentes.⁷⁸

Além da menor quantidade de pressupostos necessários em relação à regressão linear, a utilização de um grande número de variáveis dicotômicas como preditoras não viola nenhum pressuposto da regressão logística, diferentemente do que ocorre na análise discriminante, o que a torna preferível nessas situações.

⁷⁸ “1. The independent variables be interval, ratio, or dichotomous;

2. All relevant predictors be included, no irrelevant predictors be included, and the form of the relationship is linear;

3. The expected value of the error term is zero;

4. There is no autocorrelation;

5. There is no correlation between the error and the independent variables;

6. There is an absence of perfect multicollinearity between the independent variables.”

3.4.3 Estimação do Modelo

A seguinte observação é apresentada por Hosmer e Lemeshow (2000, p.8):

O método de estimação geral que leva à função dos mínimos quadrados no modelo de regressão linear (quando os termos de erro são normalmente distribuídos) é chamado de máxima verossimilhança. Esse método fornece a base para nossa abordagem à estimação para o modelo de regressão logística. De um modo bem geral, o método da máxima verossimilhança gera valores para os parâmetros desconhecidos que maximizam a probabilidade de obter o conjunto de dados observados. Para aplicar esse método, deve-se, primeiramente, construir uma função, chamada função de verossimilhança. Essa função expressa a probabilidade dos dados observados como uma função dos parâmetros desconhecidos. Os estimadores de máxima verossimilhança desses parâmetros são escolhidos de modo a maximizarem essa função. Deste modo, os estimadores resultantes são aqueles que mais se ajustam aos dados observados.⁷⁹

Ou seja, o método dos mínimos quadrados nada mais é do que um caso específico do método da máxima verossimilhança. Na regressão linear, as equações de verossimilhança obtidas são lineares nos parâmetros desconhecidos e, portanto, simples de resolver. Mas, na regressão logística, as equações são não-lineares nos parâmetros, sendo necessário utilizar métodos iterativos (HOSMER; LEMESHOW, 2000, p. 9).

Com relação à escolha das variáveis explicativas para comporem o modelo estatístico, Hosmer e Lemeshow (2000, p. 92) sugerem:

O critério para incluir uma variável em um modelo pode variar de um problema para outro e de uma disciplina científica para outra. A abordagem tradicional para a construção do modelo estatístico consiste em procurar pelo modelo mais parcimonioso que ainda explica os dados. A razão de minimizar o número de variáveis no modelo é que o modelo resultante tem maior probabilidade de ser numericamente estável, e é mais fácil de generalizar. Quanto mais variáveis são incluídas no modelo, maiores se tornam os desvios padrões estimados, e mais dependente o modelo se torna dos dados observados [...] O superajustamento é tipicamente caracterizado por coeficientes e/ou desvios padrões estimados muito grandes. Isso pode ser especialmente problemático em casos onde o número de variáveis no modelo é grande em relação ao número de elementos e/ou quando a proporção de eventos no total ($y=1$) é próxima ou de 0 ou de 1.⁸⁰

⁷⁹ “The general method of estimation that leads to the least squares function under the linear regression model (when the error terms are normally distributed) is called maximum likelihood. This method will provide the foundation for our approach to estimation with the logistic regression model. In a very general sense the method of maximum likelihood yields values for the unknown parameters which maximize the probability of obtaining the observed set of data. In order to apply this method we must first construct a function, called the likelihood function. This function expresses the probability of the observed data as a function of the unknown parameters. The maximum likelihood estimators of these parameters are chosen to be those values that maximize this function. Thus, the resulting estimators are those which agree most closely with the observed data.”

⁸⁰ “The criteria for including a variable in a model may vary from one problem to the next and from one scientific discipline to another. The traditional approach to statistical model building involves seeking the most parsimonious model that still explains the data. The rationale for minimizing the number of variables in the model is that the resultant model is more likely to be numerically stable, and is more easily generalized. The more variables included in a model, the greater the estimated standard errors become, and the more dependent the model becomes on the observed data [...] Overfitting is typically characterized by unrealistically large estimated coefficients and/or estimated standard errors. This may be especially troublesome in problems where

Apesar da disponibilidade de um grande número de registros para a modelagem, o baixo percentual de *default* causa preocupação.

Para a análise dos modelos de estimativas por máxima verossimilhança, Wooldridge (2002, p. 461) apresenta três testes: teste Wald, teste da razão de verossimilhança e teste do multiplicador de Lagrange; e explica que a escolha do teste normalmente depende da simplicidade computacional. Especificamente sobre o teste da razão de verossimilhança, Wooldridge (2006, p. 588) expõe que este teste torna-se atraente quando os modelos irrestrito e restrito são fáceis de estimar. A sua base está na diferença das funções log-verossimilhança (*log-likelihood*) dos modelos irrestrito e restrito. Como a estimativa por máxima verossimilhança maximiza a função log-verossimilhança, a exclusão de variáveis geralmente leva a uma log-verossimilhança menor do que a do modelo irrestrito. Assim, para determinar se as variáveis excluídas são importantes, deve-se determinar a estatística de teste e um conjunto de valores críticos. A estatística teste é o dobro da diferença das log-verossimilhanças, que apresenta uma distribuição Qui-quadrado aproximada, sob a hipótese nula de que os coeficientes das variáveis excluídas são nulos, com os graus de liberdade representados pelo número de variáveis excluídas.

O SPSS TRAINING (2003, p. 3-18) apresenta o procedimento *stepwise* de seleção do modelo, que indica qual é o melhor subconjunto de variáveis explicativas que são boas preditoras da variável dependente. Hosmer e Lemeshow (2000, p. 117) sugerem esse procedimento quando a variável dependente é relativamente nova e as variáveis independentes não são bem conhecidas, e as suas associações não são bem entendidas. Eles apresentam como vantagem a rapidez na análise de um grande número de variáveis e o ajuste simultâneo de várias equações e explicam que qualquer passo, tanto de inclusão como de exclusão de variáveis no modelo, é baseado em um algoritmo estatístico que verifica a relevância dessas variáveis. Um aspecto classificado como crucial é a determinação do nível de significância (α), tanto na inclusão como na exclusão de uma variável (esses dois dados devem ser estabelecidos).

Um aspecto levantado pelo SPSS TRAINING (2003, p. 3-18) com relação ao procedimento *stepwise* é que os algoritmos buscam a maximização da verossimilhança, o que não significa a

the number of variables in the model is large relative to the number of subjects and/or when the overall proportion responding ($y=1$) is close to either 0 or 1."

maximização da acurácia preditiva. Assim, especialmente quando o foco da análise for a acurácia preditiva, a validação é importante quando se utiliza desse procedimento e existem dados suficientes para tal teste. Como critério de eliminação da variável no procedimento *stepwise*, sugere-se a variação da razão de verossimilhança (*likelihood ratio*)⁸¹.

Como este estudo trata de uma variável dependente não muito conhecida (evento de inadimplência no SFN), tal como as relações com as variáveis independentes (poucos estudos disponíveis), o procedimento *stepwise* seria uma opção cômoda, mas, devido a restrições em seu trato com as variáveis categóricas⁸², que são importantes neste estudo, entende-se que a melhor alternativa consiste em trabalhar manualmente na seleção das variáveis explicativas, utilizando o teste da razão de verossimilhança.

3.4.4 Avaliação do Ajuste do Modelo

Após a construção do modelo, é necessário verificar quão bem ele consegue estimar a variável dependente (PD). Esse procedimento pode ser chamado de teste de qualidade do ajuste⁸³ do modelo. Hosmer e Lemeshow (2000, p. 155) mostram que essa tarefa não é simples:

[...] uma completa avaliação do ajuste é uma investigação multifacetada que envolve testes resumos e medidas, como também estatísticas de diagnóstico. É especialmente importante ter em mente quando utilizar testes de qualidade do ajuste globais. O resultado desejado para a maioria dos investigadores é a decisão de não rejeitar a hipótese nula de que o modelo se ajusta. Com essa decisão, pode-se submeter à possibilidade do erro Tipo II e, daí, o poder do teste torna-se um problema. [...] nenhum dos testes de qualidade do ajuste é especialmente poderoso para amostras de tamanho pequeno a moderado, ou seja, $n < 400$.⁸⁴

Wooldridge (2002, p. 465) expõe que uma das medidas da qualidade do ajuste do modelo que é usualmente apresentada é o percentual de predições corretas, ou seja, o percentual de casos

⁸¹ O procedimento de seleção de variáveis *stepwise* possui algumas variações (versão 15 do SPSS) que podem ser agrupadas em: *Forward* (utiliza a estatística de escore para seleção da variável, e três métodos de eliminação de variáveis: estatística Wald, variação na razão de verossimilhança e estatística condicional) e *Backward* (utiliza os mesmos três métodos de eliminação de variáveis do *Forward* (SPSS TRAINING, 2003, p. 3-18).

⁸² Não é possível fazer agrupamento de categorias no método *stepwise* (agrupar tipos de controle acionário, quando esses não são significantes isoladamente, mas são significantes quando considerados em conjunto, por exemplo).

⁸³ “Goodness-of-Fit Test”.

⁸⁴ “[...] a complete assessment of fit is a multi-faceted investigation involving summary tests and measures as well as diagnostic statistics. This is especially important to keep in mind when using overall goodness-of-fit tests. The desired outcome for most investigators is the decision not to reject the null hypothesis that the model fits. With this decision one is subjected to the possibility of the Type II error and hence the power of the test becomes a issue. [...] none of the overall goodness-of-fit is especially powerful for small to moderate sample sizes $n < 400$.”

em que o resultado predito está de acordo com o resultado observado. No entanto, em muitos casos é mais fácil prever um resultado do que o outro (como neste estudo), tornando essa estatística “enganadora”. Assim, o autor diz que é mais informativo apresentar os percentuais corretos de cada resultado do que o percentual correto global.

Hosmer e Lemeshow (2000, p. 147) apresentam os testes de Hosmer e Lemeshow, explicando que eles fazem agrupamentos baseados nos valores das probabilidades estimadas, geralmente chamados de “decis de risco”. Da comparação com os valores observados e utilizando a estatística do χ^2 de Pearson, obtém-se a estatística dos testes. Nesse caso, a hipótese testada é a de que não existem diferenças entre os valores preditos e observados. A vantagem destes testes é que eles geram um número simples e fácil de interpretar, que pode avaliar o ajuste do modelo⁸⁵. A grande desvantagem é que o processo de agrupamento pode ocultar um desvio importante devido a um pequeno número de dados individuais. Deste modo, eles sugerem que, antes de concluir que o modelo se ajusta, devem-se analisar os resíduos individuais. Nesse sentido, o SPSS TRAINING (2003, p. 3-7) aponta: “Medidas de influência ajudam a determinar como um caso influi nos resultados globais da regressão. As estatísticas residuais ajudam a determinar se um caso é um *outlier*. Técnicas gráficas podem ser usadas para encontrar casos que merecem mais estudo.”⁸⁶ Dentre as medidas de influência estão as de Cook, valores de influência⁸⁷ e DfBeta (medidas de influência), e, dentre as estatísticas residuais, estão o resíduo logit, o resíduo não padronizado e o resíduo padronizado.

Hosmer e Lemeshow (2000, p. 67) comentam sobre outras medidas resumo existentes, como os valores R^2 , que apresentam a seguinte característica:

Infelizmente, valores de R^2 baixos na regressão logística são a norma e isso representa um problema quando se reportam os seus valores a uma audiência acostumada a ver valores de regressão linear. [...] Assim, não se recomenda publicar rotineiramente os valores de R^2 com os resultados dos modelos ajustados de regressão linear. Entretanto, eles podem ser úteis no estágio de construção do modelo como uma estatística para avaliar modelos concorrentes.⁸⁸

⁸⁵ Segundo o SPSS TRAINING (2003, p. 3-7), esse teste de ajuste é mais robusto do que os testes tradicionais aplicados à regressão logística, especialmente nos casos com variáveis independentes contínuas e com amostras pequenas.

⁸⁶ “Influence measures help you to determine how influential a case is on the overall regression results. Residual statistics help you to determine whether a case is an outlier. Graphical techniques can be used with all of these to find cases that deserve more study.”

⁸⁷ “Leverage values”.

⁸⁸ “Unfortunately low R^2 values in logistic regression are the norm and this presents a problem when reporting their values to an audience accustomed to seeing linear regression values. [...] Thus we do not recommend routine publishing of R^2 values with results from fitted logistic regression models. However, they may be helpful

Podemos citar algumas medidas resumo: $-2LL^{89}$, Cox e Snell pseudo R^2 e Nagelkerke pseudo R^2 , sendo que a última tem a vantagem de possuir o valor máximo de “1”.

A avaliação do ajuste através de validação externa é apresentada por Hosmer e Lemeshow (2000, p. 186):

Em algumas situações, pode ser possível excluir uma subamostra de nossas observações, desenvolver um modelo baseado nos elementos remanescentes, e então testar o modelo nos elementos originalmente excluídos. Em outras situações, pode ser possível obter uma nova amostra de dados para avaliar a qualidade do ajuste de um modelo desenvolvido anteriormente. Esse tipo de avaliação é freqüentemente chamado de validação do modelo, e pode ser especialmente importante quando o modelo ajustado é utilizado para prever resultados de futuros elementos. A razão para considerar esse tipo de avaliação de desempenho do modelo é que o modelo ajustado sempre apresenta melhor desempenho no conjunto de dados de desenvolvimento.⁹⁰

Este método é muito citado na literatura e é de interesse para o nosso trabalho, que é focado na acurácia do resultado. No entanto, devido à necessidade de uma quantidade maior de elementos para compor as duas amostras (de desenvolvimento e de validação), torna-se necessário verificar se o tamanho da amostra e a quantidade de eventos existentes são suficientes para efetuar essa validação.

3.5 População, Amostras e Origem dos Dados

Sinteticamente, Gilberto Martins (2002, p. 43) conceitua população: “[...] trata-se do conjunto de indivíduos ou objetos que apresentam em comum determinadas características definidas para o estudo [...]”, e amostra: “[...] é um subconjunto da população.” Ainda, esclarece que:

É compreensível que o estudo de todos os elementos da população possibilita preciso conhecimento das variáveis que estão sendo pesquisadas; todavia, nem sempre é possível obter as informações de todos os elementos da população. Limitações de tempo, custo e as vantagens do uso das técnicas estatísticas de inferências justificam o uso de planos amostrais. Torna-se claro que

in the model building stage as a statistic to evaluate competing models.”

⁸⁹ Estatística da razão de verossimilhança.

⁹⁰ *“In some situations it may be possible to exclude a subsample of our observation, develop a model based on the remaining subjects, and then test the model in the originally excluded subjects. In other situations it may be possible to obtain a new sample of data to assess the goodness-of-fit of a previous developed model. This type of assessment is often called model validation, and may be especially important when the fitted model is used to predict outcome for future subjects. The reason for considering this type of assessment of model performance is that the fitted model always performs better in an optimistic manner on the developmental data set.”*

a representatividade da amostra dependerá do seu tamanho (quanto maior, melhor) e de outras considerações de ordem metodológica. Isto é, o investigador procurará acercar-se de cuidados, visando à obtenção de uma amostra significativa, ou seja, que de fato represente “o melhor possível” toda a população.

Com relação aos métodos de amostragem, Gilberto Martins (2005, p. 196) expõe que métodos não probabilísticos⁹¹ são aqueles em que “[...] há uma escolha deliberada dos elementos da amostra. Não é possível generalizar os resultados da amostra para a população, pois as amostras não probabilísticas não garantem a representatividade da população.” Dentre eles, encontra-se a amostragem intencional, ou seja, “De acordo com determinado critério, é escolhido intencionalmente um grupo de elementos que irão compor a amostra.” (MARTINS, Gilberto, 2005, p. 196).

Para estipular a população objeto do nosso estudo, considerou-se o exposto por Silva (2000, p. 277):

[...] a análise de crédito envolve variáveis quantitativas e qualitativas, ao mesmo tempo em que há uma tendência de que as empresas de maior porte apresentem demonstrações financeiras mais informativas [...] Na empresa muito pequena, as informações quantitativas tendem a ser deficientes.

Deste modo, decidiu-se trabalhar apenas com grandes empresas (*corporate*), para as quais as informações contábeis são, em princípio, mais confiáveis. Mais especificamente, a população de estudo é representada pelas empresas participantes das 500 Maiores, nos exercícios de 2002 a 2005, e a amostra utilizada é composta pelas empresas que possuem dados no SCR (tomadoras de crédito no SFN) no período de 12/2002 a 12/2006. Assim, classifica-se o método de amostragem como intencional, ou seja, não será possível generalizar os resultados para a população, tendo a sua validade restrita a esse contexto específico.

Quanto ao tamanho da amostra, o SPSS TRAINING (2003, p. 3-4) explica que, sob as mesmas condições, a regressão logística requer uma amostra maior do que a regressão linear. A regra prática de trabalhar com uma amostra com pelo menos 30 elementos para cada parâmetro a ser estimado será utilizada neste trabalho.

⁹¹ “Uma amostra é probabilística quando os elementos amostrais são escolhidos com probabilidades conhecidas.” (MARTINS, Gilberto, 2005, p. 21).

3.6 Descrição dos Métodos de Análise dos Dados

Para efeito das análises, considera-se cada elemento da amostra (equivalente a um registro) como a combinação de uma empresa e uma IF⁹² (CNPJ-IF)⁹³. Assim, a mesma empresa poderá ser considerada, ao mesmo tempo, adimplente e inadimplente, tendo em vista situações diversas em diferentes IFs.

3.6.1 Determinação da Parcela Material da Dívida no Evento de *Default*

Dada a definição do evento de *default*: “atraso superior a 90 dias de parcela material da dívida de uma empresa em relação a uma IF”, torna-se necessário estabelecer o significado do termo parcela material. Assim, neste trabalho, parcela material significa o percentual da carteira de crédito⁹⁴ em atraso acima de 90 dias, a partir do qual há chances consideráveis de piora (aumento do percentual de atraso) no decorrer do tempo (após três meses).

Para evitar trabalhar com carteiras de crédito não relevantes, ou seja, com valores absolutos pequenos que poderiam distorcer o resultado do modelo, decidiu-se por excluir da análise os casos onde os valores de carteira de crédito são inferiores a R\$ 5 mil⁹⁵.

