



Departamento de Métodos Quantitativos

Modelos multinível para dados binários: O efeito do sector de  
actividade no incumprimento de empresas portuguesas

Helena Margarida Bico Marques

Dissertação submetida como requisito parcial para obtenção do grau de  
Mestre em Prospecção e Análise de Dados

Orientadora:

Professora Doutora Manuela Magalhães Hill

Co-orientador:

Professor Doutor José Gonçalves Dias

Maio de 2010

## **Resumo e palavras-chave**

O estudo do incumprimento das empresas é um tópico que continua a despertar o interesse de investigadores e académicos, estando rodeado de um inegável interesse prático. Porém, a sua modelação parece ser um problema especialmente marcado por uma forte componente empírica e o próprio conceito de incumprimento pode até ser definido de diferentes formas. No presente trabalho, são acomodadas questões relativas à importância do sector de actividade como factor diferenciador da probabilidade de incumprimento e à inovação metodológica, tendo-se aplicado um modelo multinível com função de ligação *logit*, para além de uma regressão logística “tradicional”. A modelação foi em grande parte conduzida pelos dados disponíveis, tendo-se privilegiado uma lógica exploratória das variáveis identificadas em estudos recentes portugueses sobre o incumprimento de empresas.

Tanto com a utilização da regressão logística como com a modelação multinível, encontraram-se evidências de que o sector de actividade em que a empresa se insere constitui um factor diferenciador da probabilidade de incumprimento. Na regressão logística resultaram ainda como variáveis significativas rácios económico-financeiros de liquidez, estrutura de custos, solvência, rentabilidade e dimensão. A aplicação do modelo multinível com função de ligação *logit* permitiu detectar a existência de efeitos não observados ao nível do sector de actividade que implicam diferentes probabilidades de incumprimento das empresas portuguesas, possibilitando o cálculo da correlação intra-grupo (i.e., entre empresas do mesmo agrupamento de sectores de actividade).

*Classificação JEL:* C25, G33

*Palavras-chave:* incumprimento de empresas; sector de actividade; modelo *logit* multinível; regressão logística.

## **Abstract and keywords**

The study of company failure is a topic that continues to attract the attention of researchers and scholars, being surrounded by undeniable practical importance. However, it is not possible to identify a theoretical formulation of failure, which is an event that can even be defined in different ways. In fact, its modeling seems to be a problem especially characterized by a large empirical component in the sense that there is not a test whether there is an adhesion of a theoretical formulation to the data. In this study, one is bringing questions concerning the importance of the business sector as a differentiating factor in the probability of failure and methodological innovation, as it applies a multilevel logit, in addition to a "traditional" logistic regression. The modeling was largely driven by the available data and the exploring of variables identified in recent studies on the failure of Portuguese companies was privileged.

Both logistic regression and multilevel modeling provide evidence that the business sector in which the company operates is a key differentiator of the probability of failure. Logistic regression also allowed for the identification of significant variables like financial ratios concerning liquidity, cost structure, solvency, profitability, and size. The multilevel logit enabled the detection of unobserved effects at the business sector level and the calculation of intra-class correlation.

*Classificação JEL:* C25, G33

*Keywords:* company failure; business sector; multilevel logit model; logistic regression.

## **Agradecimentos**

Sem qualquer ordem específica, pois a falta de qualquer um destes elementos levaria a que o resultado não fosse este.

Aos Professores Manuela Magalhães Hill e José Gonçalves Dias, que tiveram a paciência de esperar por mim e de me auxiliar sempre que necessitei.

À Coface Serviços Portugal, S.A., que disponibilizou a informação fundamental para o trabalho que consegui desenvolver.

À minha entidade empregadora, que me disponibilizou um dos mais preciosos bens: tempo.

A todas as pessoas próximas, familiares e amigos, que fazem questão de perguntar como é que correm as coisas.

À pessoa que me permite fazer a seguinte citação:

“O que era preciso, então, era um subtil trabalho de ajustamento que trouxesse graduais correcções ao modelo, para o aproximar de uma possível realidade, e à realidade, para a aproximar do modelo. De facto, o grau de ductibilidade da natureza humana não é limitado como ele pensava num primeiro momento; e, em compensação, até mesmo o modelo mais rígido pode dar provas de uma qualquer elasticidade inesperada. Em suma, se o modelo não consegue transformar a realidade, a realidade deveria conseguir transformar o modelo”.

Palomar, Italo Calvino (Trad. João Reis)

## **Resumo executivo**

O estudo do incumprimento das empresas é um tópico que continua a despertar o interesse de investigadores e académicos, estando rodeado de um inegável interesse prático. Porém, a sua modelação parece ser um problema especialmente marcado por uma grande componente empírica e o próprio conceito de incumprimento pode até ser definido de diferentes formas. Neste trabalho são acomodadas questões relativas à importância do sector de actividade como factor diferenciador da probabilidade de incumprimento e à inovação metodológica, tendo-se aplicado um modelo multinível com função de ligação *logit*, para além de uma regressão logística “tradicional”. A modelação foi em grande parte conduzida pelos dados disponíveis, tendo-se privilegiado uma lógica exploratória das variáveis identificadas em estudos recentes portugueses sobre o incumprimento de empresas.

A base de dados utilizada no presente estudo foi disponibilizada pela Coface Serviços Portugal, S.A., dela constando elementos caracterizadores do tecido empresarial português, informações sobre processos de insolvência e dados de balanços. São consideradas como empresas em incumprimento as que registaram processos de insolvência em 2009, independentemente do estado do processo ou de quem o interpõe, entendendo-se que a existência de acções judiciais deste tipo constitui um indício forte de que existirá uma considerável probabilidade de a empresa não respeitar na íntegra as suas responsabilidades perante terceiros.

Tanto com a utilização da regressão logística quer com a modelação multinível, encontraram-se evidências de que o sector de actividade em que a empresa se insere constitui um factor diferenciador da probabilidade de incumprimento. Ainda que subsistam dúvidas sobre a qualidade do ajustamento global dos modelos estimados, é possível detectar a significância individual das variáveis que reflectem o sector de actividade. Os sectores que aparecem como mais problemáticos, independentemente da técnica estatística utilizada, são a Fabricação de Têxteis, a Indústria do Vestuário e Indústria de Couro, Madeira e Cortiça. Na regressão logística resultaram ainda como variáveis significativas rácios económico-financeiros de liquidez, estrutura de custos, solvência, rentabilidade e dimensão.

A aplicação do modelo multinível com função de ligação *logit* permitiu detectar a existência de efeitos não observados ao nível do sector de actividade que implicam diferentes probabilidades de incumprimento das empresas portuguesas, possibilitando o cálculo da correlação intra-grupo, no modelo estimado apenas com termo independente. Conclui-se pela não significância estatística do efeito fixo, mas foi possível obter as soluções para os efeitos aleatórios correspondentes aos sectores de actividade, sem que nenhum desses sectores tivesse de servir como sector de referência, como acontece no caso da regressão logística. A tentativa de estimar um modelo multinível com função de ligação *logit* considerando como potenciais variáveis explicativas as variáveis que resultaram estatisticamente significativas na regressão logística acabou por não ser totalmente bem sucedida. Apenas se conseguiram resultados considerando um rácio referente à estrutura de custos (juros e custos similares/dívidas a terceiros), e apenas se estimou um coeficiente aleatório estatisticamente significativo associado a essa variável no sector Comércio por grosso (inclui agentes), excepto de veículos automóveis e motociclos.

Em aberto fica a possibilidade de estudar a adequabilidade de um modelo multinível considerando no nível superior variáveis relativas ao sector de actividade distintas da mera média das variáveis consideradas ao nível das empresas, mas outras que constituam, por exemplo, indicadores de risco sectorial.

## Índice

Índice de Tabelas	ix
Índice de Figuras	xi
Introdução	1
1. O incumprimento das empresas	2
1.1. Apresentação da problemática e sua importância	2
1.2. Definição de incumprimento	4
1.3. Evolução das técnicas utilizadas na modelação do incumprimento de empresas	5
1.4. Abordagens portuguesas recentes	11
2. Modelos multinível	15
2.1. Especificação geral da modelação multinível	15
2.2. O modelo de regressão linear de dois níveis com uma variável explicativa em cada nível	16
2.3. O modelo de regressão de dois níveis com várias variáveis explicativas nos dois níveis	19
2.4. Modelos multinível para dados de resposta binária	20
2.5. A centragem das variáveis explicativas	23
3. Especificação e estimação do modelo	27
3.1. Seleção das variáveis explicativas e a consideração do sector de actividade	27
3.2. Estimação e diagnóstico da regressão logística	32
3.3. Estimação e diagnóstico do modelo multinível com função de ligação <i>logit</i>	34
4. Caracterização dos dados utilizados	38
4.1. Descrição da base de dados	38
4.2. Análise de valores omissos e de multicolinearidade	43
4.3. Análise descritiva e obtenção da amostra	46
4.4. Distribuição das potenciais variáveis explicativas na amostra final	54
5. Resultados empíricos	59
5.1. Regressão logística	59
5.2. Modelo multinível com função de ligação <i>logit</i>	63
5.3. Discussão comparativa dos resultados	71

Conclusão	75
Bibliografia	78
Anexos	84
Anexo 1 - Transformação proposta por Wolfinger e O'Connell (1993)	85
Anexo 2 – Matriz de correlações	86
Anexo 3 – Estudo da multicolinearidade	89
Anexo 4 – Valores médios das variáveis potencialmente explicativas na amostra final	90
Anexo 5 – Resultados da regressão logística com variáveis explicativas não centradas	93
Anexo 6 – Resultados da regressão logística com variáveis explicativas centradas considerando a média do grupo	96
Anexo 7 – Resultados do modelo <i>logit</i> multinível apenas com termo independente	99
Anexo 8 - Resultados do modelo <i>logit</i> multinível com as variáveis que resultaram estatisticamente significativas na regressão logística	101
Anexo 9 – Resultados do modelo <i>logit</i> multinível com a variável JCS_DAT	103



## **Índice de Tabelas**

Tabela 1 – Rácios económico-financeiros utilizados em estudos anteriores	29
Tabela 2 – Descrição da base de dados	38
Tabela 3 – Processos de insolvência	39
Tabela 4 – Disponibilidade de informação financeira das empresas com processos de insolvência interpostos em 2009	41
Tabela 5 – Análise de valores omissos	43
Tabela 6 – Diagnóstico de colinearidade das variáveis explicativas	45
Tabela 7 – Valores das estatísticas de diagnóstico de colinearidade após selecção de variáveis	46
Tabela 8 – Distribuição de empresas por sector de actividade e por dimensão	47
Tabela 9 – Distribuição de empresas por volume de negócios (em Euros)	48
Tabela 10 – Distribuição do incumprimento por sector de actividade	50
Tabela 11 – Distribuição do incumprimento por dimensão da empresa	50
Tabela 12 – Distribuição do incumprimento por sector de actividade e dimensão da empresa	51
Tabela 13 – Distribuição do incumprimento por novos escalões de volume de negócios (em Euros)	53
Tabela 14 – Distribuição dos incumprimentos por sector de actividade (agrupamentos considerados)	53
Tabela 15 – Estatísticas descritivas das variáveis potencialmente explicativas na amostra final	56
Tabela 16 – Ajustamento global do modelo (Teste de Hosmer e Lemeshow)	59
Tabela 17 – Matriz de classificação	60
Tabela 18 – Resumo dos modelos estimados (regressão logística)	61
Tabela 19 – Estatísticas de ajustamento (modelo multinível apenas com termo independente)	64
Tabela 20 – Estimativa da variância do termo independente aleatório (modelo multinível apenas com termo aleatório)	64
Tabela 21 – Efeitos fixos (modelo multinível apenas com termo independente)	65
Tabela 22 – Efeitos aleatórios (modelo multinível apenas com termo independente)	65

Tabela 23 – Resumo das soluções dos efeitos aleatórios estatisticamente significativos	67
Tabela 24 – Estimativas das variâncias dos efeitos aleatórios (modelo multinível com variável JCS_DAT)	67
Tabela 25 – Efeitos fixos (modelo multinível com com variável JCS_DAT)	68
Tabela 26 – Efeitos aleatórios (modelo multinível com com variável JCS_DAT)	68
Tabela 27 – Resumo das estimativas dos efeitos aleatórios (modelo multinível com com variável JCS_DAT)	71
Tabela 28 – Quadro-resumo das estimativas dos coeficientes das soluções obtidas	73
Tabela 29 – Resumo das probabilidades de incumprimento associadas aos sectores de actividade	74