Com a base de dados que atende ao que podemos chamar de “materialidade absoluta” (valores de carteira de crédito igual ou superior a R\$ 5 mil), parte-se para a determinação da “materialidade relativa”, ou seja, o termo parcela material do nosso estudo.

O procedimento adotado para a estimativa do percentual que equivale à parcela material é o seguinte:

- parte-se das 500 Maiores que possuem algum registro⁹⁶ no SCR, no período de 12/2002 a 12/2006;

⁹² Por exemplo, se a Empresa A possui dívidas no Banco X e no Banco Y, teremos dois elementos na amostra: Empresa A_Banco X e Empresa A_Banco Y.

⁹³ Simbologia utilizada para designar cada registro, onde CNPJ é a sigla de Cadastro Nacional da Pessoa Jurídica.

⁹⁴ Soma de todas as dívidas, considerando parcelas vencidas, vincendas e créditos baixados como prejuízo.

⁹⁵ A escolha desse critério baseia-se na Circular nº. 2.999 de 24/08/2000 (BCB, 2008a), que altera para R\$ 5 mil o valor mínimo da responsabilidade total do cliente em uma IF, que obriga a identificação deste no CRC (atual SCR).

⁹⁶ Um registro no SCR (ou elemento da amostra) representa a existência de dívida de uma empresa em uma IF.

- selecionam-se apenas os registros que não apresentavam carteira de crédito com atraso acima de 90 dias na data-base inicial de 12/2002⁹⁷;
- identifica-se, para cada registro, a primeira data-base com carteira de crédito em atraso acima de 90 dias, e somente são selecionados aqueles que possuam informações no SCR no terceiro mês subsequente⁹⁸;
- os percentuais de atraso relativos ao primeiro evento de atraso são classificados em decis, cujos intervalos determinam as classes de 1 a 10, que são utilizadas para classificar os registros nos dois momentos: “t” (primeiro evento de atraso) e “t+3” (terceiro mês subsequente ao primeiro atraso), conforme mostra a Ilustração 1:

Registro	Classe em "t"	Classe em "t+3"

Ilustração 1 - Esquema de Classificação de Registros

- partindo da classificação dos registros, conforme apresentado na Ilustração 1, constrói-se uma matriz que espelha a “migração” de classes dos registros do momento “t” para o momento “t+3”, onde cada célula apresenta a quantidade de registros que apresentaram o mesmo comportamento de mudança ou manutenção de classes. Na Ilustração 2, é apresentado o esquema desta matriz de migração (observe-se que as células hachuradas correspondem aos registros que apresentaram a mesma classificação nos dois momentos, à esquerda delas, os que mudaram para classes melhores, ou seja, com menores percentuais de atraso e, à esquerda delas, os que mudaram para classes piores, ou seja, com maiores percentuais de atraso):

⁹⁷ Não serão consideradas as empresas que já apresentavam atraso em 12/2002, pois o intuito é identificar o primeiro evento de atraso acima de 90 dias, partindo dessa data-base inicial.

⁹⁸ O critério utilizado tomou como base a Resolução nº. 2.682, no artigo 4º., inciso I, alínea “a”, que determina que as operações de crédito com atraso acima de 180 dias devem ser classificadas como “H”, ou seja, é necessária a constituição de provisão de 100% da operação. Assim, de maneira simplista, como a primeira ocorrência refere-se a atraso acima de 90 dias, três meses (ou 90 dias) depois, se não houvesse pagamentos, completariam 180 dias de atraso, ensejando 100% de provisão.

		Classe em "t+3"									
		1	2	3	4	5	6	7	8	9	10
Classe em "t"	1										
	2										
	3										
	4										
	5										
	6										
	7										
	8										
	9										
	10										

Ilustração 2 - Matriz de migração do percentual em atraso acima de 90 dias

- com base na matriz de migração, são selecionados apenas os casos em que houve piora na classificação por atraso (células à direita das células hachuradas, na Ilustração 2), que podem ser tabeladas em função da classe em "t", como apresentado na Ilustração 3:

Classe em "t"	% Piora em relação ao total da classe	Variação de piora em relação à classe anterior (p.p.)
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		

Ilustração 3 – Esquema de classes em função dos casos de piora de classificação de atraso

- a classe que indica a maior variação em p.p. (maior incremento de “piora”), em relação à classe anterior, é aquela que determina o percentual considerado como parcela material para efeito de caracterização do evento de *default*, uma vez que cada classe corresponde a um intervalo de percentuais de atraso acima de 90 dias (referentes aos decis anteriormente citados).

Para melhor entendimento dessa metodologia, uma configuração numérica simplificada encontra-se no Apêndice 1.

3.6.2 Seleção dos Eventos de *Default* e Não-*Default*

Definido o percentual que representa a parcela material da dívida, é possível selecionar os eventos de *default* e não-*default*, através dos seguintes passos:

- partindo-se da data-base inicial de 12/2002, tomam-se apenas os registros que não estavam em *default*, nessa data-base;
- para cada registro selecionado, procura-se pelo primeiro evento de *default*;
- se o registro não apresentar nenhum evento de *default*, ele será selecionado para compor o grupo de não-*default*, e a data-base de referência será escolhida aleatoriamente dentre aquelas em que houver dados no SCR⁹⁹.

Como o intuito é prever o evento de *default* no horizonte de um ano, para determinar o exercício da demonstração contábil a ser utilizada na construção do modelo, serão considerados:

- casos com *default* – a demonstração do ano imediatamente anterior à ocorrência do primeiro *default*;
- casos de não-*default* - a demonstração do ano imediatamente anterior à data-base selecionada como referência.

Pretendeu-se trabalhar com todos os registros disponíveis, mas deve-se atentar para possíveis problemas. Bessis (1998, p. 43) apresenta algumas taxas observadas de *default*, sendo a de tomadores *corporate* uma delas: “Taxas de *default* de tomadores *corporate* acima de 1% são freqüentemente consideradas especulativas”¹⁰⁰. O Relatório de Economia Bancária e Crédito de 2006 do BCB (BCB, 2008c, p. 12) apresenta como taxa de inadimplência¹⁰¹ o percentual de 2,7% para tomadores de crédito pessoas jurídicas. Note-se que aí se incluem micros e pequenas empresas, que, em princípio, possuem qualidade creditícia inferior às empresas *corporates*, de onde se deduz que a taxa de inadimplência das *corporate* é inferior a 2,7%. Esse baixo percentual de *default* pode causar problemas na construção do modelo, conforme exposto por Caouette *et al.* (1999, p. 169, grifos nossos):

⁹⁹ O mesmo procedimento foi adotado por Ohlson (1980).

¹⁰⁰ “*Default rates of corporate borrowers above 1% are often considered speculative.*”

¹⁰¹ Considerando os atrasos superiores a 90 dias.

No que se refere aos segmentos de qualidade superior da população, onde as inadimplências são mais raras, é difícil chegar a conclusões com base somente na capacidade de previsão de inadimplência. E é especialmente no que se refere às empresas de melhor qualidade que mais freqüentemente se pretende utilizar um modelo para classificar os riscos que elas representam para fins de precificação ou administração de carteiras. Os modelos podem exibir semelhantes poderes de previsão de inadimplência para empresas de maior qualidade, mas oferece classificações surpreendentemente diferentes de risco de inadimplência, simplesmente porque as taxas de inadimplência previstas são baixas e há poucas inadimplências subsequentes para validar uma abordagem em relação à outra. [...] Em vez de comparar taxas de erro de previsão entre modelos, uma avaliação mais direta poderia ser realizada por uma simulação de negociação baseada em regras e limitações decisórias consistentes. Os números de lucro e prejuízos resultantes poderiam ser, então, comparados. Infelizmente, este método também depende de dados de precificação de qualidade superior à atualmente disponível.

Servigny e Renault (2004, p. 111) apresentam quatro critérios cruciais para a avaliação da qualidade dos dados utilizados na modelagem (grifos nossos):

- Uso de uma grande quantidade de dados
- Disponibilidade de um número suficientemente grande de empresas em *default*
- Presença de um número limitado de dados faltantes entre os fatores de crédito selecionados
- Recurso de um procedimento transparente e “não invasivo” com relação à eliminação de outliers¹⁰²

Hosmer e Lemeshow (2000, p. 347) também fazem referência à importância da quantidade de eventos disponíveis (grifos nossos):

[...] possuir um tamanho adequado de amostra é tão importante no ajuste dos modelos de regressão logística quanto de qualquer modelo de regressão. Entretanto, o desempenho das estimativas baseadas no modelo pode ser determinado mais em função do número de eventos do que o tamanho total da amostra.¹⁰³

Deste modo, devido aos problemas apontados com relação ao baixo percentual de eventos de *default*, é provável que não se possa construir o modelo com a divisão da amostra em amostra de desenvolvimento e amostra de validação, sendo necessária a utilização de todos os registros existentes em uma única amostra, tal como feito por Ohlson (1980). Com relação à alternativa de compor a amostra com todos os eventos de *default* e igual número de eventos de não-*default*, tal como realizado por Brito (2005), também fica condicionada ao número de eventos de *default* existentes, em razão do grande número de variáveis explicativas a serem testadas.

¹⁰² “• *Use of large data set*

- *Availability of a sufficiently high number of defaulted companies*
- *Presence of a limited number of missing values among selected credit factors*
- *Recourse to a transparent and “noninvasive” procedure regarding the elimination of outliers”*

¹⁰³ “[...] having an adequate sample size is just as important when fitting logistic regression models as any other regression model. However, the performance of model-based estimates may be determined more by the number of events rather than the total sample size.”

3.6.3 Variáveis Contábeis e Cadastrais

Os indicadores contábeis a serem utilizados na construção do modelo foram selecionados com base nos principais índices utilizados na análise de balanços, segundo Matarazzo (2003, p. 153). Além desses, foram selecionados os indicadores mais comuns nos estudos sobre previsão de *default*¹⁰⁴. Deste modo, não se pretendeu encontrar novos indicadores, mas utilizar aqueles que já se mostraram relevantes em estudos anteriores, resultando na seguinte relação:

- $(PC + PELP) / PL$
- $PC / (PC + PELP)$
- AP / PL
- $AP / (PL + PELP)$
- $(AC - PC) / AT$
- $(AC + ARLP) / (PC + PELP)$
- $(AC - estoques) / PC$
- V / AT
- AC / V
- LL / V
- $LL / (PL - LL)$ ¹⁰⁵
- $EBITDA / (PC + PELP)$
- $EBITDA / AT$
- PL / AT
- PC / AT
- $LL / (PC + PELP)$

onde:

AC – ativo circulante;

¹⁰⁴ Alguns deles estão expostos no item 2.11.

¹⁰⁵ O indicador da rentabilidade do PL deveria ser medido em relação ao PL inicial do período (MATARAZZO, 2003, p. 182). Assim, além das demonstrações contábeis do período, seriam necessárias as demonstrações do período anterior, o que restringiria ainda mais a quantidade de observações disponíveis para a amostra, motivo pelo qual foi adotada essa *proxy*.

AP – ativo permanente;

ARLP – ativo realizável a longo prazo;

AT – ativo total;

PC – passivo circulante;

PELP – passivo exigível a longo prazo;

PL – patrimônio líquido;

V – receitas líquidas;

LL –lucro líquido;

EBTIDA – *earnings before interest, taxes, depreciation and amortization* (lucro antes de descontar os juros, os impostos sobre o lucro, a depreciação e a amortização).

A variável LN_AT (logaritmo natural do ativo total) é utilizada como *proxy* de tamanho da empresa, em função de estudos existentes, tais como: Ohlson (1980), Westgaard e Wijst (2001) e Bandyopadhyay (2007).

Algumas observações devem ser feitas sobre os dados contábeis das 500 Maiores:

- O ativo total foi ajustado para representar o total de recursos à disposição da empresa, sendo assim, as duplicatas descontadas não são reduzidas do ativo circulante e são reclassificadas no passivo circulante;
- Os valores foram ajustados, considerando a variação da inflação pelo IGP-M (Índice Geral de Preços – Mercado), para a mesma data-base, e expressos em dólares.

Além dos dados contábeis, foram fornecidos alguns dados cadastrais que também participam como possíveis variáveis explicativas do modelo, tais como: localização da sede (Estado), setor econômico, controle acionário¹⁰⁶ e negócios na bolsa (sim/não). Nesses casos, cada categoria é representada por uma variável *dummy*.

Com relação a variáveis não financeiras, Grunert *et al.* (2005, p. 509) encontraram evidências de que a utilização conjunta de variáveis financeiras e não financeiras gera uma previsão mais precisa de eventos futuros de *default*, comparando-se com o uso individual tanto das variáveis financeiras como das não financeiras. Note-se que se tratavam de variáveis comportamentais,

¹⁰⁶ “Indica o país de origem do acionista controlador. Empresas multinacionais controladas por holding constituída no Brasil são classificadas pelo país de origem do acionista controlador final.” (EXAME, 2007).

mais do que cadastrais.

Para complementar essas informações cadastrais, utiliza-se a data do registro do CNPJ das empresas para mensurar o tempo de funcionamento da empresa até a data da avaliação do seu risco de crédito. A variável explicativa equivale ao logaritmo natural¹⁰⁷ do tempo de abertura da empresa, LN_TA.

3.6.4 Variáveis do SCR

Os dados do SCR consistem na carteira de crédito por vencimento, ou seja, créditos a vencer, créditos vencidos e créditos baixados como prejuízo, para cada empresa, em cada IF. Esses dados constituem a base para a determinação do estado da variável dependente (*default* ou *não-default*). Além deles, a data de início de relacionamento¹⁰⁸ da empresa com a IF serve de base para criar uma variável explicativa que representa esse relacionamento. Com relação a esse assunto, Saunders (2000, p. 208, grifos do original) explica:

Uma relação antiga entre um tomador e um financiador gera um **contrato implícito** em termos de captação e pagamento que vai além do documento legal explícito no qual as relações entre tomador e financiador estão baseadas. A importância da reputação, que só pode ser construída com o passar do tempo, graças ao pagamento de compromissos e ao comportamento observado, é uma desvantagem para tomadores novos e de menor porte.

Ou seja, em princípio, quanto maior o tempo de relacionamento entre um cliente e uma IF, menor a chance de ocorrer um evento de *default*.

Como utilizamos os indicadores da demonstração contábil imediatamente anterior ao evento de *default* ou à data-base selecionada (no caso de *não-default*), existe a possibilidade de a primeira operação da empresa na IF (no caso, início do relacionamento) ocorrer no dia 31/12/X1, por exemplo, quando serão avaliadas as demonstrações contábeis de 31/12/X0. Deste modo, o tempo de relacionamento da empresa com a IF em relação à data da demonstração analisada resulta em um valor negativo de 360 dias (31/12/X0 - 31/12/X1). Assim, a informação a ser utilizada é “Tempo de Relacionamento+360”, e a variável

¹⁰⁷ A utilização do logaritmo para tempo baseou-se em estudos tais como Lawrence e Smith (1992) e Bandyopadhyay (2006).

¹⁰⁸ Segundo o leiaute do documento 3020 do SCR (BCB, 2007e), a data de início de relacionamento está assim definida: “Data de abertura de conta corrente ou outra data considerada relevante para avaliação do risco do crédito.”

explicativa, LN_TR, o logaritmo natural¹⁰⁹ dessa informação.

3.6.5 Variáveis Macroeconômicas

Com o intuito de incorporar o impacto das condições macroeconômicas no modelo, selecionaram-se, de maneira exploratória, três variáveis que poderiam influir no evento de *default*: taxa de juros, taxa de inflação e taxa de variação do PIB (Produto Interno Bruto). De maneira simples, apresenta-se como essas variáveis poderiam afetar a PD das empresas, além de indicar as fontes das informações utilizadas:

- o aumento da taxa de juros atua na diminuição dos recursos disponíveis para investimentos e para consumo, já que parte deles deve ser direcionada para a poupança, com tendência a afetar negativamente a situação econômica das empresas. A taxa de juros adotada neste estudo é a taxa de juros SELIC (Sistema Especial de Liquidação e de Custódia), divulgada pelo BCB, em % a.a.;
- o aumento da taxa de inflação aumenta os custos das empresas, sendo que muitas não conseguem repassar todo esse aumento para os preços dos seus produtos que, de qualquer modo, tendem ter a sua demanda diminuída, afetando negativamente a situação econômica da empresa. A taxa de inflação adotada é o IGP-M (Índice Geral de Preços – Mercado), divulgado pela FGV (Fundação Getúlio Vargas), em % a.a.;
- a variação positiva do PIB indica crescimento econômico, que tende a afetar positivamente a situação econômica das empresas. A taxa utilizada é a taxa de variação do PIB real, divulgada pelo IBGE (Instituto Brasileiro de Geografia e Estatística), em % a.a.

3.6.6 Seleção de Variáveis

Segundo Morrison (2006, p. 25): “Um dos métodos mais diretos de reduzir o número de preditivas em um problema de regressão talvez seja criar vários critérios de filtragem por regras através de correlações bivariadas e casadas.” As correlações casadas referem-se à correlação de cada variável preditiva com uma outra, enquanto que as correlações bivariadas referem-se à correlação da variável preditiva com a variável dependente. Nesse processo,

¹⁰⁹ A utilização de logaritmo para essa variável tem a mesma justificativa apresentada para a variável relativa ao tempo de abertura da empresa (LN_TA).

estabelece-se um limite de correlação casada e eliminam-se as variáveis que apresentem informações duplicadas, escolhendo apenas aquela com maior correlação com a variável dependente¹¹⁰. O autor explica que esse procedimento oferece uma visão simplista da seleção de variáveis, não incorporando testes de significância estatística e tratando apenas de um par de variáveis por vez, sendo necessário, em certo ponto, considerar uma abordagem mais multivariada.

Nesse ponto, levantamos o problema da multicolinearidade. Conforme Maroco (2003, p. 414):

Quando as variáveis estão fortemente correlacionadas entre si – condição designada por multicolinearidade - a análise do modelo de regressão ajustado pode ser extremamente confusa e desprovida de significado, fazendo desta condição – que as variáveis independentes o sejam de facto – um dos principais pressupostos a validar durante a análise de regressão.