## **Índice de Figuras**

Figura 1 – Diagrama da selecção preliminar da base de empresas a estudar	42
Figura 2 – Distribuição das micro-empresas por volume de negócios	49
Figura 3 – Distribuição das pequenas empresas por volume de negócios	49
Figura 4 – Distribuição das médias empresas por volume de negócios	49
Figura 5 – Distribuição das grandes empresas por volume de negócios	49
Figura 6 – Curva ROC	60



## Introdução

O objectivo desta tese é testar a aplicação de um modelo multinível com função de ligação *logit* ao estudo do incumprimento de empresas portuguesas, como alternativa à regressão logística “tradicional”. As variáveis explicativas são as usualmente consideradas neste tipo de problemática, nomeadamente variáveis de índole financeira, testando-se o sector de actividade como factor diferenciador da probabilidade de incumprimento.

Em termos de estrutura, no Capítulo 1 é feita uma breve revisão das abordagens que têm sido seguidas no âmbito do estudo do incumprimento das empresas e das técnicas que têm vindo a ser utilizadas na sua modelação, incluindo a referência a trabalhos recentes sobre a realidade do tecido empresarial português. No Capítulo 2 é introduzida a modelação multinível, que pode ser considerada uma inovação metodológica no estudo do incumprimento. É apresentada a especificação genérica do modelo para de seguida se apresentar a especificação de um modelo de escolha discreta de dois níveis, com função de ligação *logit*, sendo este o modelo aplicado e cujos resultados serão comparados com os resultados obtidos através da aplicação de uma regressão logística “tradicional”. Com o primeiro testar-se-á empiricamente a existência de efeitos não observados ao nível do sector de actividade que impliquem diferentes probabilidades de incumprimento das empresas. Na regressão logística, os sectores de actividade vão ser considerados como variáveis *dummy*. É feita ainda uma breve discussão em torno da questão da centragem das variáveis explicativas na modelação multinível. No Capítulo 3 apresenta-se o modelo a estimar, sendo mencionadas não apenas as variáveis a considerar como também a estimação e diagnóstico na regressão logística e no modelo multinível com função de ligação *logit*. No Capítulo 4 descrevem-se os dados utilizados e o respectivo processo de tratamento que foi necessário conduzir até se chegar a uma base de dados adequada para efeitos da análise pretendida. A aplicação empírica consta no Capítulo 5, finalizando-se com as principais conclusões e limitações deste estudo. Apresentam-se ainda sugestões para investigação futura.

## 1. O incumprimento das empresas

### 1.1. Apresentação da problemática e sua importância

O estudo da problemática do incumprimento das empresas e seus factores explicativos tem suscitado ao longo do tempo um grande interesse de investigação. Apresentam-se de seguida alguns dos factores apontados em trabalhos recentes.

Balcaen e Ooghen (2006) referem o trabalho de vários autores (Doumpos e Zopoudinis, 1999; Bickerdyke *et al.*, 1999; Amrhein e Katz, 1998) que apresentam motivos económicos para o interesse na temática do incumprimento das empresas, uma vez que a sua ocorrência gera vários tipos de custos, não apenas para os accionistas, administradores e trabalhadores, mas também no ambiente directo da empresa – investidores, clientes, fornecedores, Estado – e para a economia como um todo. Devido aos efeitos de contágio, os custos de incumprimento de uma empresa, principalmente quando existe uma rede considerável de empresas relacionadas, pode causar um efeito em espiral descendente para a economia de um país. Consequentemente, a previsão de incumprimento de empresas é importante não apenas do ponto de vista "individual", mas também ao nível da "sociedade como um todo".

Altman e Saunders (1998) apresentam, entre outros motivos para o aumento do interesse nesta temática, o aumento estrutural de falências registadas mundialmente e o declínio do valor dos activos reais (e, consequentemente, do colateral) em muitos mercados. Já em 1976, Deakin havia referido que a importância de um modelo sobre a situação financeira de uma empresa estaria muito ligada à sua utilidade para a tomada de decisão, podendo o interesse aumentar consideravelmente se o modelo de classificação puder ser utilizado para estimar a probabilidade de as empresas pertencerem a determinado grupo. Apesar de se esperar que os auditores independentes e outros decisores consigam fazer avaliações correctas sobre a saúde financeira das empresas, a investigação tem mostrado que, na prática, não têm um tão bom desempenho como os modelos de previsão.

Mais recentemente, o Novo Acordo de Basileia (Basel Committee on Banking Supervision, 2006), comumente referido como Basileia II, também é muitas vezes apontado como um dos motivos para o estudo da modelação do incumprimento, já que impôs que os cálculos de requisitos mínimos de capital dos bancos sejam mais sensíveis ao seu verdadeiro perfil de

risco, podendo os bancos ser autorizados a utilizar as suas notações internas para esse efeito de cálculo de requisitos mínimos de capital, havendo assim um incentivo para o desenvolvimento de modelos internos de avaliação de risco (Balcaen e Ooghen, 2006; Vermeulen, 2008).

Chan-Lau (2006) refere que a probabilidade de incumprimento constitui o primeiro passo na avaliação da exposição ao risco e as perdas potenciais enfrentadas por um investidor ou pelas instituições financeiras. Segundo Antunes *et al.* (2005), o cálculo da probabilidade de incumprimento das obrigações de crédito das sociedades não financeiras é um dos instrumentos de que as instituições financeiras necessitam para levar a cabo a sua actividade de concessão de empréstimos, podendo ser aplicadas na constituição de provisões para crédito mal-parado.

Chan-Lau (2006) menciona ainda a importância das probabilidades de incumprimento como um *input* fundamental para a avaliação do risco sistémico e para os exercícios de *stress testing* dos sistemas financeiros ao nível nacional, regional e global. A questão da análise da estabilidade financeira é também apresentada em outros estudos, como em Antunes *et al.* (2005) que mencionam a possibilidade de utilização de modelos de incumprimento para calcular o desempenho do sistema financeiro no caso de a economia ser sujeita a choques macroeconómicos negativos e de grande amplitude (mas verosímeis). Vermeulen (2008) afirma que o risco de crédito das empresas pode variar com alterações na rentabilidade, liquidez e solvência das empresas, as quais satisfazem as suas necessidades de financiamento nos mercados de capitais, e esse risco pode materializar-se se as empresas falharem no cumprimento das suas obrigações e no pagamento das suas dívidas. Se não for antecipado e ocorrer em termos sistémicos, um aumento do número de incumprimentos de empresas pode pôr em causa a estabilidade financeira. A modelação do risco de crédito das empresas deve, assim, constituir um elemento chave das análises de estabilidade financeira.

Balcaen e Ooghen (2006) mencionam ainda o facto de o aumento da disponibilidade de dados e de técnicas estatísticas ter vindo a oferecer, ao longo do tempo, maiores possibilidades para a investigação deste tema e a multiplicação de trabalhos sobre o incumprimento das empresas.

## 1.2. Definição de incumprimento

De acordo com Soares (2006), os modelos que têm sido desenvolvidos em vários países têm subjacente um conceito de incumprimento que não é único. O evento incumprimento pode corresponder à falência efectiva da empresa mas pode também ter outra natureza, como seja a do não pagamento de dívida, quer esta tenha a forma de obrigações ou de empréstimos bancários. De facto, Platt e Platt (2006) referem que a maior parte dos trabalhos de investigação sobre incumprimento de empresas tem-se debruçado sobre amostras de empresas que registaram falência, sendo menos comuns modelos de previsão de problemas financeiros.

O trabalho pioneiro de Beaver (1966) assenta numa definição de *failure* ou incumprimento que não se esgota nas situações de falência, sendo apresentada como a incapacidade de uma empresa cumprir as suas obrigações financeiras no momento da sua maturidade. Em termos operacionais, essa situação corresponde a falência, falta de pagamento a credores, existência de contas bancárias a descoberto ou a um não pagamento de dividendos a capital preferencial. Outros trabalhos, como o de Altman (1968) ou Ohlson (1980) recorreram à falência jurídica.

Contudo, Soares (2006) refere que, muito embora o incumprimento de responsabilidades de crédito por parte de uma empresa não evolua necessariamente para a falência desta, as situações de falência de sociedades não financeiras são, em geral, antecedidas por episódios de incumprimento que, ao persistirem e adquirirem severidade, culminam na extinção da empresa.

Mais recentemente, Basileia II estabelece que se deve considerar que determinado mutuário incorre em incumprimento (*default*) se se verificar uma das seguintes situações (ou ambas):

- i) O banco atribuir uma probabilidade reduzida à possibilidade de o devedor respeitar na íntegra as suas obrigações perante o grupo bancário, se não recorrer a medidas como a execução de eventuais garantias;
- ii) O devedor registar um atraso superior a 90 dias relativamente a uma obrigação de pagamento significativa para o grupo bancário. Quando se trate de descobertos significativos, o atraso deve começar a ser contado no momento em que o devedor tiver infringido um



limite autorizado, tiver sido notificado da fixação de um limite inferior aos seus montantes em dívida ou tiver utilizado, de forma não autorizada, montantes de crédito.

Em relação à alínea i), entre as circunstâncias que constituem indícios de uma possível situação de incumprimento apresentadas no Aviso do Banco de Portugal n.º5/2007 (Banco de Portugal, 2007), que é um dos instrumentos regulamentares que transpõem para o quadro nacional as disposições em matéria de cálculo regulamentar de requisitos de fundos próprios, neste caso para efeitos de cobertura de risco de crédito, contam-se a solicitação da declaração de falência do devedor por sua iniciativa ou por parte da instituição de crédito mútua<sup>1</sup>.

Perante a não existência de uma definição completamente fechada de incumprimento, importa, pois, definir com precisão o que está a ser estudado quando se fala em incumprimento de empresas. No presente estudo, utilizar-se-á como referência as empresas que registaram processos de insolvência, independentemente do estado do processo ou de quem o interpõe, no ano de 2009. Entende-se que a existência de acções judiciais deste tipo constitui um indício forte de uma situação de incumprimento, no sentido em que existe uma considerável probabilidade de a empresa não respeitar na íntegra as suas responsabilidades perante terceiros. Esta questão é retomada na Secção 3.1.

### **1.3. Evolução das técnicas utilizadas na modelação do incumprimento de empresas**

Chan-Lau (2006) classifica os modelos de estimação de probabilidades de incumprimento em duas categorias: *market-based models*, que se baseiam nos preços dos títulos (*security prices*) e *fundamentals-based models*, que se baseiam em dados das demonstrações financeiras, factores sistemáticos do mercado e factores económicos e informação de *ratings*. Os *fundamental-based models*, classe em que se insere o modelo utilizado no presente trabalho, são entendidos como especialmente úteis no caso da estimação de probabilidades de incumprimento associadas a empréstimos ou a empresas privadas, não cotadas, sendo habitual o recurso a técnicas econométricas relativamente simples (análise discriminante, modelo *logit*, modelo *probit*). Ainda de acordo com o autor, dentro dos *fundamental-based models* é possível encontrar: (i) modelos macroeconómicos, normalmente utilizados na

---

<sup>1</sup> Cf. ponto 46 do Aviso do Banco de Portugal n.º 5/2007 (Banco de Portugal, 2007).

estimação de probabilidades de incumprimento ao nível sectorial; (ii) modelos de *credit scoring* ou *accounting-based*, os quais geram probabilidades de incumprimento ou *ratings* para empresas com base em informação financeira; (iii) modelos baseados em *ratings*, que podem ser utilizados para inferir probabilidades de incumprimento quando existe informação de *rating*; (iv) finalmente, os modelos híbridos, os quais geram probabilidades de incumprimento utilizando como variáveis explicativas uma combinação de variáveis económicas, rácios financeiros e informação de *ratings*.