O SPSS TRAINING (2003, p. 3-18) explica que, embora não exista um diagnóstico de multicolinearidade disponível para a regressão logística, é possível testar a sua existência. Como a multicolinearidade envolve relações entre as variáveis explicativas, sugere-se utilizar o procedimento da regressão linear com as mesmas variáveis da logística (dependente e independentes) e utilizar as estatísticas de colinearidade, enfatizando que todos os outros resultados devem ser ignorados.

Assim, utilizamo-nos desses dois procedimentos em conjunto, com o objetivo de evitar problemas com multicolinearidade das variáveis explicativas, selecionando aquelas que não apresentam informações “duplicadas”, ao mesmo tempo a redução do número de variáveis explicativas resulta em um modelo mais parcimonioso. No entanto, deve-se ter o cuidado de não eliminar variáveis significantes para o modelo.

Com isso, parte-se do modelo completo (com todas as variáveis explicativas importantes e uma constante) e faz-se a exclusão de variáveis que possuem menor significância no modelo, ou seja, um processo *backward* só de exclusão. Nesse processo, também são agrupados tipos de categorias quando possível, com o intuito de manter categorias que não são significantes individualmente, mas são significantes quando consideradas em conjunto. O modelo resultante é avaliado através do teste da razão de verossimilhança, e esse processo é repetido

¹¹⁰ No nosso caso, como a variável dependente é dicotômica, um critério que pode ser utilizado é a comparação do teste *t* de igualdade das médias das variáveis explicativas, em função da variável dependente.

sucessivamente até que todas as variáveis sejam significantes, a um nível de significância preestabelecido.

3.6.7 Limitações das Bases de Dados

As principais limitações das bases de dados utilizadas são:

- os dados do SCR são por IF, e não por conglomerado financeiro;
- os dados contábeis são por empresa, e não por conglomerado econômico;
- os dados das operações de crédito estão consolidados, ou seja, as características especiais de cada modalidade de crédito não estão sendo consideradas;
- expectativa de baixo percentual de eventos de *default* na amostra¹¹¹.

¹¹¹ Verrone (2007, p. 110) expôs em seu trabalho, realizado com executivos de 11 bancos no Brasil, que: “Foi apontada como uma questão importante a dificuldade de modelagem de risco para carteiras *corporate*, dada a pouca relevância estatística dos eventos de default, [...]”.

4 ANÁLISE DOS RESULTADOS

4.1 Estatísticas Descritivas da População e Amostras

4.1.1 Dados das 500 Maiores

Os dados fornecidos das 500 Maiores, no período de 2002 a 2005 (edições de 2003 a 2006), totalizam 2.000 registros, que correspondem a 638 empresas distintas. A participação dessas empresas nas quatro edições da Revista Exame está distribuída conforme apresentado na Tabela 1:

Tabela 1 – Participação das 500 Maiores nas edições de 2003 a 2006

Número de Edições	Frequência	Participação
1 ano	114	17,9%
2 anos	72	11,3%
3 anos	66	10,3%
4 anos	386	60,5%
Total	638	100,0%

FONTE: Base de dados das 500 Maiores

Salienta-se que as estatísticas descritivas apresentadas a seguir consideram cada empresa apenas uma vez, ou seja, embora a grande maioria delas (82,1%) tenha participado em mais de uma edição da revista, elas representarão apenas uma ocorrência nesse levantamento.

Com relação às informações cadastrais, para atender ao compromisso de confidencialidade dos dados utilizados, as categorias que representavam menos do que três empresas foram agrupadas com a denominação de “Demais”, para se evitar a possível identificação de empresas.

A maior parte das empresas participantes (55,9%) possuem sede nos estados de São Paulo e do Rio de Janeiro, conforme mostra a Tabela 2:

Tabela 2 – Sede das 500 Maiores nas edições de 2003 a 2006

Sede	Frequência	Participação
AL	4	0,6%
AM	18	2,9%
BA	21	3,3%
CE	11	1,7%
DF	13	2,1%
ES	19	3,0%
GO	6	1,0%
MG	50	7,9%
MT	3	0,5%
PA	8	1,3%
PE	7	1,1%
PR	39	6,2%
RJ	75	11,9%
RS	50	7,9%
SC	20	3,2%
SP	277	44,0%
Demais	9	1,4%
Total	630	100,0%

FONTE: Base de dados das 500 Maiores

Os setores de Energia, Bens de Consumo e Química e Petroquímica são os mais representativos, somando 29,9% das empresas, conforme podemos observar na Tabela 3:

Tabela 3 – Setor econômico das 500 Maiores nas edições de 2003 a 2006

Setor Econômico	Frequência	Participação
Atacado	38	6,0%
Auto-indústria	48	7,6%
Bens de Capital	10	1,6%
Bens de Consumo	64	10,2%
Comunicações	8	1,3%
Indústria da Construção	21	3,3%
Indústria Digital	21	3,3%
Diversos	4	0,6%
Eletroeletrônico	29	4,6%
Energia	66	10,5%
Farmacêutico	15	2,4%
Mineração	13	2,1%
Papel e Celulose	14	2,2%
Pendente - Nova Classif.	13	2,1%
Produção Agropecuária	20	3,2%
Química e Petroquímica	58	9,2%
Serviços	33	5,2%
Siderurgia e Metalurgia	49	7,8%
Telecomunicações	32	5,1%
Têxteis	10	1,6%
Transporte	19	3,0%
Varejo	45	7,1%
Total	630	100,0%

FONTE: Base de dados das 500 Maiores

Quanto ao controle acionário, as empresas brasileiras e americanas somam 63,9% do total das ocorrências, conforme mostra a Tabela 4:

Tabela 4 – Controle acionário das 500 Maiores nas edições de 2003 a 2006

Controle Acionário	Frequência	Participação
Alemão	22	3,5%
Americano/Brasileiro	4	0,6%
Americano	89	14,1%
Anglo-Holandês	5	0,8%
Belga	6	1,0%
Brasileiro	314	49,8%
Espanhol	13	2,1%
Estatal	41	6,5%
Francês	22	3,5%
Holandês	6	1,0%
Inglês	9	1,4%
Italiano	15	2,4%
Japonês	9	1,4%
Luxemburguês	8	1,3%
Mexicano	9	1,4%
Português/Espanhol	9	1,4%
Português	4	0,6%
Sueco	5	0,8%
Suíço	10	1,6%
Demais	30	4,8%
Total	630	100,0%

FONTE: Base de dados das 500 Maiores

A grande maioria das empresas participantes (82,1%) não possui ações negociadas em bolsa de valores, conforme pode ser visto na Tabela 5:

Tabela 5 – Situação de negociação em bolsa das 500 Maiores nas edições de 2003 a 2006

Negociação em Bolsa	Frequência	Participação
Não	506	80,3%
Pendente	11	1,7%
Sim	113	17,9%
Total	630	100,0%

FONTE: Base de dados das 500 Maiores

4.1.2 Dados do SCR

Do total de 638 empresas relativas às 500 Maiores, 8 empresas não possuem dados no SCR no período de 12/2002 a 12/2006 (dados mensais), ou seja, não possuem operações de crédito no SFN nesse período.

No total, existem 12.058 registros CNPJ_IF, sendo que cada registro representa os dados agrupados das operações de crédito de uma empresa em uma determinada IF.

Considerando esses registros, temos as seguintes estatísticas:

Tabela 6 – Dados das 500 Maiores no SCR

	Carteira de Crédito (R\$ mil)			Carteira de Crédito vencida há mais de 90 dias (R\$ mil)			Atraso acima de 90 dias no período
	dez/02	dez/06	Média do Período	dez/02	dez/06	Média do Período	
Registros	4.138	3.977	4.361	128	144	160	3,7%
Soma	84.786.312	108.155.181	94.027.836	839.488	633.574	1.064.802	1,1%
Média	20.490	27.195	21.578	6.558	4.400	6.636	
Máximo	2.203.643	2.035.137	2.505.660	357.929	160.194	161.972	

FONTE: Base de dados do SCR

Verifica-se que o percentual da carteira de crédito vencida há mais de 90 dias é de 1,1% (média no período em estudo). Esse percentual é inferior à taxa de inadimplência de 2,7%, apontada no Relatório de Economia Bancária e Crédito de 2006 do BCB (BCB, 2008, p. 12), como já observado anteriormente.

Com relação ao tempo de relacionamento, que é a informação que diferencia o comportamento de pagamento de uma mesma empresa perante diferentes IFs, os registros que não possuíam informação de data de início de relacionamento foram descartados. Para os restantes, utilizou-se, como critério para estabelecer a data de início de relacionamento, a data mais antiga informada no SCR¹¹², sendo que os registros que apresentam valores iguais ou inferiores a zero para “Tempo de Relacionamento+360” foram excluídos da análise.

¹¹² Como existe a possibilidade de informar a data de abertura de conta corrente na IF ou outra data considerada relevante para avaliação do risco do crédito, foi necessário definir um critério para processar essa informação, já que existem datas de início de relacionamento diversas para um mesmo cliente em uma mesma IF, dependendo da data-base considerada.

4.1.3 Determinação da Parcela Material da Dívida no Evento de *Default*

Partindo-se dos requisitos de “materialidade absoluta”, dos 12.058 registros iniciais, apenas 10.819 registros possuíam carteira de crédito maior ou igual a R\$ 5 mil no período. Desses, foram excluídos os casos com atrasos em 12/2002 (data-base inicial sem atraso), restando 10.694 registros.

Partindo desses registros, foram identificados aqueles que possuíam algum atraso até 09/2006 (para que se pudesse avaliar a situação após 3 meses dessa ocorrência). Assim, selecionamos 730 registros com algum evento de atraso, que foram classificados em decis em função do percentual de atraso, conforme tabela abaixo:

Tabela 7 – Percentual do Primeiro Evento de Atraso no Período de 01/2003 a 09/2006

Decil	Atraso
1	0,0003%
2	0,0068%
3	0,0352%
4	0,2007%
5	0,7980%
6	2,4637%
7	7,9979%
8	41,8000%
9_10	100,0000%

FONTE: Base de dados do SCR

Considerando cada decil como uma classe (de 1 a 9, sendo que a classe 9 representa os decis 9 e 10), podemos classificar o percentual de atraso de uma empresa em uma IF em “t” (mês do primeiro evento de atraso) e em “t+3” (terceiro mês subsequente ao primeiro evento de atraso), verificando o comportamento em relação ao percentual de atraso da dívida nesse período (melhora, manutenção ou piora).

Procedendo dessa maneira para cada registro CNPJ_IF, obteve-se a matriz de migração de classes de percentual de atraso, como mostra a Tabela 8:

Tabela 8 – Matriz de Migração de Classes de Percentual de Atraso

		Classe em "t+3"									Total
		1	2	3	4	5	6	7	8	9	
Classe em "t"	1	64	4		3		1				72
	2	51	14	5	1		1		2	1	75
	3	52	2	9	4	1	1	1	1	1	72
	4	46	2	2	12	6	1	1	1	2	73
	5	55	4	2	1	3	3	1	2	2	73
	6	47	3		2	2	7	3	6	3	73
	7	35		1	3	1		9	16	8	73
	8	35	1	2	1	1	2		11	20	73
	9	82	2		1			1	3	57	146
Total		467	32	21	28	14	16	16	42	94	730

FONTE: Base de dados do SCR

Para determinar o percentual que representaria uma “situação de dificuldade” do cliente (aumento do percentual de atraso), separamos a evolução do atraso em três casos: melhora, manutenção ou piora da situação em relação ao primeiro atraso, conforme mostra a Tabela 9:

Tabela 9 – Tabela Resumo da Migração das Classificações por Atraso

Classe em "t"	Total	Classe em "t+3"			Piora	
		Melhor	Igual	Pior	Casos	Variação
1	72	0	64	8	11%	
2	75	51	14	10	13%	2p.p.
3	72	54	9	9	13%	-1p.p.
4	73	50	12	11	15%	3p.p.
5	73	62	3	8	11%	-4p.p.
6	73	54	7	12	16%	5p.p.
7	73	40	9	24	33%	16p.p.
8	73	42	11	20	27%	-5p.p.
9	146	89	57	0	0%	-27p.p.
Total	730	442	186	102		

A classe 7 (que corresponde ao decil 7) foi aquela que apresentou a maior variação de piora do percentual de atraso. Como ela se refere aos percentuais de atraso entre 2,5% a 8,0%, adota-se, para a nossa definição do evento de *default*, o percentual de atraso igual ou superior a 2,5% da carteira de crédito como sendo a parcela material da dívida na caracterização do evento de *default*.

4.1.4 Dados da Amostra

Partindo-se da base de dados inicial de 10.819 registros que possuíam carteira de crédito maior ou igual a R\$ 5 mil no período em estudo, selecionam-se os registros que atendem aos seguintes requisitos:

- não possuem atraso superior a 90 dias em 12/2002;
- possuem pelo menos uma informação de carteira de crédito igual ou superior a R\$ 5 mil no SCR, no período de 01/2003 a 12/2006.

Os 10.609 registros selecionados são classificados em 10.221 casos de *não-default* e 388 casos de *default*¹¹³.

Dos casos de *não-default*, são excluídos aqueles que tiveram a primeira operação após 08/2006 (não possuem dados suficientes para apresentar atraso acima de 90 dias no período), restando 9.998 casos de *não-default* que, juntos com os 388 casos de *default*, totalizam 10.386 registros.

Assim, fazendo a associação dos 10.386 registros com a base de dados das 500 Maiores (em função do CNPJ da empresa e da data da demonstração contábil necessária¹¹⁴), além dos dados de “Tempo de Relacionamento+360”, o resultado obtido foi um conjunto de 7.918 registros, sendo 205 casos de *default* e 7.713 casos de *não-default*, que possuem todas as informações necessárias para a construção do modelo.

Deste modo, os registros distribuem-se nos exercícios das demonstrações contábeis utilizadas para a modelagem, conforme mostra a Tabela 10:

¹¹³ Considera-se caso de *default* o registro que possui algum evento de *default* e caso de *não-default* aquele registro que não possui nenhum evento de *default*.

¹¹⁴ Nos casos de *default*, a data do evento já está definida. Nos casos de *não default*, verificou-se em quais datas-base existiam informações nas 500 Maiores e no SCR e, dentre essas, foi escolhida, aleatoriamente, uma data-base de referência para o estudo desse caso.

Tabela 10 – Distribuição dos registros da amostra pelo exercício das demonstrações financeiras

Ano	Não-Default		Default		Total	
	Registros	Partic.	Registros	Partic.	Registros	Partic.
2002	2.176	96,8%	71	3,2%	2.247	100,0%
2003	2.004	97,8%	45	2,2%	2.049	100,0%
2004	1.743	96,9%	56	3,1%	1.799	100,0%
2005	1.790	98,2%	33	1,8%	1.823	100,0%
TOTAL	7.713	97,4%	205	2,6%	7.918	100,0%

Com relação à distribuição dos registros por estado da sede das empresas, verifica-se que a maior incidência de eventos de *default* ocorre no Amazonas, seguido por Mato Grosso, como pode visto na Tabela 11:

Tabela 11 – Distribuição dos registros da amostra em função da sede

Sede	Não-Default		Default		Total	
	Registros	Partic.	Registros	Partic.	Registros	Partic.
AL	66	100,0%	-	0,0%	66	100,0%
AM	115	92,7%	9	7,3%	124	100,0%
BA	316	98,8%	4	1,3%	320	100,0%
CE	158	98,8%	2	1,3%	160	100,0%
DF	91	96,8%	3	3,2%	94	100,0%
ES	264	98,9%	3	1,1%	267	100,0%
GO	128	100,0%	-	0,0%	128	100,0%
MG	739	97,1%	22	2,9%	761	100,0%
MT	60	95,2%	3	4,8%	63	100,0%
PA	107	100,0%	-	0,0%	107	100,0%
PE	84	100,0%	-	0,0%	84	100,0%
PR	522	97,8%	12	2,2%	534	100,0%
RJ	893	97,7%	21	2,3%	914	100,0%
RS	705	97,5%	18	2,5%	723	100,0%
SC	299	96,1%	12	3,9%	311	100,0%
SP	3.033	97,1%	91	2,9%	3.124	100,0%
Demais	133	96,4%	5	3,6%	138	100,0%
TOTAL	7.713	97,4%	205	2,6%	7.918	100,0%

Os registros classificados por setores econômicos das empresas apresentam maior percentual de eventos de *default* nos setores de diversos e de comunicações, conforme descrito na Tabela 12:

Tabela 12 – Distribuição dos registros da amostra em função do setor econômico

Setor Econômico	Não-Default		Default		Total	
	Registros	Partic.	Registros	Partic.	Registros	Partic.
Atacado	486	98,6%	7	1,4%	493	100,0%
Auto-indústria	312	98,7%	4	1,3%	316	100,0%
Bens de Capital	108	96,4%	4	3,6%	112	100,0%
Bens de Consumo	809	97,4%	22	2,6%	831	100,0%
Comunicações	66	90,4%	7	9,6%	73	100,0%
Indústria da Construção	343	95,0%	18	5,0%	361	100,0%
Indústria Digital	84	95,5%	4	4,5%	88	100,0%
Diversos	37	84,1%	7	15,9%	44	100,0%
Eletroeletrônico	235	95,9%	10	4,1%	245	100,0%
Energia	933	97,7%	22	2,3%	955	100,0%
Farmacêutico	95	100,0%	-	0,0%	95	100,0%
Mineração	138	97,2%	4	2,8%	142	100,0%
Papel e Celulose	285	97,6%	7	2,4%	292	100,0%
Pendente - Nova Classif.	50	96,2%	2	3,8%	52	100,0%
Produção Agropecuária	475	98,1%	9	1,9%	484	100,0%
Química e Petroquímica	800	97,9%	17	2,1%	817	100,0%
Serviços	284	97,9%	6	2,1%	290	100,0%
Siderurgia e Metalurgia	769	97,5%	20	2,5%	789	100,0%
Telecomunicações	473	98,3%	8	1,7%	481	100,0%
Têxteis	152	92,7%	12	7,3%	164	100,0%
Transporte	180	96,3%	7	3,7%	187	100,0%
Varejo	599	98,7%	8	1,3%	607	100,0%
TOTAL	7.713	97,4%	205	2,6%	7.918	100,0%

A empresas com controle acionário anglo-holandês, português-espanhol e francês apresentaram os maiores percentuais de registros com evento de *default*, conforme mostra a Tabela 13:

Tabela 13 – Distribuição dos registros da amostra em função do controle acionário

Controle Acionário	Não-Default		Default		Total	
	Registros	Partic.	Registros	Partic.	Registros	Partic.
Alemão	150	96,8%	5	3,2%	155	100,0%
Amer_Brasileiro	50	98,0%	1	2,0%	51	100,0%
Americano	425	95,9%	18	4,1%	443	100,0%
Anglo_Holandês	19	95,0%	1	5,0%	20	100,0%
Belga	58	100,0%	-	0,0%	58	100,0%
Brasileiro	5.169	97,4%	136	2,6%	5.305	100,0%
Espanhol	198	98,5%	3	1,5%	201	100,0%
Estatat	216	98,2%	4	1,8%	220	100,0%
Francês	239	95,2%	12	4,8%	251	100,0%
Holandês	88	98,9%	1	1,1%	89	100,0%
Inglês	56	100,0%	-	0,0%	56	100,0%
Italiano	218	97,8%	5	2,2%	223	100,0%
Japonês	36	100,0%	-	0,0%	36	100,0%
Luxemburguês	83	97,6%	2	2,4%	85	100,0%
Mexicano	57	100,0%	-	0,0%	57	100,0%
Port_Espanhol	76	95,0%	4	5,0%	80	100,0%
Português	75	98,7%	1	1,3%	76	100,0%
Sueco	28	96,6%	1	3,4%	29	100,0%
Suíço	131	97,8%	3	2,2%	134	100,0%
Demais	341	97,7%	8	2,3%	349	100,0%
TOTAL	7.713	97,4%	205	2,6%	7.918	100,0%

Praticamente não há diferenciação entre as empresas com ações negociadas ou não em bolsa, quando se verifica o percentual de registros com ocorrência de *default*, como mostra a Tabela 14:

Tabela 14 – Distribuição dos registros da amostra em função da negociação em bolsa

Negociação em Bolsa	Não-Default		Default		Total	
	Registros	Partic.	Registros	Partic.	Registros	Partic.
Não	5.453	97,48%	141	2,52%	5.594	100,0%
Sim	2.260	97,25%	64	2,75%	2.324	100,0%
TOTAL	7.713	97,41%	205	2,59%	7.918	100,0%

As estatísticas descritivas das variáveis explicativas métricas, encontram-se no Apêndice 3.