Pode-se situar o início da utilização de rácios financeiros para a previsão do incumprimento das empresas e classificação por qualidade de crédito no trabalho de Fitzpatrick, em 1932 (citado por Chan-Lau, 2006), que concluiu que a probabilidade de incumprimento estaria relacionada com as características individuais das empresas. Umhas décadas mais tarde, Beaver (1966), Altman (1968) e Merton (1974) desenvolveram estudos sobre a temática que viriam a ser considerados pioneiros, e desde então tem-se assistido a uma proliferação de trabalhos nesta área.

Beaver (1966) introduziu a versão univariada da análise discriminante para empresas públicas e cotadas em bolsa, pertencentes a sectores relacionados com actividades industriais<sup>2</sup>. No seu artigo, Beaver define incumprimento (*failure*) como a incapacidade de uma empresa cumprir as suas obrigações financeiras no momento da sua maturidade. Em termos operacionais, essa situação corresponderia a falência, *default* de obrigações, uma conta bancária a descoberto (*overdrawn bank account*) ou a um não pagamento de dividendos de acções preferenciais. À data, os rácios financeiros já eram comumente utilizados como variáveis preditivas do incumprimento, embora ao nível de sistemas periciais, por analistas, tendo sido considerado pelo autor como um ponto de partida para a verificação empírica da análise de rácios.

Balcaen e Ooghen (2006) referem que, em resposta a Beaver, Tamari (1966) constatou que a avaliação da saúde financeira de uma empresa não se deveria basear em apenas uma variável, tendo introduzido um índice de risco (*risk index*) com base na atribuição de pontos entre 0 e

---

<sup>2</sup> “Moody’s Industrial Manual contains the financial-statement data for industrial, publicly owned corporations. The population excludes firms of noncorporate form, privately held corporations, and nonindustrial firms (e.g., public utilities, transportation companies, and financial institutions). (...) Strictly speaking inferences drawn from this study apply only to firms that are members of the population.” (Beaver, 1966, pp. 72).

100, de acordo com os valores de vários rácios financeiros (quanto maior o número de pontos, melhor a situação financeira da empresa).

Em 1968, Altman introduziu a análise discriminante na forma multivariada, combinando linearmente cinco rácios financeiros referentes a empresas da indústria transformadora (*manufacturing corporations*) para produzir o que se designou por *z-score*<sup>3</sup>. Neste artigo, a definição de incumprimento correspondia a situações de falência, mais concretamente, “(...) *manufacturers that filed a bankruptcy petition under Chapter X of the National Bankruptcy Act during the period 1946-1965.*” Em 1977, Altman, Hadelman e Narayanan ajustaram o modelo *z-score* original, de modo a que reflectisse explicitamente novos desenvolvimentos relativos ao incumprimento de empresas e a incorporar melhorias relativas às técnicas de análise discriminante, criando um modelo denominado *zeta analysis* (Altman, 2000).

De acordo com Altman e Saunders (1998), até à década de 70, e até à generalização da utilização de técnicas mais objectivas, era uma prática comum das instituições procederem à avaliação do risco de crédito das posições em risco sobre empresas (empréstimos concedidos) através de *expert judgement* ou sistemas periciais, os quais, embora assentes em informação objectiva (informação financeira que reflectia várias características dos devedores, como a reputação (“*character*”), capital (alavancagem), capacidade (volatilidade dos ganhos) e colateral - os chamados 4 Cs do crédito), resultavam em análises altamente subjectivas (somente a opinião dos analistas) sobre a decisão de conceder ou não crédito. Nas décadas de 80 e 90, as instituições financeiras começam a afastar-se progressivamente dos sistemas subjectivos/periciais para se aproximarem de sistemas mais objectivos. A evolução ao nível das técnicas/métodos aplicados foi possível com o desenvolvimento da informática relacionada com a estatística. Uma vez que o principal problema consiste na identificação de padrões contrastantes nas empresas “más” e nas empresas “boas”, a escolha de técnicas de classificação ocorreu naturalmente (Lacerda e Moro, 2008). Entre as principais técnicas que têm vindo a ser utilizadas nos modelos de *credit scoring* contam-se a já referida análise discriminante, os modelos de probabilidade condicionada (modelos *logit*, *probit* e probabilístico linear), redes neuronais, árvores de decisão e *support vector machines* (SVM) (Altman e Saunders, 1998; Chan-Lau, 2006).

---

<sup>3</sup> O *z-score* de Altman é uma combinação dos seguintes rácios: *cash-flow*/activo total, lucros retidos/activo total, resultados antes de juros e impostos/activo total, capitalização de mercado/dívidas totais e vendas/activo total.

Refira-se que o modelo de Altman e as suas modificações mantiveram a sua popularidade durante muito tempo e ainda são frequentemente utilizadas na prática, embora os pressupostos em que assentam – a normalidade e a igualdade das matrizes de variância-covariância das empresas “boas” e “más” – sejam alvo de bastantes críticas (Lacerda e Moro, 2008). Até aos anos 80, a análise discriminante dominou a literatura sobre modelos de incumprimento de empresas para, a partir daí, a sua utilização e o método ser de alguma forma preterido em relação a técnicas menos exigentes em termos de pressupostos – os modelos de probabilidade condicionada, como os modelos *logit*, *probit* e probabilístico linear (Balcaen e Ooghe, 2006).

Os modelos de probabilidade condicionada permitem estimar a probabilidade de uma empresa entrar em incumprimento, condicionada a um conjunto de características das empresas, através da estimação de máxima verosimilhança. Os modelos baseiam-se em determinados pressupostos em relação à distribuição de probabilidade. Os modelos *logit* assumem uma distribuição logística, enquanto os modelos *probit* assumem uma distribuição normal (acumulada). O modelo probabilístico linear assume que a relações entre as variáveis e a probabilidade de incumprimento é linear.

Ohlson (1980) foi pioneiro na utilização de uma análise *logit* com rácios financeiros para prever o incumprimento de empresas, materializado no acontecimento de falência<sup>4</sup>, e Zmijewski (1984) foi pioneiro na aplicação de uma análise *probit*, embora a primeira se mantenha mais popular. O modelo *logit* combina várias características ou atributos num *score* de probabilidade (multivariado) para cada empresa, o qual indica a probabilidade de incumprimento ou a vulnerabilidade face ao incumprimento, o que explica o sucesso da utilização desta técnica aplicada a um estudo desta natureza (Lacerda e Moro, 2008). Na aplicação do modelo *logit* não se assumem pressupostos no que respeita à distribuição das variáveis independentes nem no que respeita às probabilidades de incumprimento *a priori* (Balcaen e Ooghe, 2006). Os pressupostos do modelo *logit* são que a variável dependente

---

<sup>4</sup> As variáveis consideradas por Ohlson (1980) foram: dimensão, medida pelo logaritmo do rácio activo total/índice de preços do PNB, passivo total/activo total, cash-flow/activo total, passivo de curto prazo/activo de curto prazo, variável *dummy* que assume o valor 1 quando o passivo total excede o activo total e 0 caso contrário, resultado líquido/activo total, fundos provenientes de operações/passivo total, variável *dummy* que assume o valor 1 quando resultado líquido é negativo nos dois anos anteriores e 0 caso contrário e, finalmente, uma variável que pretende medir as variações no resultado líquido.

seja dicotómica, com grupos discretos, sem sobreposições e identificáveis; os custos dos erros tipo I e tipo II<sup>5</sup> devem ser considerados na selecção da probabilidade de corte.

Platt e Platt (1991) utilizam o modelo *logit* para testar se os rácios financeiros dos sectores de actividade, e não apenas os rácios das empresas, predizem melhor a falência das empresas. Em geral, o modelo que considerava os rácios financeiros dos sectores de actividade teve um desempenho melhor do que o modelo não ajustado, tendo-se chegado a conclusões semelhantes no contexto de modelos de análise discriminante (Izan, 1984).

Técnicas mais complexas, nomeadamente procedimentos de aprendizagem automática, têm vindo a ser aplicadas ao problema do incumprimento de empresas, como a utilização de redes neuronais, árvores de decisão ou *support vector machines*. Contudo, a utilização de técnicas mais complexas e os aperfeiçoamentos metodológicos nem sempre foram acompanhados de uma melhoria nos resultados (Pereira *et al.*, 2007).

No que respeita às redes neuronais, a elaboração destes modelos supõe, por um lado, a dedução das características essenciais dos neurónios e suas conexões e, por outro, a implementação do modelo num computador de forma a que se possa simular um tal sistema. Os princípios que ainda hoje vigoram sobre as redes neuronais (artificiais) foram apresentados pela primeira vez por Warren McCulloch e Walter Pitts em 1943, e demonstraram que as redes neuronais artificiais podem calcular qualquer função aritmética ou lógica (Hagan *et al.*, 1996). No início da década de 90, Tam e Kiang (1992) utilizaram dados de incumprimentos de empresas para comparar o desempenho das redes neuronais com o de outros métodos, incluindo a regressão logística, tendo o método sido considerado pelos autores como promissor em termos de precisão das previsões, adaptabilidade e robustez. Entre outras vantagens desta técnica destaca-se a possibilidade de modelação não linear dos dados, existindo trabalhos que destacam evidências de existência de relações não lineares entre as variáveis. Kiang e Tam (1992) citavam o trabalho de Shepanski (1983) segundo o qual a avaliação humana (como é o caso do trabalho efectuado pelos analistas de risco) é melhor aproximada por uma função linear. Contudo, um dos inconvenientes mais apontados às redes neuronais é a dificuldade para o utilizador dos modelos resultantes entender o

---

<sup>5</sup> O erro de tipo I corresponde à rejeição da hipótese nula quando esta é verdadeira, sendo o erro de tipo II correspondente à não rejeição da hipótese nula quando esta é falsa (Greene, 2003).

raciocínio inerente, sendo estes sistemas normalmente entendidos como "caixas negras" que dão respostas, mas não transmitem conhecimento sobre o processo que conduziu à obtenção das mesmas (Pereira *et al.*, 2006). A necessidade de grandes conjuntos de dados de treino e de ciclos de aprendizagem das redes (iterações) também podem ser apontadas como inconvenientes das redes neuronais (Kumar e Ravi, 2007).

De acordo com Zhang e Härdle (2008), as técnicas de indução de regras e árvores de decisão constituem outro método de classificação não linear, além das redes neuronais, sobre o qual recaiu um grande interesse de investigadores e académicos desde que Breiman *et al.* (1984) introduziram o CART (*Classification And Regression Tree*). As árvores de decisão são uma forma de representação de um conjunto de regras que seguem uma hierarquia de classes ou valores, expressando uma lógica condicional simples. Em termos gráficos podem ser semelhantes a uma árvore, consistindo numa estrutura que interliga um conjunto de nós através de *ramos* resultantes de uma partição recursiva dos dados, desde o nó *raiz* até aos nós terminais (*folhas*), que fornecem a classificação para a instância.

A indução de regras aparece muitas vezes associada às árvores de decisão, sendo usadas para expressar o conhecimento representado nestas. Uma vez que uma árvore de decisão se encontra na forma normal disjuntiva é relativamente fácil traduzir este classificador para um conjunto de regras de decisão. A apresentação das regras é efectuada tendo em consideração a seguinte prioridade: a regra mais importante é representada na árvore como o primeiro nó, seguindo-se as restantes regras de acordo com o seu grau de relevância ou prioridade.

As árvores de decisão apresentam como vantagens a sua versatilidade e uma leitura facilmente compreensível do tipo "se-então", sendo a necessidade de grandes volumes de dados para a sua construção e a possibilidade de sobreajustamento normalmente apontados como desvantagens (Kumar e Ravi, 2007).

Os SVM são um método de classificação não linear recente, baseado na separação de duas classes de observações com a máxima margem ou *gap* entre as classes de observação (no caso em estudo, as empresas em incumprimento e as empresas sem incumprimento) (Lacerda e Moro, 2008). O SVM tem sido utilizado com muito sucesso em problemas de classificação de muitos domínios e normalmente revelam um melhor desempenho do que outras técnicas (Zhang and Härdle, 2008). Contudo, apresenta desvantagens relacionadas com uma elevada

complexidade em termos algorítmicos e a exigência em termos de memória (Kumar e Ravi, 2007).