4.2 Construção do Modelo

Devido ao baixo percentual de eventos de *default* (2,6%) e à grande quantidade de variáveis

explicativas a serem analisadas (79), decidiu-se por construir o modelo com toda a amostra, não fazendo a divisão em amostra de desenvolvimento e amostra de validação. Também se descartou a alternativa de trabalhar com amostras emparelhadas, pois, devido ao pequeno número de casos (205 eventos de *default*), não seria possível construir o modelo partindo de um total de 79 variáveis explicativas¹¹⁵.

Inicialmente, com o intuito de reduzir o número de variáveis explicativas e eliminar aquelas que apresentassem multicolinearidade, utilizou-se o procedimento sugerido por Morrison (2006) para levantar as variáveis com alta correlação e determinar a mais significativa (procedimento de redução de variáveis já descrito), e o procedimento do SPSS para identificar a multicolinearidade entre as variáveis explicativas. Assim, foi descartada, do modelo inicial, a variável PCAT (passivo circulante sobre ativo total), restando 78 variáveis.

O método de seleção das variáveis foi o *backward*¹¹⁶, no qual inicialmente todas as variáveis explicativas são incluídas e a cada passo são eliminadas as menos significantes. Também foram agrupadas as variáveis cadastrais, quando apropriado. O processo terminou quando todas as variáveis tornaram significantes no modelo, a um nível de significância de 10%.¹¹⁷

O modelo completo constituído pelas 78 variáveis não apresentou solução final, sendo necessário eliminar 9 variáveis não significantes ($p\text{-value}=1$) para a convergência no processo de iteração. Deste modo, o primeiro modelo foi constituído por 69 variáveis explicativas.

Assim, foram testados 16 modelos sucessivamente, obtendo-se as estatísticas apresentadas na Tabela 15:

¹¹⁵ A lista com as variáveis explicativas testadas encontra-se no Apêndice 2.

¹¹⁶ O processo utilizado foi apenas de eliminação de variáveis, sem a inclusão.

¹¹⁷ Mayer (1995, p. 100) questiona a utilização do padrão de 5% para nível de significância.

Tabela 15 – Estatísticas dos modelos testados

Modelos	-2 Log likelihood	Parâmetros	Razão de Verossimilhança	Graus de Liberdade	<i>p-value</i>
Modelo 1	1.746,944	69			
Modelo 2	1.746,949	65	0,005	4	1,00
Modelo 3	1.747,109	61	0,161	4	1,00
Modelo 4	1.747,332	51	0,223	10	1,00
Modelo 5	1.747,692	48	0,360	3	0,95
Modelo 6	1.748,186	40	0,494	8	1,00
Modelo 7	1.749,662	35	1,477	5	0,92
Modelo 8	1.753,317	30	3,655	5	0,60
Modelo 9	1.758,326	25	5,009	5	0,41
Modelo 10	1.761,713	23	3,387	2	0,18
Modelo 11	1.763,543	22	1,830	1	0,18
Modelo 12	1.765,735	21	2,192	1	0,14
Modelo 13	1.767,517	20	1,781	1	0,18
Modelo 14	1.768,833	19	1,316	1	0,25
Modelo 15	1.770,747	18	1,914	1	0,17
Modelo 16	1.772,948	17	2,202	1	0,14

No processo de exclusão de variáveis explicativas, o *p-value* manteve-se sempre superior a 10%, não rejeitando a hipótese de que as variáveis excluídas não eram relevantes, a esse nível de significância.

A Tabela 16 descreve o modelo final (modelo 16):

Tabela 16 – Descrição do modelo final obtido

Variables in the Equation									
		B	S.E.	Wald	df	Sig.	Exp(B)	95,0% C.I. for EXP(B)	
								Lower	Upper
Step 1 ^a	LN_TR	,222	,071	9,883	1	,002	1,249	1,087	1,435
	APPL	,019	,010	3,870	1	,049	1,020	1,000	1,040
	ACPCAT	-1,221	,433	7,954	1	,005	,295	,126	,689
	PLAT	-,555	,238	5,423	1	,020	,574	,360	,916
	LN_AT	,136	,063	4,599	1	,032	1,145	1,012	1,297
	IGPM	,015	,008	3,357	1	,067	1,015	,999	1,032
	PIB	,090	,045	3,980	1	,046	1,094	1,002	1,195
	s_3_4	-,734	,425	2,987	1	,084	,480	,209	1,103
	s2	1,643	,370	19,702	1	,000	5,171	2,503	10,682
	C_1_3_4	,596	,230	6,713	1	,010	1,814	1,156	2,848
	C_10_19	,821	,301	7,443	1	,006	2,274	1,260	4,103
	C17	2,207	,730	9,146	1	,002	9,090	2,174	37,999
	Set_20_5	1,517	,263	33,187	1	,000	4,558	2,720	7,636
	Set6	1,087	,267	16,584	1	,000	2,965	1,757	5,003
	Set8	2,375	,433	30,077	1	,000	10,753	4,601	25,129
	Set19	-,924	,519	3,163	1	,075	,397	,143	1,099
	Constant	-6,440	,876	54,058	1	,000	,002		

a. Variable(s) entered on step 1: LN_TR, APPL, ACPCAT, PLAT, LN_AT, IGPM, PIB, s_3_4, s2, C_1_3_4, C_10_19, C17, Set_20_5, Set6, Set8, Set19.

Onde:

- LN_TR: logaritmo natural do tempo de relacionamento em anos¹¹⁸;
- APPL: ativo permanente sobre patrimônio líquido;
- ACPCAT: capital circulante líquido sobre ativo total;
- PLAT: patrimônio líquido sobre ativo total;
- LN_AT: logaritmo natural do ativo total;
- IGPM: índice geral de preços – mercado;
- PIB: variação do PIB real;
- s_3_4: sede na Bahia ou no Ceará;
- s2: sede no Amazonas;
- C_1_3_4: controle acionário alemão, americano-brasileiro ou anglo-holandês;
- C_10_19: controle acionário francês ou sueco;
- C17: controle acionário português-espanhol;
- Set_20_5: setor têxtil ou de comunicações;
- Set6: setor da indústria da construção;
- Set8: setor de diversos

¹¹⁸ O tempo de relacionamento é acrescido de 360 dias (um ano), conforme já explicado anteriormente.

- Set19: setor de telecomunicações.

Para a análise dos resíduos obtidos, verificamos o resíduo padronizado (resíduo dividido pela estimativa do seu desvio padrão), conforme o Gráfico 7:

Gráfico 7 - Gráfico do Resíduo Padronizado x Registro

Verifica-se que poucos casos apresentam resíduos padronizados superiores a “+3” e nenhum inferior a “-3”. Esses casos foram analisados e decidiu-se pela manutenção desses registros.

Com relação a medidas de influência, verifica-se a estatística de Cook (mede quanto o resíduo de todos os casos mudaria se um caso particular fosse excluído do cálculo dos coeficientes da regressão), através do Gráfico 8:

Gráfico 8 – Estatística de Cook x Registro

Se um caso não tivesse nenhuma influência nos resultados da regressão, a estatística de Cook teria valor zero. Os valores dessa estatística obtidos são baixos, indicando a não existência de casos influentes no resultado global.

As estatísticas resumo do modelo são apresentadas na Tabela 17:

Tabela 17 – Estatísticas resumo

Model Summary			
Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	1772,948 ^a	,016	,076

a. Estimation terminated at iteration number 7 because parameter estimates changed by less than ,001.

Como já comentado, os valores de R^2 (ou melhor, pseudos R^2) da regressão logística são baixos se comparados com a estatística equivalente da regressão linear (coeficiente de determinação). Salienta-se que esses valores não podem ser comparados com os resultados obtidos em estudos similares, tendo em vista que essa comparação só é possível quando se

utilizam a mesma amostra e as mesmas variáveis.

Na Tabela 18, apresentam-se as estatísticas obtidas com o teste de Hosmer e Lemeshow:

Tabela 18 – Estatísticas do teste de Hosmer e Lemeshow

Hosmer and Lemeshow Test			
Step	Chi-square	df	Sig.
1	3,469	8	,902

As estatísticas do teste não possibilitam a rejeição da hipótese nula de que não há diferenças significativas entre as classificações efetuadas pelo modelo e as classificações observadas (nas 10 classes). Ou seja, os resultados do modelo apresentam aderência adequada aos dados observados. Como não foram detectados casos influentes, isso representa um bom resultado, tendo em vista que o foco deste trabalho é a previsão.

Na Tabela 19, apresenta-se a tabela de contingência do teste de Hosmer e Lemeshow:

Tabela 19 – Tabela de contingência do teste de Hosmer e Lemeshow

Contingency Table for Hosmer and Lemeshow Test						
		DEFAULT = 0		DEFAULT = 1		Total
		Observed	Expected	Observed	Expected	
Step	1	786	787,105	6	4,895	792
1	2	783	784,328	9	7,672	792
	3	784	783,112	9	9,888	793
	4	777	780,064	15	11,936	792
	5	776	778,048	16	13,952	792
	6	776	775,734	16	16,266	792
	7	777	772,836	15	19,164	792
	8	770	768,702	22	23,298	792
	9	765	760,720	27	31,280	792
	10	719	722,351	70	66,649	789

Verifica-se, para cada uma das 10 classes, a quantidade de elementos preditos e observados. Note-se que todas as classes atendem ao requisito de possuir pelo menos 5 eventos esperados em cada classe.

Na Tabela 20, apresenta-se a tabela de classificação quando se considera valor de corte 0,5:

Tabela 20 – Tabela de classificação com valor de corte de 0,5

Classification Table ^a					
Observed			Predicted		Percentage Correct
			DEFAULT		
			0	1	
Step 1	DEFAULT	0	7713	0	100,0
		1	205	0	,0
Overall Percentage					97,4

a. The cut value is ,500

Nota-se o alto percentual global de eventos corretamente classificados (97,4%), mas, conforme já comentado, esse percentual é “enganador”, uma vez que nenhum caso de *default* foi corretamente classificado. Temos alta especificidade¹¹⁹ (100,0%) e baixa sensibilidade¹²⁰ (0,0%), o que não é razoável, pois não seria possível prever nenhum evento de *default*.

Segundo Crook *et al.* (2007, p. 1450), o valor de corte deve ser escolhido em função de sensibilidade e especificidade adequadas, uma vez que o aumento de uma implica na diminuição da outra. Com o intuito de encontrar um valor de corte adequado, apresentamos a curva ROC, através do Gráfico 9:

¹¹⁹ Probabilidade de verdadeiros negativos, ou seja, não-*defaults* corretamente classificados como não-*default*.

¹²⁰ Probabilidade de verdadeiros positivos, ou seja, *defaults* corretamente classificados como *default*.

Gráfico 9 – Curva ROC

Verifica-se que a área abaixo da curva ROC é de 0,697 e que, ao nível de significância de 5%, rejeita-se a hipótese nula de que seja igual a 0,5 (área abaixo da reta, que representa poder discriminante nulo). Ou seja, o modelo apresenta um razoável poder discriminante.

Para determinar o valor de corte mais adequado, adotou-se o critério de maximização da sensibilidade e da especificidade. Assim, para cada valor de corte, foi somada a probabilidade de *não-defaults* serem corretamente classificados como *não-defaults* com a probabilidade de *defaults* serem corretamente classificados como *defaults*. O valor máximo dessa soma ocorreu para o valor de corte de 0,028. Note-se que, deste modo, desconsidera-se que existam diferentes custos de classificações incorretas de *default* e de *não-default*.

Assim, o valor de corte adotado foi de 2,8% (os casos com PD calculada inferiores a 2,8% são classificados como *não-default* e aqueles com PD igual ou superior a 2,8% são classificados como *default*). Com isso, obtivemos percentuais de especificidade e sensibilidade de 73,1% e 57,1%, respectivamente, e percentual de acerto global de 72,7%, conforme mostra a Tabela 21:

Tabela 21 – Tabela de classificação com valor de corte de 0,028

Classification Table ^a					
Observed			Predicted		
			DEFAULT		Percentage Correct
			0	1	
Step 1	DEFAULT	0	5641	2072	73,1
		1	88	117	57,1
Overall Percentage					72,7

a. The cut value is ,028

Comparando com os resultados anteriores (valor de corte 0,5), percebe-se considerável diminuição do percentual de acerto global (de 97,4% para 72,7%). Obteve-se aumento da sensibilidade (de 0,0% para 57,1%) em troca de diminuição da especificidade (de 100,0% para 73,1%). Embora tenhamos, por simplicidade, desconsiderado os custos diferenciados das classificações incorretas de *default* e de não-*default* na determinação do valor de corte, sabe-se que o custo de um caso de *default* classificado incorretamente é maior do que a de um não-*default* classificado incorretamente. Isso por si só já justificaria o sacrifício da especificidade em prol da sensibilidade.

Com o intuito de comparar o procedimento utilizado com um procedimento alternativo, foi executada a modelagem pelo *stepwise*, método *forward LR*¹²¹, com nível de significância de 5% para a inclusão de variáveis e nível de significância de 10% para a eliminação. O modelo obtido¹²² apresentou 12 variáveis explicativas, mais a constante. Como o modelo obtido manualmente possui 16 variáveis explicativas, além da constante, é natural que possua um melhor ajuste. Para que seja possível comparar a qualidade do ajuste dos dois modelos, é necessário fazer a correção pelo número de variáveis. Assim, utiliza-se a estatística AIC (Akaike Information Criteria)¹²³ que resulta em 1.818,23 para o *stepwise* e 1.806,95 para o processo manual, o que indica que o procedimento manual apresenta melhor qualidade de ajuste do que o procedimento *stepwise*.

¹²¹ Utiliza a estatística de escore para a seleção das variáveis a serem incluídas, e a variação na razão da verossimilhança para a eliminação de variáveis.

¹²² A descrição do modelo encontra-se no Apêndice 4.

¹²³ Critério utilizado para comparar diferentes modelos que utilizam os mesmos dados. Quanto menor a estatística AIC, melhor a qualidade do ajuste do modelo.

4.3 Interpretação do Modelo

Na Tabela 22, são apresentadas as associações das variáveis explicativas métricas, com a variável dependente (PD), esperadas e obtidas:

Tabela 22 – Associações esperadas e obtidas das variáveis explicativas contínuas

Variável	Sinal do Coeficiente	
	Esperado	Obtido
LN_TR ¹²⁴	indeterminado	(+)
APPL ¹²⁵	(+)	(+)
ACPCAT ¹²⁶	(-)	(-)
PLAT ¹²⁷	(-)	(-)
LN_AT ¹²⁸	(-)	(+)
IGPM ¹²⁹	(+)	(+)
PIB ¹³⁰	(-)	(+)

FONTE: Elaboração da autora

A variável explicativa LN_TR, no caso desta amostra, não possuía um comportamento esperado em relação à PD. Com o resultado obtido, denota-se que empresas com tempo de relacionamento mais longo com a IF são mais propensas a inadimplir do que as empresas com tempo de relacionamento mais curto. Isso poderia indicar uma avaliação de risco mais rígida nas primeiras operações de uma empresa em uma IF, com exigência de melhores garantias, o que aumentaria a chances de pagamento. No entanto, é necessário salientar que, como já exposto, essa variável provém da informação de data de início de relacionamento da empresa, que depende da escolha de cada IF (data de abertura da conta corrente ou data relevante para a avaliação de risco de crédito).

As variáveis explicativas APPL, ACPCAT, PLAT e IGPM apresentaram comportamento condizente com o esperado, enquanto que as variáveis explicativas LN_AT e PIB apresentaram comportamento contrário ao senso comum. Nesse sentido, pode-se levantar

¹²⁴ LN_TR: logaritmo natural do tempo de relacionamento em anos.

¹²⁵ APPL = ativo permanente / patrimônio líquido.

¹²⁶ ACPCAT = capital circulante líquido / ativo total.

¹²⁷ PLAT = patrimônio líquido / ativo total.

¹²⁸ LN_AT = ln (ativo total).

¹²⁹ IGPM: índice geral de preços – mercado.