#### **1.4. Abordagens portuguesas recentes**

Nos últimos anos tem-se assistido à produção de alguns trabalhos de autores portugueses sobre o incumprimento de empresas nacionais, alguns impulsionados pelos desenvolvimentos do Novo Acordo de Basileia. Não é objectivo desta secção proceder a uma enumeração exaustiva de todos os trabalhos que incidiram sobre o estudo da previsão do incumprimento de empresas portuguesas, mas antes perceber em que consistiram algumas dessas aplicações, com enfoque para as desenvolvidas mais recentemente. Refira-se que os trabalhos aqui mencionados serviram de base à selecção das variáveis a considerar como potencialmente explicativas do incumprimento das empresas portuguesas<sup>6</sup>, as quais serão detalhadas mais à frente.

Em trabalhos académicos é possível encontrar alguns estudos empíricos sobre a falência das empresas portuguesas. A título de exemplo, Tristão (1997) estudou a aplicabilidade dos modelos *z-score* e *zeta* de Altman à realidade nacional; Santos (2000) aplicou a análise discriminante e a regressão logística à previsão de pequenas e médias empresas do sector têxtil e do vestuário; Gamelas (2005) faz uma análise económica e de eficiência do processo de falência de empresas portuguesas assente numa análise empírica; e Barros (2008) estimou modelos *logit*, *probit* e *gombit* de previsão da falência de pequenas e médias empresas portuguesas.

Um conjunto de artigos claramente impulsionado pelos desenvolvimentos de Basileia II é apresentado seguidamente de forma mais detalhada.

Antunes *et al.* (2005) estimaram um modelo de resposta binária com especificação *probit*, tendo o evento de interesse sido definido como o incumprimento num determinado empréstimo por parte de uma empresa mutuária. O nível de estudo é o empréstimo, não a empresa, ainda que o modelo estatístico usado incorpore variáveis ao nível do empréstimo e do devedor, bem como variáveis que traduzem a envolvente macroeconómica. Ao nível do

---

<sup>6</sup> Apenas não se consideraram as variáveis do trabalho de Santos (2000) pelo facto de o objecto de estudo ser apenas o sector têxtil e do vestuário.

empréstimo, foram usadas duas variáveis dicotómicas: a primeira é um indicador de incumprimento do devedor em outros empréstimos; a segunda variável é um indicador de que o devedor do empréstimo entrou em incumprimento em mais de 50 por cento dos seus empréstimos (para além daquele sob observação) durante o trimestre corrente. A nível da empresa, são usadas variáveis categóricas para o sector de actividade e para a dimensão do crédito total da empresa. Ao nível da envolvente macroeconómica, foi utilizada a taxa de desemprego, a taxa de juro de curto prazo e o desvio do Produto Interno Bruto (PIB) relativamente à tendência e, para acomodar o possível impacto da introdução do euro em 1999, foi incluída uma variável dicotómica que assume o valor 1, antes de 1999, e o valor 0, posteriormente. Foi ainda considerado o efeito cruzado desta variável com a taxa de juro, uma vez que até 1999 as taxas de juro eram tipicamente altas num contexto de inflação elevada. São incluídos no modelo termos de interacção entre o sector de actividade e as variáveis macroeconómicas, e entre a classe de dimensão do crédito e as variáveis macroeconómicas, dado que diferentes sectores e empresas de diferentes dimensões poderão responder de forma diferenciada aos desenvolvimentos macroeconómicos. Finalmente, incluíram-se variáveis para correcção de variações sazonais.

Soares (2006) recorreu à análise discriminante para modelar um indicador de notação da qualidade de crédito de sociedades não financeiras, tendo comparado os resultados obtidos com modelos *logit* e *probit*. O evento de incumprimento foi definido como correspondendo à falência efectiva da empresa e ao não pagamento da dívida, quer esta tenha a forma de obrigações ou de empréstimos bancários. Foi estimado um modelo apenas para empresas da indústria transformadora e, para além dos rácios financeiros, os modelos estimados tiveram em conta factores transversais às observações, como sejam os associados ao ciclo económico e ao sector de actividade das empresas (incluíram-se uma medida do *output gap*<sup>7</sup> e a taxa de juro de curto prazo como variáveis independentes). A inclusão de variáveis *dummy* sectoriais foi também testada quando se consideraram as observações para todos os sectores de actividade. O contributo de algumas variáveis *dummy* para a função discriminante mostrou-se pouco significativo, embora permitisse melhorar marginalmente a percentagem de

---

<sup>7</sup> O *output gap* diz respeito à diferença entre o *output* realizado e o *output* potencial, e para o seu cálculo a autora refere que foi aplicado o filtro HP sobre dados extraídos das “Séries trimestrais para a economia portuguesa” publicadas no Boletim Económico do Verão de 2006 do Banco de Portugal, remetendo para a metodologia apresentada em Almeida e Félix (2006).



classificações correctas na amostra. Também no contexto do modelo *logit*, a inclusão de variáveis *dummy* sectoriais foi rejeitada. A dimensão da empresa, que em alguns estudos desta natureza surge como variável relevante para a determinação da probabilidade de incumprimento (contribuindo para a reduzir), neste trabalho não apresentou um contributo significativo para a função discriminante.

Bonfim (2006) baseou-se em matrizes de correlação para estabelecer a relação, a nível agregado, entre variáveis macroeconómicas e o incumprimento. Ao nível das empresas, foram utilizados modelos de escolha discreta (*logit* e *probit*) para melhor entender o que explica o incumprimento das empresas ao nível dos empréstimos. Depois, complementou-se a análise com modelos de duração, que permitem perceber o tempo que demora um empréstimo a entrar em incumprimento e não apenas considerar se ocorre ou não um incumprimento, que neste estudo é definido como o montante (capital e juros) vencidos nos últimos 3 a 6 meses. A autora refere que a desvantagem desta definição é a impossibilidade de avaliar separadamente a evolução do crédito vencido das empresas e das famílias. Os resultados obtidos sugerem que as probabilidades de incumprimento são influenciadas por várias características específicas das empresas, tais como a sua estrutura financeira, rentabilidade e liquidez, bem como pelo seu desempenho em termos de vendas ou a sua política de investimento. Quando se consideram variáveis relativas ao tempo (*time-effect controls*) ou variáveis macroeconómicas em conjunto com as características específicas das empresas, os resultados parecem melhorar substancialmente. Assim, apesar de a situação financeira e operacional das empresas ter um papel central na explicação das probabilidades de incumprimento ao nível micro, as condições macroeconómicas são também muito importantes na avaliação das probabilidades de incumprimento ao longo do tempo. A autora refere que, ainda que as variáveis idiossincráticas das empresas pareçam ter um importante papel na previsão do incumprimento, elas devem ser vistas como dependendo da dimensão da empresa e do sector de actividade em que se inserem, dado que algumas variáveis podem ser mais ou menos importantes dependendo do tipo de empresas. Por este motivo, num segundo modelo, foram consideradas variáveis *dummy* relativas ao sector de actividade, tendo os resultados sugerido a existência de algumas diferenças nas variáveis que influenciam o risco de crédito entre os vários sectores de actividade (com os coeficientes associados aos rácios financeiros das empresas a manterem-se robustos).