¹³⁰ PIB: variação do PIB real.

algumas hipóteses para o ocorrido:

- a variável LN_AT é uma *proxy* do tamanho da empresa e como a base de dados é composta apenas pelas maiores empresas brasileiras, essa informação passa a não ter a mesma importância que teria quando são consideradas empresas de portes diversos (caso de estudos anteriores);
- a variável PIB, embora seja contínua na sua concepção, apresenta apenas quatro valores neste estudo, já que, por simplicidade, apenas utilizamos o seu valor da data de fechamento das demonstrações financeiras. Além disso, ao colocar as variáveis macroeconômicas no mesmo nível hierárquico das variáveis microeconômicas (indicadores contábeis), assumiu-se uma grande restrição ao modelo, que pode ter distorcido a estimativa do coeficiente da variável em questão.

Neste ponto, cabe citar Gujarati (2000, p. 14):

[...] o pesquisador deve sempre estar ciente de que os resultados da pesquisa são apenas tão bons quanto a qualidade dos dados. Portanto, se em determinadas situações os pesquisadores acharem que os resultados da pesquisa são "insatisfatórios", a causa pode não ter sido o uso do modelo errado, mas a baixa qualidade dos dados. Infelizmente, por causa da natureza não-experimental dos dados usados na maior parte dos estudos da ciência social, os pesquisadores muitas vezes não têm outra escolha senão contar com os dados disponíveis. Mas eles devem estar sempre cientes de que os dados utilizados podem não ser os melhores. E devem tentar não ser excessivamente dogmáticos em relação aos resultados obtidos em determinado estudo, especialmente quando a qualidade dos dados for duvidosa.

Para viabilizar a realização deste estudo, dados foram consolidados e simplificações foram feitas. Quando isso ocorre, há perda de qualidade, o que prejudica a modelagem e, conseqüentemente, os resultados obtidos. No entanto, o modelo construído apresenta boa aderência aos dados observados, que é importante quando o foco é a previsão, e razoável poder discriminante na identificação de eventos de *default*.

Analisando as variáveis explicativas categóricas que compõem o modelo, podemos dizer que as características que têm influência no aumento da PD são: sede no Amazonas; controles acionários alemão, americano-brasileiro, anglo-holandês, francês, sueco ou português-espanhol; e setores têxtil, de comunicações, da indústria da construção, ou de diversos. Por outro lado, as características que têm influência na diminuição da PD são: sede na Bahia ou no Ceará; e setor de telecomunicações.

Com o intuito de facilitar o entendimento, considere-se o caso de uma empresa que possua como variáveis explicativas métricas os valores das medianas da amostra, e categorias que impliquem nas menores PD, como, por exemplo, sede no Ceará, controle acionário brasileiro e setor de telecomunicações. Essa empresa, pelo modelo, apresenta PD de 0,3%. Se apenas mudar a sede para o Amazonas, a PD sobe para 3,5%. E se, no caso anterior, alterar também o controle acionário para português-espanhol, a PD sobe para 24,8%. Para completar, se mudar também o setor para diversos, a PD passaria para 89,9%¹³¹. Ou seja, mantendo-se os mesmos valores das variáveis explicativas métricas (medianas) e considerando as categorias que mais aumentem a PD, como sede no Amazonas, controle acionário português-espanhol e setor de diversos, a PD passa de 0,3% para 89,9%. São dois casos “extremos” e fictícios, mas que mostram como o modelo pode ser utilizado e interpretado.

4.4 Limitações do Estudo

Para a concretização deste trabalho, houve necessidade de simplificação em diversos aspectos. As principais limitações na construção do modelo foram:

- o modelo utilizado pressupõe que as observações sejam independentes entre si, mas, no nosso caso, a mesma empresa possui operações de crédito em várias IFs, ou seja, representa diversas observações, enfraquecendo essa independência;
- o modelo refere-se ao SFN como um todo, o que não considera as características individuais de cada IF;
- desconsiderou-se a ocorrência de fusão/aquisição de empresas e/ou IFs;
- a definição de início de relacionamento de uma empresa em uma IF abrange conceitos diferentes (abertura de conta corrente na IF ou data relevante para a avaliação do risco de crédito), prejudicando a utilização dessa informação para fins de um modelo global do risco do cliente, uma vez que diferentes IFs (ou até a mesma IF) podem informar tipos de dados diferentes para o mesmo cliente, dependendo da modalidade ou época da operação;
- a variável relativa ao tempo de abertura da empresa pode não representar tempo de funcionamento da mesma quando há mudança de CNPJ;

¹³¹ Salienta-se que não existe empresa com essas características na amostra de modelagem.

- apesar de a avaliação do risco deste estudo focar no cliente e não na operação, a consolidação de informações de modalidades diversas pode distorcer os resultados, uma vez que características específicas estão sendo desconsideradas¹³²;
- a consolidação de categorias de variáveis explicativas, para evitar a identificação das empresas, acarreta perda no poder preditivo do modelo;
- não foi possível realizar a validação externa do modelo;
- as variáveis macroeconômicas foram incluídas no modelo de maneira simplista, no mesmo nível hierárquico das variáveis microeconômicas;
- não foi considerada a diferença entre os custos de classificação incorreta de *default* e de *não-default*;
- a maior parte dos registros referem-se a empresas que não têm a obrigatoriedade de que suas demonstrações contábeis sejam avaliadas por auditores independentes. Na ausência de auditoria independente, a credibilidade dos dados informados fica prejudicada.

¹³² Segundo Sicsú (2003, p. 329): “Em geral, quanto maior sua abrangência, menor será o seu poder discriminador. Modelos que atendem a vários produtos ou mercados podem perder eficiência, pois características específicas de uma operação ou um mercado não podem ser consideradas na fórmula de cálculo de escore.”

5 CONCLUSÕES

O risco de crédito é uma das principais preocupações quando se pensa na saúde financeira de uma IF. Apesar de ser um risco inerente ao negócio bancário e, portanto, objeto de inúmeros estudos desde o surgimento dessa atividade, a sua mensuração permanece um constante problema, já que a conjuntura econômica é dinâmica e as inovações financeiras aumentam a complexidade da sua avaliação.

O Basileia II representa uma grande mudança na mensuração do risco de crédito. De uma estrutura rígida de atribuição de níveis de risco para a ponderação no cálculo do capital mínimo requerido das IFs, cria-se a possibilidade de que modelos internos sejam utilizados para a mensuração do risco de crédito. Nesse contexto, a PD passa a ser um dos principais elementos na determinação do requerimento mínimo de capital de uma IF.

Além dessa aplicação, a PD tem papel importante na constituição de provisão para créditos de liquidação duvidosa, na precificação e no estabelecimento de limites para as operações de crédito de um cliente, já que compõe o cálculo da perda esperada.

Assim, o objetivo principal deste trabalho foi construir um modelo com variáveis contábeis e cadastrais de grandes empresas, juntamente com variáveis macroeconômicas, para estimar a PD dessas empresas no SFN. Um modelo com essas características tem utilidade para as IFs que necessitam desenvolver modelos internos para auxiliarem na gestão de riscos e que, também, atendam ao Basileia II, bem como para o órgão supervisor do SFN, que deverá avaliar e autorizar os modelos internos das IFs.

Apesar da existência de uma crescente variedade de técnicas utilizadas para esse tipo de modelagem, a técnica estatística selecionada foi a da regressão logística. Dentre os vários fatores que influenciaram na escolha, podemos destacar: facilidade em trabalhar com variáveis categóricas, maior flexibilidade quanto às suposições iniciais, escala de resultados entre zero e um (indispensável quando se trata de cálculo de probabilidades) e a relativa facilidade da interpretação dos resultados dada a sua semelhança com a regressão linear múltipla.

As bases de dados utilizadas foram a das 500 Maiores e a do SCR que, devido ao compromisso de confidencialidade, tiveram alguns dos seus dados consolidados. Apesar da grande quantidade de registros disponíveis (7.918), o baixo percentual de eventos de inadimplência (2,6%) e o grande número de variáveis explicativas testadas (79) dificultaram a modelagem e impossibilitaram a avaliação do ajuste através da validação externa, que é recomendada nos casos em que se pretende prever os resultados de futuros elementos.

Antes da modelagem em si, foi necessário determinar a parcela material que define o evento de *default* (atraso superior a 90 dias de parcela material da dívida de uma empresa, em relação a uma IF). Na falta de uma definição objetiva desse termo, através de uma matriz de migração de classificações por percentuais de atraso, estimou-se o percentual da carteira de crédito em atraso acima de 90 dias, a partir do qual há chances consideráveis de piora (aumento do percentual de atraso) no decorrer do tempo. Dessa maneira, a parcela material foi definida como 2,5% da dívida.

Embora existam vários trabalhos que tratam da probabilidade de insolvência, utilizando informações de falência ou concordata, são raros os trabalhos sobre probabilidade de inadimplência, especialmente quando se trata do SFN como um todo e não somente de uma IF. Assim, as suas associações com as variáveis explicativas não são bem conhecidas e a utilização do procedimento *stepwise* de pacotes estatísticos existentes seria uma opção prática e fácil, em função da rapidez na análise de um grande número de variáveis e ao ajuste simultâneo de várias equações. No entanto, devido a restrições desse procedimento no tratamento das variáveis categóricas (não é possível o agrupamento entre elas, eliminando aquelas que poderiam ser significantes no conjunto, mas não o são individualmente), decidiu-se por um processo manual de seleção de variáveis. A partir de um modelo inicial com todas as variáveis explicativas, eliminaram-se as variáveis menos significantes e agruparam-se as variáveis com características semelhantes, em várias etapas, até que todas as variáveis contidas no modelo fossem significantes. Analisaram-se 16 modelos, consecutivamente. Salienta-se que não existe o modelo ótimo, mas modelos que podem atender às necessidades de cada usuário específico.

O modelo final obtido, com valor de corte de 2,8%, apresentou 72,7% de classificações globais corretas, sendo 73,1% de classificações corretas de eventos de não-*default* e 57,1%, de eventos de *default*. Considerando que um modelo que classifique todos os eventos como não-

default (sem nenhuma variável explicativa, mas apenas uma constante igual a zero) apresentaria percentual de acertos globais de 97,4%, poderíamos chegar à conclusão enganosa de que esse último modelo é melhor do que o elaborado neste trabalho. No entanto, o modelo concorrente não teria qualquer utilidade em um processo de gestão de riscos, cujo interesse é identificar futuros eventos de *default*.

Através de vários testes estatísticos, verificou-se que o modelo construído apresenta boa qualidade de ajuste aos dados observados, que é importante quando o foco do modelo é a previsão, e razoável poder discriminante, o que representa um bom resultado, tendo em vista o baixo percentual de eventos de *default* existente (2,6%). Além do aspecto do baixo percentual, várias limitações prejudicaram o desempenho do modelo, tais como: consolidação de todas as operações de crédito de uma empresa em uma IF, sem separação por modalidade; consideração de independência dos registros da amostra, trabalhando com todo o SFN e não com IF; e utilização de variáveis macroeconômicas de maneira “pontual” (apenas no fechamento do exercício contábil) e no mesmo nível hierárquico das variáveis microeconômicas.

Com relação às variáveis explicativas métricas, o modelo indicou que as variáveis ACPCAT¹³³ e PLAT¹³⁴ possuem associação negativa com a PD, comportamento já esperado para essas variáveis. Por outro lado, aponta-se que as variáveis LN_TR¹³⁵, APPL¹³⁶, LN_AT¹³⁷, IGPM¹³⁸ e PIB¹³⁹ possuem associação positiva com a PD. Esse comportamento era esperado para as variáveis APPL e IGPM, mas não para LN_AT e PIB. Os resultados não esperados podem ter explicação: na composição da amostra pelas maiores empresas do país, fazendo com que o porte não possua o poder discriminante esperado nesse grupo - no caso da variável LN_AT; e pela simplicidade no trato das variáveis macroeconômicas, utilizando apenas dados pontuais nas datas das demonstrações contábeis e considerando-as no mesmo nível hierárquico das variáveis microeconômicas - no caso da variável PIB. Com relação à variável LN_TR, há indicação de que empresas com tempo de relacionamento mais longo

¹³³ ACPCAT = capital circulante líquido / ativo total.

¹³⁴ PLAT = patrimônio líquido / ativo total.

¹³⁵ LN_TR = ln (tempo de relacionamento da empresa na IF).

¹³⁶ APPL = ativo permanente / patrimônio líquido.

¹³⁷ LN_AT = ln (ativo total).

¹³⁸ IGPM: índice geral de preços – mercado.

¹³⁹ PIB: variação do PIB real.

com a IF são mais propensas a entrar em *default* do que as empresas com tempo de relacionamento mais curto. No entanto, salienta-se que a definição da data de início de relacionamento da empresa depende da escolha de cada IF, ou seja, essa informação pode não ser uniforme no SFN.

Quanto às variáveis explicativas categóricas, o modelo indicou que as características que têm influência no aumento da PD são: sede no Amazonas; controles acionários alemão, americano-brasileiro, anglo-holandês, francês, sueco ou português-espanhol; e setores têxtil, de comunicações, da indústria da construção, ou de diversos. Por outro lado, as características que têm influência na diminuição da PD são: sede na Bahia ou no Ceará; e setor de telecomunicações.

Como exemplo, compara-se uma empresa “A” (sede no Ceará, controle acionário brasileiro e do setor de telecomunicações) com uma empresa “B” (sede no Amazonas, controle acionário português-espanhol e do setor de diversos), assumindo, para ambos os casos, que todas as variáveis explicativas métricas possuem os valores das medianas da amostra. Através do modelo construído, obtém-se PD de 0,3% para a empresa “A” e de 89,9%, para a empresa “B”. Esses casos são fictícios, mas proporcionam uma visão de como as características das empresas (no caso, de dois exemplos extremos) têm influência na estimativa da PD.

Para fins de comparação, realizou-se a modelagem com a mesma amostra utilizando o procedimento *stepwise forward LR*, que resultou em um modelo com qualidade de ajuste inferior ao modelo obtido através do processo manual de seleção de variáveis.

As principais contribuições deste trabalho consistem na apresentação de uma metodologia para o cálculo do percentual que representa a parcela material que caracteriza o evento de *default*; na exploração da modelagem com regressão logística através de um procedimento manual de seleção de variáveis, como uma alternativa melhor do que o procedimento automatizado do *stepwise*; e na utilização de variáveis oriundas do SCR, com a visão do conjunto do SFN e não apenas de uma única IF.

Este trabalho serve potencialmente como referência para outros estudos de previsão de inadimplência bancária, mostrando as dificuldades enfrentadas e apresentando sugestões para aprimoramento. As principais sugestões são: trabalhar com dados das operações de crédito

das empresas por modalidades, obter uma amostra com maior número de eventos de *default* para que seja possível efetuar a validação externa, utilizar um modelo que não prescindia da independência dos elementos da amostra (GEE - *Generalized Estimating Equations*) e trabalhar com modelos multiníveis para incorporar, de forma mais adequada, as variáveis macroeconômicas.

REFERÊNCIAS

ALTMAN, Edward I. *Corporate financial distress and bankruptcy: a complete guide to predicting & avoiding distress and profiting from bankruptcy*. 2nd ed. New York: Wiley, 1993.

ALTMAN, Edward I.; NARAYANAN, Paul. *An international survey of business failure classification models*. **Financial Markets, Institutions & Instruments**, v. 6, n. 2, may 1997.

ASSAF NETO, Alexandre. **Mercado financeiro**. 6. ed. São Paulo: Atlas, 2005.

AZIZ, M. Adnan; DAR, Humayon A. *Predicting corporate bankruptcy: whither do we stand?* **Loughborough University Working Paper**, 2004. Disponível em <<http://hdl.handle.net/2134/325>>. Acesso em: 07 jul. 2007.

BANCO CENTRAL DO BRASIL – BCB. **O Banco Central do Brasil – BACEN**. Disponível em: <<http://www.bcb.gov.br/pre/composicao/bacen.asp>>. Acesso em 28 jul. 2007a.

BANCO CENTRAL DO BRASIL. **Circular n. 2.999**, de 24 de agosto de 2000. Disponível em: <<http://www5.bcb.gov.br/normativos/detalhamentocorreio.asp?N=100159246&C=2999&ASS=CIRCULAR+2.999>>. Acesso em: 13 jan. 2008a.

BANCO CENTRAL DO BRASIL - BCB. **Circular n. 3.098**, de 20 de março de 2002. Disponível em: <<http://www5.bcb.gov.br/normativos/detalhamentocorreio.asp?N=102046610&C=3098&ASS=CIRCULAR+3.098>>. Acesso em: 01 abr. 2007b.

BANCO CENTRAL DO BRASIL - BCB. **Composição e evolução do SFN**. Disponível em <<http://www.bcb.gov.br/?SFNCOMP>>. Acesso em 28 jul. 2007c.

BANCO CENTRAL DO BRASIL - BCB. **O Conselho Monetário Nacional – CMN**. Disponível em <www.bcb.gov.br/pre/composicao/CMN.asp>. Acesso em 28 jul. 2007d.

BANCO CENTRAL DO BRASIL - BCB. **Leiaute do Documento 3020**. Disponível em: <http://www.bcb.gov.br/fis/crc/ftp/SCR_Leiaute_fev07-v12.xls>. Acesso em 05 ago. 2007e.

BANCO CENTRAL DO BRASIL - BCB. **Planejamento Estratégico do BCB**. Disponível em: <<http://www.bcb.gov.br/?PLANIBC>>. Acesso em 28 jul. 2007f.

BANCO CENTRAL DO BRASIL - BCB. **PROER**. Disponível em: <<http://www.bcb.gov.br/?PROER>>. Acesso em: 28 jul. 2007g.

BANCO CENTRAL DO BRASIL - BCB. **Regimento interno do Banco Central do Brasil**.

Disponível em: <http://www.bcb.gov.br/Pre/sobre/regimento_interno_bcb.pdf>. Acesso em: 01 abr. 2007h.

BANCO CENTRAL DO BRASIL - BCB. **Relatório de economia bancária e crédito - 2005**. Disponível em: <http://www.bcb.gov.br/pec/spread/port/rel_econ_ban_cred.pdf>. Acesso em: 26 jan. 2008b.

BANCO CENTRAL DO BRASIL - BCB. **Relatório de economia bancária e crédito - 2006**. Disponível em: <http://www.bcb.gov.br/Pec/spread/port/relatorio_economia_bancaria_credito.pdf>. Acesso em: 17 jan. 2008c.

BANCO CENTRAL DO BRASIL - BCB. **Resolução n. 1.748**, de 30 de agosto de 1990. Disponível em: <<http://www5.bcb.gov.br/normativos/detalhamentocorreio.asp?N=090131116&C=1748&ASS=RESOLUCAO+1.748>>. Acesso em: 01 abr. 2007i.

BANCO CENTRAL DO BRASIL - BCB. **Resolução n. 2.099**, de 24 de fevereiro de 2000. Disponível em: <<http://www5.bcb.gov.br/normativos/detalhamentocorreio.asp?N=094163143&C=2099&ASS=RESOLUCAO+2.099>>. Acesso em: 16 maio 2007j.

BANCO CENTRAL DO BRASIL - BCB. **Resolução n. 2.682**, de 21 de dezembro de 1999. Disponível em: <<http://www5.bcb.gov.br/normativos/detalhamentocorreio.asp?N=099294427&C=2682&ASS=RESOLUCAO+2.682>>. Acesso em: 01 abr. 2007.

BANCO CENTRAL DO BRASIL - BCB. **Resolução n. 2.692**, de 24 de fevereiro de 2000. Disponível em: <<http://www5.bcb.gov.br/normativos/detalhamentocorreio.asp?N=100037112&C=2692&ASS=RESOLUCAO+2.692>>. Acesso em: 16 maio 2007k.

BANCO CENTRAL DO BRASIL - BCB. **Séries Temporais**. Disponível em <<https://www3.bcb.gov.br/sgspub/localizarseries/localizarSeries.do?method=prepararTelaLocalizarSeries>>. Acesso em 2007l.

BANCO CENTRAL DO BRASIL - BCB. **Sistema de Informações de Crédito do Banco Central – SCR**. 2004. Disponível em: <http://www.bcb.gov.br/fis/crc/ftp/SCR_VisaoGeral_v1.00.pdf>. Acesso em 05 ago. 2007m.

BANDYOPADHYAY, Arindam. *Mapping corporate drift towards default – part 2: a hybrid credit-scoring model*. **Journal of Risk Finance**, [S.l.], v. 8, n. 1, p. 46-55, 2007.

BANDYOPADHYAY, Arindam. *Predicting probability of default of Indian corporate bonds: logistic and Z-score model approaches*. **Journal of Risk Finance**, [S.l.], v. 7, n. 3, p. 255-272, 2006

BASEL COMMITTEE ON BANKING SUPERVISION - BCBS. *An explanatory note on the Basel II IRB risk weight functions*. Basel, 2005. Disponível em: <<http://www.bis.org/bcbs/irbriskweight.pdf>>. Acesso em: 05 ago. 2007.

BASEL COMMITTEE ON BANKING SUPERVISION - BCBS. *International convergence of capital measurement and capital standards: a revised framework comprehensive version*. Basel, 2006. Disponível em: <<http://www.bis.org/publ/bcbs128.pdf>>. Acesso em: 26 jan. 2007.

BESSIS, Joel. *Risk management in banking*. Chichester: John Wiley & Sons, 1998.

BRASIL. Presidência da República. **Lei n. 4.595**, de 31 de dezembro de 1964. Disponível em: <http://www.planalto.gov.br/ccivil_03/LEIS/L4595.htm>. Acesso em: 27 set. 2007.

BRASIL. Presidência da República. **Lei n. 6.404**, de 15 de dezembro de 1976. Disponível em: <http://www.planalto.gov.br/ccivil_03/Leis/L6404compilada.htm>. Acesso em 30 ago. 2007.

BRITO, Giovani A. S. **Mensuração de risco de portfolio para carteira de crédito a empresas**. 2005. 83 f. Dissertação (Mestrado em Controladoria e Contabilidade) – Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo, São Paulo.

CANUTO, Otaviano; LIMA, Gilberto Tadeu. **Basiléia 2: da regulação substantiva à regulação procedimental**. In: FENDT, Roberto; LINS, Maria Antonieta Del Tedesco (Org.). **Arquitetura assimétrica: o espaço dos países emergentes e o sistema financeiro internacional**. São Paulo: Fundação Konrad Adenauer, 2002.

CAOUTTE, John B. *et al.* **Gestão do risco de crédito: o próximo grande desafio financeiro**. Rio de Janeiro: Qualitymark, 1999.

CAPELLETTO, Lúcio Rodrigues. **Mensuração do risco sistêmico no setor bancário com utilização de variáveis contábeis e econômicas**. 2006. 259 f. Tese (Doutorado em Ciências Contábeis) – Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo, São Paulo.

CARLING, Kenneth *et al.* *Corporate credit risk modeling and the macroeconomy*. **Journal of Banking and Finance**, [S.l.], v. 31, p. 845-868, 2007.

CARNEIRO, Fábio Lacerda *et al.* **Novo acordo de Basiléia: estudo de caso para o contexto brasileiro**. **Resenha BM&F**, São Paulo, 163, 2005.

CASEY, Cornelius; BARTCZAK, Norman. *Using operating cash flow data to predict financial distress: some extensions*. **Journal of Accounting Research**, [S.l.], v. 23, n. 1, Spring 1985.

COSSIN, Didier; PIROTTE, Hugues. *Advanced credit risk analysis: financial approaches and mathematical models to assess, price, and manage credit risk*. Baffins Lane: John Wiley & Sons, 2000.

CROOK, Jonathan N. *et al. Recent developments in consumer credit risk assessment*. **European Journal of Operational Research**, [S.l.], v. 28, n. 3, p. 1447-1465, 2007.

DERMINE, Jean; CARVALHO, Cristina Neto de. *Bank loan losses-given-default: a case study*. **Journal of Banking and Finance**, [S.l.], v. 30, p. 1219-1243, 2006.

DIAS FILHO, José Maria. **Gestão tributária na era da responsabilidade fiscal**: propostas para otimizar a curva da receita utilizando conceitos de semiótica e regressão logística. 2003. 251 f. Tese (Doutorado em Controladoria e Contabilidade) – Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo, São Paulo.

EXAME. Melhores e Maiores: as 500 maiores empresas do país. São Paulo: Editora Abril, 2007.

FERREIRA, Aurélio Buarque de Holanda. **Novo dicionário Aurélio da língua portuguesa**. 3. ed. Curitiba: Positivo, 2004.

FITCH, Thomas P. *Dictionary of banking terms*. 4th ed. New York: Barron's, 2000.

GITMAN, Lawrence J. **Princípios de administração financeira**. 10. ed. São Paulo: Pearson Addison Wesley, 2004.

GONÇALVES, Eric Bacconi. **Análise de risco de crédito com o uso de modelos de regressão logística, redes neurais e algoritmos genéticos**. 2005. 96 f. Dissertação (Mestrado em Administração) – Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo, São Paulo.

GRUNERT, Jens *et al. The role of non-financial factors in internal credit ratings*. **Journal of Banking and Finance**. [S.l.], v. 29, p. 509-531, 2005.

GUJARATI, Damodar N. **Econometria básica**. 3. ed. São Paulo: Pearson Education, 2000.

HAIR JR., Joseph F. *et al. Multivariate Data Analysis*. 6th ed. New Jersey: Pearson Prentice Hall, 2005.

HILL, R. Carter *et al. Econometria*. São Paulo: Saraiva, 2000.

HORNBY, A. S. *Oxford Advanced Learner's Dictionary*. 6th ed. Oxford: Oxford University Press, 2000.

HOSMER, David W.; LEMESHOW, Stanley. *Applied logistic regression*. 2nd ed. New York: John Wiley & Sons, 2000.

IUDÍCIBUS, Sérgio *et al.* **Manual de contabilidade das sociedades por ações**. 6. ed. São Paulo: Atlas, 2003.

JOHNSTON, Jack; DINARDO, John. *Econometric methods*. 4th ed. Singapore: McGraw-Hill, 1997.

JORION, Philippe. *Value at risk*: a nova fonte de referência para a gestão do risco financeiro. 2. ed. São Paulo: Bolsa de Mercadorias e Futuros, 2003.

LAWRENCE, Edward C.; SMITH, L. Douglas. *An analysis of default risk in mobile home credit*. **Journal of Banking and Finance**, [S.l.], v. 16, p. 299-312, 1992.

LIMA, Iran Siqueira *et al.* (Org.). **Curso de mercado financeiro: tópicos especiais**. São Paulo: Atlas, 2006.

MARCONI, Marina de Andrade; LAKATOS, Eva Maria. **Fundamentos de metodologia científica**. 6. ed. São Paulo: Atlas, 2005.

MAROCO, João. **Análise estatística com utilização do SPSS**. 2. ed. Lisboa: Edições Silabo, 2003.

MARTINS, Eliseu (Org.). **Avaliação de empresas**: da mensuração contábil à econômica. São Paulo: Atlas, 2001.

MARTINS, Gilberto de Andrade. **Estatística geral e aplicada**. 3. ed. São Paulo: Atlas, 2005.

MARTINS, Gilberto de Andrade. **Manual para elaboração de monografias e dissertações**. 3. ed. São Paulo: Atlas, 2002.

MARTINS, Márcio Severo. **A previsão de insolvência pelo modelo de Cox**: uma contribuição para a análise de companhias abertas brasileiras. 2003. 103 f. Dissertação (Mestrado em Administração) – Universidade Federal do Rio Grande do Sul, Porto Alegre.

MATARAZZO, Dante Carmine. **Análise financeira de balanços**: abordagem básica e gerencial. São Paulo: Atlas, 2003.

MAYER, Thomas. **Doing economic research**: essays on the applied methodology of economics. Aldershot: Edward Elgar, 1995.

MORRISON, Jeffrey S. Seleção de variáveis no desenvolvimento de modelos. **Tecnologia de**

Crédito - Serasa. [S.l.], n. 57, p. 24-40, 2006.

NEOPHYTOU, Evridiki *et al.* *Predicting corporate failure: empirical evidence for the UK. University of Southampton Working Paper*, 2000. Disponível em: <<http://www.management.soton.ac.uk/research/publications/documents/01-173.pdf>>. Acesso em: 23 dez. 2007.

NIYAMA, Jorge K.; GOMES, Amaro L. O. **Contabilidade de instituições financeiras**. São Paulo: Atlas, 2005.

OHLSON, James A. *Financial ratios and the probabilistic prediction of bankruptcy*. **Journal of Accounting Research**, [S.l.], v. 18, n. 1, p. 109-131, Spring 1980.

PARENTE, Guilherme G.C. As novas normas de classificação de crédito e o disclosure das provisões: uma abordagem introdutória. In: SEMANA DE CONTABILIDADE DO BANCO CENTRAL DO BRASIL, 9., **Anais...** São Paulo, 2000.

PESTANA, Maria Helena; GAGEIRO, João Nunes. **Análise de dados para ciências sociais: a complementaridade do SPSS**. 3. ed. Lisboa: Edições Silabo, 2003.

PLATT, Harlan D.; PLATT, Marjorie B. *Business cycle effects on state corporate failure rates*. **Journal of Economics and Business**, [S.l.], v. 46, p. 113-127, 1994.

SANTOS, José Odílio. **Análise de crédito: empresas e pessoas físicas**. São Paulo: Atlas, 2000.

SAUNDERS, Anthony. **Administração de instituições financeiras**. São Paulo: Atlas, 2000.

SAUNDERS, Anthony. *Comments on "The macroeconomic impact of bank capital requirements in emerging economies: past evidence to assess the future"*. **Journal of Banking and Finance**, [S.l.], v. 26, p. 905-907, 2002.

SCHRICKEL, Wolfgang Kurt. **Análise de crédito: concessão e gerência de empréstimos**. 5. ed. São Paulo: Atlas, 2000.

SECURATO, José Roberto. **Decisões financeiras em condições de risco**. São Paulo: Atlas, 1996.

SERVIGNY, Arnaud de; RENAULT, Oliver. *Measuring and managing credit risk*. New York: McGraw-Hill, 2004.

SICSÚ, Abraham Laredo. Desenvolvimento de um sistema de *credit scoring*. In: DUARTE Jr.; VARGA, Gyorgy (Org.). **Gestão de riscos no Brasil**. Rio de Janeiro: Financial Consultoria, 2003.

SILVA, José Pereira da. **Gestão e análise de risco de crédito**. 3. ed. São Paulo: Atlas, 2000.

SPSSTRAINING. *Advanced statistical analysis using SPSS*. Apostila do curso de SPSS ministrado em 2003.

THOMAS, Lyn C. *et al.* **Credit scoring and its applications**. Philadelphia: SIAM, 2002.

VEIGA, Rafael Paschoarelli. **Um modelo para o cálculo da probabilidade de default empregando árvores binomiais**. 2006. 195 f. Tese (Doutorado em Administração) – Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo, São Paulo.

VERRONE, Marco Antônio Guimarães. **Basiléia II no Brasil**: uma reflexão com foco na regulação bancária para risco de crédito – Resolução CMN 2.682/99. 2007. 155 f. Dissertação (Mestrado em Administração) – Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo, São Paulo.

VICENTE, Ernesto Fernando Rodrigues. **A estimativa do risco na constituição da PDD**. 2001. 163 f. Dissertação (Mestrado em Controladoria e Contabilidade) - Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo, São Paulo.

WESTGAARD, Sjur; WIJST, Nico van der. *Default probabilities in a corporate bank portfolio: A logistic model approach*. **European Journal of Operational Research**, [S.l.], v. 135, p. 338-349, 2001.

WITKOWSKA, Dorota. *Discrete choice model application to the credit risk evaluation*. **International Advances in Economic Research**, [S.l.], v. 12, p. 33-42, 2006.

WOOLDRIDGE, Jeffrey M. **Econometric analysis of cross section and panel data**. Cambridge: MIT, 2002.

WOOLDRIDGE, Jeffrey M. **Introductory econometrics – a modern approach**. 3 ed. Ohio: Thomson South-Western, 2006.

APÊNDICES

- APÊNDICE 1 – CONFIGURAÇÃO SIMPLIFICADA DA DETERMINAÇÃO DA PARCELA MATERIAL
- APÊNDICE 2 – LISTA DAS VARIÁVEIS EXPLICATIVAS TESTADAS
- APÊNDICE 3 – ESTATÍSTICAS DESCRITIVAS DAS VARIÁVEIS EXPLICATIVAS MÉTRICAS
- APÊNDICE 4 – MODELO OBTIDO ATRAVÉS DO PROCEDIMENTO *STEPWISE MÉTODO FORWARD LR*

APÊNDICE 1 – CONFIGURAÇÃO SIMPLIFICADA DA DETERMINAÇÃO DA PARCELA MATERIAL

Considere um conjunto de 20 registros (CNPJ_IF) com informações de percentual de atraso¹⁴⁰ em 10 datas-base (DB), como mostra a Tabela 1:

Tabela 1 – Percentual da dívida em atraso nas datas-base, por registro

Registro	DB_1	DB_2	DB_3	DB_4	DB_5	DB_6	DB_7	DB_8	DB_9	DB_10
CNPJ_IF_1					1%	1%	1%	0%	1%	1%
CNPJ_IF_2			1%	2%	4%	2%	2%			
CNPJ_IF_3			2%	3%	2%	0%				
CNPJ_IF_4		2%	2%	2%	1%	1%	1%			
CNPJ_IF_5		2%	2%	3%	2%	1%				
CNPJ_IF_6				2%	3%	3%	4%	5%	6%	
CNPJ_IF_7			5%	6%	4%	3%	2%			
CNPJ_IF_8							5%	7%	8%	10%
CNPJ_IF_9					5%	10%	10%	25%	20%	
CNPJ_IF_10	5%	6%	7%	8%						
CNPJ_IF_11		5%	10%	30%	40%	50%	100%			
CNPJ_IF_12			10%	12%	7%	5%	2%			
CNPJ_IF_13				10%	15%	20%	35%	100%		
CNPJ_IF_14		10%	30%	40%	50%	40%				
CNPJ_IF_15				20%	12%	18%	20%	25%		
CNPJ_IF_16			20%	23%	40%	60%	50%			
CNPJ_IF_17				50%	60%	30%	30%			
CNPJ_IF_18					50%	70%	60%	60%	100%	
CNPJ_IF_19							100%	100%	100%	100%
CNPJ_IF_20						100%	50%	30%	0%	10%

Para todos os registros sem atraso na DB_1¹⁴¹, a primeira data-base com atraso (“t”) está hachurada em rosa e a terceira data-base subsequente (“t+3”), em azul. Note-se que o registro CNPJ_IF_10 não está marcado, pois já se encontrava em atraso no início do período em estudo.

Tomando os percentuais de atraso da primeira ocorrência (marcados em rosa), procede-se à formação das classes de atraso (nesta configuração, equivalentes aos quintis), conforme mostra a Tabela 2:

¹⁴⁰ Para simplificar, não será mencionado que o atraso deverá ser acima de 90 dias.

¹⁴¹ Para se determinar a primeira ocorrência de atraso, é necessário que não haja atraso na primeira data-base em estudo.

Tabela 2 – Classes por atraso – quintis

Classe	Definição
1	0% = atraso < 2%
2	2% = atraso < 5%
3	5% = atraso < 10%
4	10% = atraso < 32%
5	32% = atraso = 100%

Utilizando a mesma Tabela 2, classificam-se os percentuais de atraso de cada registro em “t” (hachurados em rosa na Tabela 1) e em “t+3” (hachurados em azul na Tabela 1), obtendo a Tabela 3:

Tabela 3 – Registros com percentuais de atraso em classes, em “t” e em “t+3”

Registro	Classe em "t"	Classe em "t+3"
CNPJ_IF_1	1	1
CNPJ_IF_2	1	2
CNPJ_IF_3	2	1
CNPJ_IF_4	2	1
CNPJ_IF_5	2	2
CNPJ_IF_6	2	2
CNPJ_IF_7	3	2
CNPJ_IF_8	3	4
CNPJ_IF_9	3	4
CNPJ_IF_11	3	5
CNPJ_IF_12	4	3
CNPJ_IF_13	4	5
CNPJ_IF_14	4	5
CNPJ_IF_15	4	4
CNPJ_IF_16	4	5
CNPJ_IF_17	5	4
CNPJ_IF_18	5	5
CNPJ_IF_19	5	5
CNPJ_IF_20	5	1

Esses mesmos dados podem ser apresentados na forma de uma matriz de migração de classes de atraso de “t” para “t+3”, conforme mostra a Tabela 4:

Tabela 4 – Matriz de migração de classes de atraso de “t” para “t+3”

		Classe em "t+3"							
		1	2	3	4	5	Total	Piora	
Classe em "t"	1	1	1				2	1	
	2	2	2				4	0	
	3		1		2	1	4	3	
	4			1	1	3	5	3	
	5	1			1	2	4	0	
	Total	4	4	1	4	6	19	7	

Cada célula apresenta o número de casos existentes naquele tipo de migração de classes (de “t” para “t+3”), sendo que as células à direita das células hachuradas representam os casos que migraram, em “t+3”, para classes com maior percentual de atraso do que em “t”. Tomando-se apenas esses casos, constrói-se a Tabela 5:

Tabela 5 – Tabela resumo dos casos em que houve migração para classes percentual de maior atraso

Classe em "t"	% Piora em relação ao total	Variação de piora em relação
1	50,0	
2	0,0	-50,0
3	75,0	75,0
4	60,0	-15,0
5	0,0	-60,0

Assim, identifica-se a classe 3 como sendo aquela que apresentou maior variação de “piora” em relação à classe anterior (75,0 p.p.).

Lembrando que a classe 3 refere-se aos percentuais de atraso entre 5% e 10% (exclusive), vide Tabela 2, adota-se 5% como parcela material, nesta configuração simplificada.

APÊNDICE 2 – LISTA DAS VARIÁVEIS EXPLICATIVAS TESTADAS

VARIÁVEL	DESCRIÇÃO
PCPELPPL	(passivo circulante + passivo exigível a longo prazo) / patrimônio líquido
PCPCPELP	passivo circulante / (passivo circulante + passivo exigível a longo prazo)
APPL	ativo permanente / patrimônio líquido
APPLPELP	ativo permanente / (patrimônio líquido + passivo exigível a longo prazo)
ACPCAT	(ativo circ. - passivo circ.) / ativo total = capital circulante líquido / ativo total
ACARLPPC	(ativo circ. + ativo realiz. longo prazo) / (passivo circ. + passivo exig. longo prazo)
ACestPC	(ativo circulante - estoques) / passivo circulante
VAT	receitas líquidas / ativo total
ACV	ativo circulante / receitas líquidas
LLV	lucro líquido / receitas líquidas
LLPLLL	lucro líquido / (patrimônio líquido - lucro líquido)
EPCPELP	EBITDA / (passivo circulante + passivo exigível a longo prazo)
EAT	EBITDA / ativo total
PLAT	patrimônio líquido / ativo total
PCAT	passivo circulante / ativo total
LLPCPELP	lucro líquido / (passivo circulante + passivo exigível a longo prazo)
LN_AT	logarítmo natural do ativo total
LN_TR	logarítmo natural do tempo de relacionamento entre a empresa e a IF
LN_TA	logarítmo natural do tempo de abertura da empresa
Selic	taxa de juros SELIC (%a.a.)
IGPM	índice geral de preços - mercado (%a.a.)
PIB	taxa de variação do PIB real (%a.a.)
s1	sede em AL
s2	sede em AM
s3	sede em BA
s4	sede em CE
s5	sede nos demais estados
s6	sede em DF
s7	sede em ES
s8	sede em GO
s9	sede em MG
s10	sede em MT
s11	sede em PA
s12	sede em PE

VARIÁVEL	DESCRIÇÃO
s13	sede em PR
s14	sede em RJ
s15	sede em RS
s16	sede em SC
s17b	sede em SP - variável basal
C1	controle acionário alemão
C2	controle acionário americano/brasileiro
C3	controle acionário americano
C4	controle acionário anglo-holandês
C5	controle acionário belga
C6b	controle acionário brasileiro - variável basal
C7	demaís controles acionários
C8	controle acionário espanhol
C9	controle acionário estatal
C10	controle acionário francês
C11	controle acionário holandês
C12	controle acionário inglês
C13	controle acionário italiano
C14	controle acionário japonês
C15	controle acionário luxemburguês
C16	controle acionário mexicano
C17	controle acionário português/espanhol
C18	controle acionário português
C19	controle acionário sueco
C20	controle acionário suíço
Set1	setor de atacado
Set2	setor de auto-indústria
Set3	setor de bens de capital
Set4	setor de bens de consumo
Set5	setor de comunicações
Set6	setor de indústria da construção
Set7	setor de indústria digital
Set8	setor de diversos

VARIÁVEL	DESCRIÇÃO
Set9	setor de eletroeletrônico
Set10b	setor de energia - variável basal
Set11	setor de farmacêutico
Set12	setor de mineração
Set13	setor de papel e celulose
Set14	setor de pendente - nova classif.
Set15	setor de produção agropecuária
Set16	setor de química e petroquímica
Set17	setor de serviços
Set18	setor de siderurgia e metalurgia
Set19	setor de telecomunicações
Set20	setor de têxteis
Set21	setor de transporte
Set22	setor de varejo
bols	negócios em bolsa

APÊNDICE 3 – ESTATÍSTICAS DESCRITIVAS DAS VARIÁVEIS EXPLICATIVAS MÉTRICAS

Variáveis	Default	Média	DP	N	t	p-value
LN_TA	não	3,148571	0,693867	7713	-1,5258	0,127099
	sim	3,223501	0,697578	205		
LN_TR	não	1,803779	1,186629	7713	-4,05941	6,77E-05
	sim	2,045303	0,829614	205		
PCPELPPL	não	2,654679	11,39223	7713	-1,80604	0,07095
	sim	4,10352	8,973114	205		
PCPCPELP	não	0,601622	0,224135	7713	2,515701	0,011899
	sim	0,561759	0,215573	205		
APPL	não	1,391469	4,706324	7713	-3,0579	0,002504
	sim	2,22449	3,824196	205		
APPLPELP	não	0,745986	3,475961	7713	-0,00407	0,996752
	sim	0,746975	0,823613	205		
ACPCAT	não	0,072333	0,193373	7713	3,300086	0,001135
	sim	0,016722	0,239207	205		
ACARLPPC	não	1,059176	0,789593	7713	1,22687	0,219908
	sim	0,989985	1,037783	205		
ACestPC	não	8,264941	141,7168	7713	0,728798	0,466147
	sim	1,050899	1,478488	205		
VAT	não	1,318249	1,097199	7713	3,493953	0,000575
	sim	1,100838	0,872784	205		
ACV	não	0,425393	0,318985	7713	-0,61175	0,540722
	sim	0,43939	0,458621	205		
LLV	não	0,043188	0,204189	7713	1,604175	0,110202
	sim	-0,0029	0,409991	205		
LLPLLL	não	0,14591	3,310893	7713	0,262515	0,792931
	sim	0,084819	2,290669	205		
EPCPELP	não	0,270387	0,310081	7713	2,02462	0,04294
	sim	0,225814	0,347561	205		
EAT	não	0,128655	0,115411	7713	3,006413	0,002652
	sim	0,104102	0,115338	205		
PLAT	não	0,382828	0,260135	7713	3,075863	0,002379
	sim	0,301072	0,378191	205		
PCAT	não	0,361411	0,188482	7713	-1,35609	0,175108
	sim	0,379557	0,210956	205		
LLPCPELP	não	0,104242	0,223342	7713	3,558886	0,000375
	sim	0,047833	0,246933	205		
LN_AT	não	13,3003	1,23501	7713	-1,92095	0,054773
	sim	13,46828	1,26444	205		
SELIC	não	19,46784	3,466713	7713	-1,9186	0,056367
	sim	19,96293	3,651156	205		
IGPM	não	12,48231	8,925284	7713	-2,81179	0,004939
	sim	14,25771	8,820824	205		
PIB	não	3,07198	1,609033	7713	-1,48634	0,137228
	sim	3,241367	1,662357	205		

**APÊNDICE 4 – MODELO OBTIDO ATRAVÉS DO PROCEDIMENTO *STEPWISE*
MÉTODO *FORWARD LR***

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95,0% C.I. for EXP(B)	
								Lower	Upper
Step 1	LN_TR	,237	,071	11,098	1	,001	1,267	1,102	1,457
	ACPCAT	-1,198	,403	8,818	1	,003	,302	,137	,666
	PLAT	-,557	,231	5,797	1	,016	,573	,364	,902
	IGPM	,016	,008	4,059	1	,044	1,016	1,000	1,033
	s2	1,664	,368	20,392	1	,000	5,278	2,564	10,866
	C3	,624	,261	5,736	1	,017	1,867	1,120	3,113
	C10	,837	,311	7,248	1	,007	2,310	1,256	4,249
	C17	1,337	,531	6,343	1	,012	3,808	1,345	10,779
	Set5	1,507	,412	13,359	1	,000	4,512	2,011	10,120
	Set6	1,127	,265	18,044	1	,000	3,085	1,835	5,189
	Set8	2,307	,430	28,730	1	,000	10,048	4,322	23,362
	Set20	1,439	,319	20,390	1	,000	4,218	2,258	7,880
	Constant	-4,392	,217	407,737	1	,000	,012		

a. Variable(s) entered on step 1: LN_TR, ACPCAT, PLAT, IGPM, s2, C3, C10, C17, Set5, Set6, Set8, Set20.

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	1792,234 ^a	,014	,065

a. Estimation terminated at iteration number 7 because parameter estimates changed by less than ,001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	7,588	8	,475

Contingency Table for Hosmer and Lemeshow Test

		DEFAULT = 0		DEFAULT = 1		Total
		Observed	Expected	Observed	Expected	
Step 1	1	785	785,408	7	6,592	792
	2	782	782,515	10	9,485	792
	3	782	781,632	11	11,368	793
	4	771	779,010	21	12,990	792
	5	778	777,438	14	14,562	792
	6	777	775,790	15	16,210	792
	7	779	773,734	13	18,266	792
	8	770	770,507	22	21,493	792
	9	767	762,487	25	29,513	792
	10	722	724,479	67	64,521	789

ANEXOS

ANEXO A – RESOLUÇÃO Nº. 2.682/99

ANEXO B – TABELA COM PESQUISAS ANTERIORES – *Predicting corporate
bankruptcy: whither do we stand?*

ANEXO C – QUADRO-RESUMO DOS ÍNDICES (MATARAZZO, 2003, p. 152)

ANEXO A**RESOLUÇÃO No. 2.682/99**

Dispõe sobre critérios de classificação das operações de crédito e regras para constituição de provisão para créditos de liquidação duvidosa.

O BANCO CENTRAL DO BRASIL, na forma do art. 9º da Lei nº 4.595, de 31 de dezembro de 1964, torna público que o CONSELHO MONETÁRIO NACIONAL, em sessão realizada em 21 de dezembro de 1999, com base no art. 4º, incisos XI e XII, da citada Lei,

R E S O L V E U:

Art. 1º Determinar que as instituições financeiras e demais instituições autorizadas a funcionar pelo Banco Central do Brasil devem classificar as operações de crédito, em ordem crescente de risco, nos seguintes níveis:

- I - nível AA;
- II - nível A;
- III - nível B;
- IV - nível C;
- V - nível D;
- VI - nível E;
- VII - nível F;
- VIII - nível G;
- IX - nível H.

Art. 2º A classificação da operação no nível de risco correspondente é de responsabilidade da instituição detentora do crédito e deve ser efetuada com base em critérios consistentes e verificáveis, amparada por informações internas e externas, contemplando, pelo menos, os seguintes aspectos:

- I - em relação ao devedor e seus garantidores:
 - a) situação econômico-financeira;
 - b) grau de endividamento;
 - c) capacidade de geração de resultados;
 - d) fluxo de caixa;
 - e) administração e qualidade de controles;

- f) pontualidade e atrasos nos pagamentos;
- g) contingências;
- h) setor de atividade econômica;
- i) limite de crédito;

II - em relação à operação:

- a) natureza e finalidade da transação;
- b) características das garantias, particularmente quanto à suficiência e liquidez;
- c) valor.

Parágrafo único. A classificação das operações de crédito de titularidade de pessoas físicas deve levar em conta, também, as situações de renda e de patrimônio bem como outras informações cadastrais do devedor.

Art. 3º A classificação das operações de crédito de um mesmo cliente ou grupo econômico deve ser definida considerando aquela que apresentar maior risco, admitindo-se excepcionalmente classificação diversa para determinada operação, observado o disposto no art. 2º, inciso II.

Art. 4º A classificação da operação nos níveis de risco de que trata o art. 1º deve ser revista, no mínimo:

I - mensalmente, por ocasião dos balancetes e balanços, em função de atraso verificado no pagamento de parcela de principal ou de encargos, devendo ser observado o que segue:

- a) atraso entre 15 e 30 dias: risco nível B, no mínimo;
- b) atraso entre 31 e 60 dias: risco nível C, no mínimo;
- c) atraso entre 61 e 90 dias: risco nível D, no mínimo;
- d) atraso entre 91 e 120 dias: risco nível E, no mínimo;
- e) atraso entre 121 e 150 dias: risco nível F, no mínimo;
- f) atraso entre 151 e 180 dias: risco nível G, no mínimo;
- g) atraso superior a 180 dias: risco nível H;

II - com base nos critérios estabelecidos nos arts. 2º e 3º:

a) a cada seis meses, para operações de um mesmo cliente ou grupo econômico cujo montante seja superior a 5% (cinco por cento) do patrimônio líquido ajustado;

b) uma vez a cada doze meses, em todas as situações, exceto na hipótese prevista no art. 5º.

Parágrafo 1º As operações de adiantamento sobre contratos de câmbio, as de financiamento à importação e aquelas com prazos inferiores a um mês, que apresentem atrasos superiores a trinta dias, bem

como o adiantamento a depositante a partir de trinta dias de sua ocorrência, devem ser classificados, no mínimo, como de risco nível G.

Parágrafo 2º Para as operações com prazo a decorrer superior a 36 meses admite-se a contagem em dobro dos prazos previstos no inciso I.

Parágrafo 3º O não atendimento ao disposto neste artigo implica a reclassificação das operações do devedor para o risco nível H, independentemente de outras medidas de natureza administrativa.

Art. 5º As operações de crédito contratadas com cliente cuja responsabilidade total seja de valor inferior a R\$50.000,00 (cinquenta mil reais) podem ter sua classificação revista de forma automática unicamente em função dos atrasos consignados no art. 4º, inciso I, desta Resolução, observado que deve ser mantida a classificação original quando a revisão corresponder a nível de menor risco.

Parágrafo 1º O Banco Central do Brasil poderá alterar o valor de que trata este artigo.

Parágrafo 2º O disposto neste artigo aplica-se às operações contratadas até 29 de fevereiro de 2000, observados o valor referido no caput e a classificação, no mínimo, como de risco nível A.

Art. 6º A provisão para fazer face aos créditos de liquidação duvidosa deve ser constituída mensalmente, não podendo ser inferior ao somatório decorrente da aplicação dos percentuais a seguir mencionados, sem prejuízo da responsabilidade dos administradores das instituições pela constituição de provisão em montantes suficientes para fazer face a perdas prováveis na realização dos créditos:

- I - 0,5% (meio por cento) sobre o valor das operações classificadas como de risco nível A;
- II - 1% (um por cento) sobre o valor das operações classificadas como de risco nível B;
- III - 3% (três por cento) sobre o valor das operações classificadas como de risco nível C;
- IV - 10% (dez por cento) sobre o valor das operações classificadas como de risco nível D;
- V - 30% (trinta por cento) sobre o valor das operações classificadas como de risco nível E;
- VI - 50% (cinquenta por cento) sobre o valor das operações classificadas como de risco nível F;
- VII - 70% (setenta por cento) sobre o valor das operações classificadas como de risco nível G;
- VIII - 100% (cem por cento) sobre o valor das operações classificadas como de risco nível H.

Art. 7º A operação classificada como de risco nível H deve ser transferida para conta de compensação, com o correspondente débito em provisão, após decorridos seis meses da sua classificação nesse nível de risco, não sendo admitido o registro em período inferior.

Parágrafo único. A operação classificada na forma do disposto no caput deste artigo deve permanecer registrada em conta de compensação pelo prazo mínimo de cinco anos e enquanto não esgotados todos os procedimentos para cobrança.

Art. 8º A operação objeto de renegociação deve ser mantida, no mínimo, no mesmo nível de risco em que estiver classificada, observado que aquela registrada como prejuízo deve ser classificada como de risco nível H.

Parágrafo 1º Admite-se a reclassificação para categoria de menor risco quando houver amortização significativa da operação ou quando fatos novos relevantes justificarem a mudança do nível de risco.

Parágrafo 2º O ganho eventualmente auferido por ocasião da renegociação deve ser apropriado ao resultado quando do seu efetivo recebimento.

Parágrafo 3º Considera-se renegociação a composição de dívida, a prorrogação, a novação, a concessão de nova operação para liquidação parcial ou integral de operação anterior ou qualquer outro tipo de acordo que implique na alteração nos prazos de vencimento ou nas condições de pagamento originalmente pactuadas.

Art. 9º É vedado o reconhecimento no resultado do período de receitas e encargos de qualquer natureza relativos a operações de crédito que apresentem atraso igual ou superior a sessenta dias, no pagamento de parcela de principal ou encargos.

Art. 10. As instituições devem manter adequadamente documentadas sua política e procedimentos para concessão e classificação de operações de crédito, os quais devem ficar à disposição do Banco Central do Brasil e do auditor independente.

Parágrafo único. A documentação de que trata o caput deste artigo deve evidenciar, pelo menos, o tipo e os níveis de risco que se dispõe a administrar, os requerimentos mínimos exigidos para a concessão de empréstimos e o processo de autorização.

Art. 11. Devem ser divulgadas em nota explicativa às demonstrações financeiras informações detalhadas sobre a composição da carteira de operações de crédito, observado, no mínimo:

I - distribuição das operações, segregadas por tipo de cliente e atividade econômica;

II - distribuição por faixa de vencimento;

III - montantes de operações renegociadas, lançados contra prejuízo e de operações recuperadas, no exercício.

Art. 12. O auditor independente deve elaborar relatório circunstanciado de revisão dos critérios adotados pela instituição quanto à classificação nos níveis de risco e de avaliação do provisionamento registrado nas demonstrações financeiras.

Art. 13. O Banco Central do Brasil poderá baixar normas complementares necessárias ao cumprimento do disposto nesta Resolução, bem como determinar:

I - reclassificação de operações com base nos critérios es-

tabelecidos nesta Resolução, nos níveis de risco de que trata o art. 1º;

II - provisionamento adicional, em função da responsabilidade do devedor junto ao Sistema Financeiro Nacional;

III - providências saneadoras a serem adotadas pelas instituições, com vistas a assegurar a sua liquidez e adequada estrutura patrimonial, inclusive na forma de alocação de capital para operações de classificação considerada inadequada;

IV - alteração dos critérios de classificação de créditos, de contabilização e de constituição de provisão;

V - teor das informações e notas explicativas constantes das demonstrações financeiras;

VI - procedimentos e controles a serem adotados pelas instituições.

Art. 14. O disposto nesta Resolução se aplica também às operações de arrendamento mercantil e a outras operações com características de concessão de crédito.

Art. 15. As disposições desta Resolução não contemplam os aspectos fiscais, sendo de inteira responsabilidade da instituição a observância das normas pertinentes.

Art. 16. Esta Resolução entra em vigor na data da sua publicação, produzindo efeitos a partir de 1º de março de 2000, quando ficarão revogadas as Resoluções nºs 1.748, de 30 de agosto de 1990, e 1.999, de 30 de junho de 1993, os arts. 3º e 5º da Circular nº 1.872, de 27 de dezembro de 1990, a alínea "b" do inciso II do art. 4º da Circular nº 2.782, de 12 de novembro de 1997, e o Comunicado nº 2.559, de 17 de outubro de 1991.

Brasília, 21 de dezembro de 1999

Arminio Fraga Neto
Presidente

ANEXO B

TABELA COM PESQUISAS ANTERIORES – *Predicting corporate bankruptcy: whither do we stand?*

Table 1
(Summary of Previous Research Attributes and Findings)

No.	Author & Year	Model	OPA (%)	TypeI (%)	TypeII (%)	ES	TS	Ind. Var.	Country	Years	Firm Type	Publishing Journal
1	Altman (1968)	MDA	95	6	3	66	25	FR	USA	46 - 65	Manufac. Ind. (ltd.)	The Jr. of Finance
2	Altman et al. (1977)	MDA	92.8	3.77	10.34	111	111	FR	USA	64 - 74	Manf. and retail (ltd.)	Jr. of Banking & Fin.
3	Altman & Varetto (1994)	MDA	NA	13.6	9.7	1212	450	FR	Italy	85 - 92	Industrial (ltd.)	Jr. of Banking & Fin.
4	Altman & Varetto (1994)	NN	NA	13.8	10.6	1212	450	FR	Italy	85 - 92	Industrial (ltd.)	Jr. of Banking & Fin.
5	Aziz et al. (1988)	MDA	88.8	NA	NA	98	NA	CF	USA	71 - 82	Mix. Ind. (ltd.)	J of Manag Studies
6	Aziz et al. (1988)	Logit	91.8	14.3	2.1	98	NA	CF	USA	71 - 82	Mix. Ind. (ltd.)	J of Manag Studies
7	Aziz et al. (1988)	BSDM	91.8	NA	NA	98	NA	CF	USA	71 - 82	Mix. Ind. (ltd.)	J of Manag Studies
8	Back et al. (1996)	MDA	85.14	13.51	16.22	74	NA	FR	Finland	86 - 89	Mix. Ind. (ltd.)	Tech. Report Turku
9	Back et al. (1996)	Logit	96.49	13.51	13.51	74	NA	FR	Finland	86 - 89	Mix. Ind. (ltd.)	Tech. Report Turku
10	Back et al. (1996)	NN	97.3	5.26	0	74	NA	FR	Finland	86 - 89	Mix. Ind. (ltd.)	Tech. Report Turku
11	Back et al. (1996)	GA	97.3	5.26	0	74	NA	FR	Finland	86 - 89	Mix. Ind. (ltd.)	Tech. Report Turku
12	Beynon and Peel (2001)	MDA	78.3	16.7	26.7	60	30	Mix	UK	NA	Manufac. Ind. (ltd.)	Omega
13	Beynon and Peel (2001)	Logit	80	16.7	23.3	60	30	Mix	UK	NA	Manufac. Ind. (ltd.)	Omega
14	Beynon and Peel (2001)	RPA	93.3	10	3.3	60	30	Mix	UK	NA	Manufac. Ind. (ltd.)	Omega
15	Beynon and Peel (2001)	RS	91.7	13.3	3.3	60	30	Mix	UK	NA	Manufac. Ind. (ltd.)	Omega
16	Booth (1983)	MDA	85	18	12	44	26	Mix	Australia	64 - 79	Mix. Ind. (ltd.)	Jr. of B. Fin. & Acc.
17	Booth (1983)	BSDM	85	18	12	44	26	Mix	Australia	64 - 79	Mix. Ind. (ltd.)	Jr. of B. Fin. & Acc.
18	Brockman and Turtle (2003)	MDA	74.5	NA	NA	NA	NA	Mix	USA	89 - 98	Mix. Ind. (ltd.)	J of Fin. Eco.
19	Brockman and Turtle (2003)	Logit	85	NA	NA	NA	NA	Mix	USA	89 - 98	Mix. Ind. (ltd.)	J of Fin. Eco.
20	Brockman and Turtle (2003)	Credit	85	NA	NA	NA	NA	Mix	USA	89 - 98	Mix. Ind. (ltd.)	J of Fin. Eco.
21	Casey & Bartczak (1984)	Univariate	75	10	27	290	NA	CF	USA	71 - 82	Mix. Ind. (ltd.)	Harvard Bus. Review
22	Casey & Bartczak (1984)	MDA	86	17	13	290	NA	FR	USA	71 - 82	Mix. Ind. (ltd.)	Harvard Bus. Review
23	Casey & Bartczak (1984)	Cash	75	10	27	290	NA	CF	USA	71 - 82	Mix. Ind. (ltd.)	Harvard Bus. Review
24	Coats and Fant (1993)	MDA	87.9	36.2	0	282	NA	FR	USA	70 - 89	Mix. Ind. (ltd.)	Fin. Management
25	Coats and Fant (1993)	NN	95	10.6	2.1	282	NA	FR	USA	70 - 89	Mix. Ind. (ltd.)	Fin. Management
26	Dimitras et al. (1999)	MDA	90	12.5	7.5	80	38	FR	Greece	86 - 93	Mix. Ind. (ltd.)	Eur J of Oper. Res.
27	Dimitras et al. (1999)	Logit	90	7.5	12.5	80	38	FR	Greece	86 - 93	Mix. Ind. (ltd.)	Eur J of Oper. Res.
28	Dimitras et al. (1999)	RS	97.5	2.5	2.5	80	38	FR	Greece	86 - 93	Mix. Ind. (ltd.)	Eur J of Oper. Res.
29	El Hennawy & Morris (1983)	MDA	97.72	4.55	0	44	44	Mix	UK	60 - 71	Mix. Ind. (ltd.)	Jr. of B. Fin. & Acc.
30	Foreman (2002)	Logit	97.4	14.29	0	77	14	FR	USA	1999	Telecom. Ind.	Jr. of Eco. & Bus.
31	Frydman et al. (1985)	MDA	74	9	17	200	NA	FR	USA	71 - 81	Mix. Ind. (ltd.)	The Jr. of Finance

No.	Author & Year	Model	OPA (%)	TypeI (%)	TypeII (%)	ES	TS	Ind. Var.	Country	Years	Firm Type	Publishing Journal
32	Frydman et al. (1985)	RPA	89	9	2	200	NA	FR	USA	71 - 81	Mix. Ind. (ltd.)	The Jr. of Finance
33	Gombola et al. (1987)	MDA	89	NA	NA	77	NA	FR	USA	70 - 82	Manf. and retail (ltd.)	Fin. Management
34	Gombola et al. (1987)	BSDM	89	NA	NA	77	NA	FR	USA	70 - 82	Manf. and retail (ltd.)	Fin. Management
35	Jo et al. (1997)	MDA	82.22	NA	NA	542	NA	Mix	Korea	91 - 93	Mix. Ind. (ltd.)	Exp. Sys. With App.
36	Jo et al. (1997)	NN	83.79	NA	NA	542	NA	Mix	Korea	91 - 93	Mix. Ind. (ltd.)	Exp. Sys. With App.
37	Jo et al. (1997)	CBR	81.52	NA	NA	542	NA	Mix	Korea	91 - 93	Mix. Ind. (ltd.)	Exp. Sys. With App.
38	Kahya & Theodossiou (1999)	MDA	77.8	31	17	189	NA	FR	USA	74 - 91	Manf. and retail (ltd.)	Rev of Q Fin & Acc
39	Kahya & Theodossiou (1999)	Logit	77.2	33	16	189	NA	FR	USA	74 - 91	Manf. and retail (ltd.)	Rev of Q Fin & Acc
40	Kahya & Theodossiou (1999)	CUSUM	82.5	18	17	189	NA	FR	USA	74 - 91	Manf. and retail (ltd.)	Rev of Q Fin & Acc
41	Keasey & McGuinness (1990)	Logit	86	14	14	86	30	FR	UK	76 - 84	Mix. Ind. (ltd.)	Jr. of B. Fin. & Acc.
42	Laitinen and Laitinen (1998)	Logit	80.49	17.07	21.95	82	NA	Mix	Finland	86 - 91	Industrial (ltd.)	Jr. of B. Fin. & Acc.
43	Laitinen and Laitinen (1998)	Par Adj.	80.49	17.07	21.95	82	NA	Mix	Finland	86 - 91	Industrial (ltd.)	Jr. of B. Fin. & Acc.
44	Laitinen and Laitinen (1998)	Cash	58.54	41.46	41.46	82	NA	CF	Finland	86 - 91	Industrial (ltd.)	Jr. of B. Fin. & Acc.
45	Lin & Piesse (2001)	Univariate	79.22	28.12	2.22	77	NA	FR	UK	85 - 94	Mix. Ind. (ltd.)	Document UOLondon
46	Lin & Piesse (2001)	Logit	87	12.5	8.89	77	NA	FR	UK	85 - 94	Mix. Ind. (ltd.)	Document UOLondon
47	McGurr and DeVaney (1998)	MDA	74.1	NA	NA	112	NA	Mix	USA	89 - 93	Retail firms (ltd.)	Jr. of Bus Res
48	McGurr and DeVaney (1998)	Logit	67.2	NA	NA	112	NA	Mix	USA	89 - 93	Retail firms (ltd.)	Jr. of Bus Res
49	McGurr and DeVaney (1998)	Cash	68.43	NA	NA	112	NA	Mix	USA	89 - 93	Retail firms (ltd.)	Jr. of Bus Res
50	McKee & Lensberg (2002)	GA	82.6	6.8	10.3	291	NA	FR	USA	91 - 97	Mix. Ind. (ltd.)	Eur J of Oper. Res.
51	McKee & Lensberg (2002)	RS	82.6	6.8	10.3	291	NA	FR	USA	91 - 97	Mix. Ind. (ltd.)	Eur J of Oper. Res.
52	Messier & Hansen (1988)	RPA	100	NA	NA	32	16	FR	USA	75 - 76	NA	Management Sc.
53	Meyer & Pifer (1970)	LPM	80	3	0	60	18	FR	USA	48 - 65	Banks	The Jr. of Finance
54	Moyer (1977)	MDA	90.48	5	14	54	NA	Mix	USA	65 - 75	NA	Fin. Management
55	Moyer (1977)	BSDM	85.19	11	18	54	NA	Mix	USA	65 - 75	NA	Fin. Management
56	Neophytou et al. (2001)	Univariate	90	NA	NA	102	52	FR	UK	88 - 94	Industrial (ltd.)	Disc Paper Sthmpn
57	Neophytou et al. (2001)	Logit	93.75	8.33	4.17	102	52	FR	UK	88 - 94	Industrial (ltd.)	Disc Paper Sthmpn
58	Neophytou et al. (2001)	MDA	93.75	NA	NA	102	52	FR	UK	88 - 94	Industrial (ltd.)	Disc Paper Sthmpn
59	Neophytou et al. (2001)	NN	95.83	NA	NA	102	52	FR	UK	88 - 94	Industrial (ltd.)	Disc Paper Sthmpn
60	Park and Han (2002)	CBR	84.52	NA	NA	2144	NA	Mix	Korea	95 - 98	Mix. Ind. (ltd.)	Exp. Sys. With App.
61	Piesse & Wood (1992)	MDA	NA	25	34	48	48	FR	UK	73 - 86	Motor Compnts. (ltd.)	British Acc Review
62	Plat & Plat (1990)	Logit	90	7	14	171	68	Mix	USA	72 - 86	Mix. Ind. (ltd.)	Jr. of B. Fin. & Acc.
63	Pompe & Feelders (1997)	MDA	70	NA	NA	288	288	FR	Belgium	88 - 94	Constrn. ind. (ltd.)	Mic.Comp. in C Eng
64	Pompe & Feelders (1997)	RPA	70	NA	NA	288	288	FR	Belgium	88 - 94	Constrn. ind. (ltd.)	Mic.Comp. in C Eng
65	Pompe & Feelders (1997)	NN	73	NA	NA	288	288	FR	Belgium	88 - 94	Constrn. ind. (ltd.)	Mic.Comp. in C Eng
66	Salchenberger et al. (1992)	Logit	93.5	10	3	200	404	FR	USA	86 - 87	S & loan Association	Decision Science
67	Salchenberger et al. (1992)	NN	97	4	2	200	404	FR	USA	86 - 87	S & loan Association	Decision Science

Appendix

(List of Abbreviations used in the Study)

BSDM	Balance Sheet Decomposition Measure (Entropy Theory)
Cash	Cash Management Theory
CBR	Case-Based Reasoning
CF	Cash Flow
Const.	Construction
Credit	Credit Risk Theories (including 'Option Pricing' and 'Macroeconomic' theories)
CUSUM	Cumulative Sums Model (Time Series)
ES	Estimation Sample
FR	Financial Ratios
GA	Genetic Algorithms
Gamb.	Gambler's ruin theory
Ind.	Industry
Ind. Var.	Independent Variables
LPM	Linear Probability Model
Manf.	Manufacturing
MDA	Multiple Discriminant Analysis
NA	Not Available
NN	Neural Networks
Non-Fin.	Non-Financial
OPA	Overall Predictive Accuracy
Par. Adj.	Partial Adjustment Model (Time Series)
RPA	Recursive Partitioning (Decision Tree) Analysis
RS	Rough Sets Model
S & Loan	Saving and Loan
Telecom.	Telecommunications
TS	Test (or holdout) Sample
Type I	Type I error of classifying failed firms as non-failed
Type II	Type II error of classifying non-failed firms as failed

ANEXO C

QUADRO-RESUMO DOS ÍNDICES (MATARAZZO, 2003, p. 152)

SÍMBOLO	ÍNDICE	FÓRMULA	INDICA	INTERPRETAÇÃO
1. CT/PL	Estrutura de Capital • Participação de Capitais de Terceiros (Endividamento)	$\frac{\text{Capitais de Terceiros}}{\text{Patrimônio Líquido}} \times 100$	Quanto a empresa tomou de capitais de terceiros para cada \$ 100 de capital próprio.	Quanto menor, melhor.
2. PC/CT	• Composição do Endividamento	$\frac{\text{Passivo Circulante}}{\text{Capitais de Terceiros}} \times 100$	Qual o percentual de obrigações a curto prazo em relação às obrigações totais.	Quanto menor, melhor.
3. AP/PL	• Imobilização do Patrimônio Líquido	$\frac{\text{Ativo Permanente}}{\text{Patrimônio Líquido}} \times 100$	Quanto a empresa aplicou no Ativo Permanente para cada \$ 100 de Patrimônio Líquido.	Quanto menor, melhor.
4. AP/PL + ELP	• Imobilização dos Recursos não Correntes	$\frac{\text{Ativo Permanente}}{\text{Patrimônio Líquido} + \text{Exigível a Longo Prazo}} \times 100$	Que percentual dos Recursos não Correntes (Patrimônio Líquido e Exigível a Longo Prazo) foi destinado ao Ativo Permanente.	Quanto menor, melhor.
5. LG	Liquidez • Liquidez Geral	$\frac{\text{Ativo Circulante} + \text{Realizável a Longo Prazo}}{\text{Passivo Circulante} + \text{Exigível a Longo Prazo}}$	Quanto a empresa possui de Ativo Circulante + Realizável a Longo Prazo para cada \$ 1 de dívida total.	Quanto maior, melhor.
6. LC	• Liquidez Corrente	$\frac{\text{Ativo Circulante}}{\text{Passivo Circulante}}$	Quanto a empresa possui de Ativo Circulante para cada \$ 1 de Passivo Circulante.	Quanto maior, melhor.
7. LS	• Liquidez Seca	$\frac{\text{Disponível} + \text{Títulos a Receber} + \text{Outros Ativos de Rápida Conversibilidade}}{\text{Passivo Circulante}}$	Quanto a empresa possui de Ativo Líquido para cada \$ 1 de Passivo Circulante.	Quanto maior, melhor.
8. V/AT	Rentabilidade (ou Resultados) • Giro do Ativo	$\frac{\text{Vendas Líquidas}}{\text{Ativo}}$	Quanto a empresa vendeu para cada \$ 1 de investimento total.	Quanto maior, melhor.
9. LL/V	• Margem Líquida	$\frac{\text{Lucro Líquido}}{\text{Vendas Líquidas}} \times 100$	Quanto a empresa obtém de lucro para cada \$ 100 vendidos.	Quanto maior, melhor.
10. LL/AT	• Rentabilidade do Ativo	$\frac{\text{Lucro Líquido}}{\text{Ativo}} \times 100$	Quanto a empresa obtém de lucro para cada \$ 100 de investimento total.	Quanto maior, melhor.
11. LL/PL	• Rentabilidade do Patrimônio Líquido	$\frac{\text{Lucro Líquido}}{\text{Patrimônio Líquido Médio}} \times 100$	Quanto a empresa obtém de lucro para cada \$ 100 de capital próprio investido, em média, no exercício.	Quanto maior, melhor.