Lacerda e Moro (2008) utilizaram a análise discriminante, o modelo *logit* e os SVM na análise das variáveis que permitem prever o incumprimento das empresas portuguesas, aqui entendido na acepção de Basileia II, pelo que uma empresa é considerada como estando em incumprimento num dado ano se, nesse ano, a Central de Responsabilidades de Crédito registar a existência de crédito vencido por mais de 3 meses. Os rácios financeiros que caracterizam a estrutura de custos, a liquidez, a actividade e a alavancagem parecem relevantes em qualquer das metodologias utilizadas, bem como a dimensão da empresa. Outras variáveis, como resultados antes de custos financeiros/custos financeiros (*earnings before interest over interest paid*) apenas parecem relevantes quando se aplica a técnica SVM, o que pode ser explicado pela sua relação não linear com o incumprimento, a qual pode ser captada pelo SVM mas não pelo modelo *logit* ou pela análise discriminante. A inclusão destas variáveis que não têm uma relação linear com o incumprimento explicam o melhor desempenho do SVM, medido por rácios de precisão superiores. Lacerda e Moro (2008) analisam também os perfis dos empregados e a concentração da dívida, confirmando a sua significância. Por um lado, a existência de relações com muitos bancos é um sinal de uma situação financeira mais débil, sendo esta variável fortemente explicativa. Também, a duração média do emprego afigura-se significativa para a explicação do incumprimento das empresas portuguesas.

## 2. Modelos multinível

### 2.1. Especificação geral da modelação multinível

A abordagem multinível aplica-se a uma população com uma estrutura hierárquica (Hox, 2002). Uma amostra de uma população com essa característica pode ser descrita como uma amostra estratificada: primeiro, considera-se uma amostra de unidades de um nível superior (e.g., escolas), para depois se retirarem subunidades dessas unidades amostrais (e.g., amostra de alunos das escolas). Nestas amostras, as observações individuais não são, geralmente, independentes, uma vez que os indivíduos interagem com os contextos sociais aos quais pertencem, *i.e.*, os indivíduos são influenciados pelos grupos sociais em que se integram, e as propriedades desses grupos são, por sua vez, influenciadas pelos indivíduos que formam o grupo. Desta forma, não se verificará o pressuposto de independência entre as observações normalmente assumido no modelo de regressão linear múltipla tradicional (Hox, 2002). De acordo com Guo e Zhao (2000), quando a estrutura hierárquica existente nos dados é ignorada e o pressuposto de independência é violado, os modelos lineares “tradicionais” tendem a subestimar os erros-padrão, o que resultará em intervalos de confiança e ensaios de hipóteses incorrectos. Segundo os mesmos autores, a estimação correcta de erros-padrão não é apenas um aspecto técnico, podendo comprometer uma conclusão importante. Os autores referem o exemplo de um estudo de Bennet (1976), que utilizou técnicas de regressão linear sem considerar o facto de estudantes estarem agrupados em professores e turmas para demonstrar que em escolas primárias do Reino Unido os estudantes beneficiavam mais de um estilo de ensino mais formal. Os resultados foram amplamente divulgados e considerados até que Aitkin *et al.* (1981) demonstraram que, uma vez considerados os agrupamentos de estudantes num modelo multinível, os resultados obtidos por Bennett relativos aos estilos de ensino já não eram estatisticamente significativos.

Assim, assume particular relevância proceder a análises multinível que permitam testar empiricamente a existência de estruturas hierárquicas em conjuntos de dados relativamente aos quais se supõe que tal acontece, bem como analisar a interacção entre as variáveis que descrevem indivíduos e as variáveis que descrevem níveis superiores de análise (*i.e.*, agrupamentos de indivíduos).

Ainda que o âmbito do presente trabalho seja a aplicação de um modelo multinível com uma função de ligação *logit* ao problema do incumprimento de empresas portuguesas, será apresentada a especificação do modelo multinível aplicável a dados lineares, para depois se apresentar a especificação do modelo *logit* multinível que será utilizado. A explicação dos vários modelos será feita com recurso ao exemplo de aplicação do presente estudo (empresas no nível inferior e sectores de actividade no nível superior).

Singer (1998) afirma que entre as principais questões a apreender em torno da modelação multinível se encontram: i) a diferença entre efeitos fixos e efeitos aleatórios; ii) a noção de níveis múltiplos numa hierarquia; iii) a noção de que a matriz de variâncias-covariâncias pode assumir várias estruturas; e iv) a centragem das variáveis explicativas. Estes aspectos vão ser abordados nas secções seguintes.

## 2.2. O modelo de regressão linear de dois níveis com uma variável explicativa em cada nível

O modelo de regressão multinível<sup>8</sup> assume que existe uma estrutura hierárquica nos dados, com a variável dependente, que se assume contínua, a ser medida no nível mais baixo, podendo existir variáveis explicativas em todos os níveis (assumindo-se erros normalmente distribuídos). Considerando que se recolhe informação sobre  $J$  sectores de actividade, cada um com  $n_j$  empresas, e tomando como exemplo o caso mais simples em que apenas existe uma variável explicativa em cada nível (um determinado rácio financeiro ao nível da empresa –  $X$  - e o volume de negócios médio do sector de actividade -  $W$ ), estabelece-se a seguinte equação de regressão para prever a variável explicativa  $Y$  - o incumprimento das empresas<sup>9</sup>:

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + e_{ij} \quad (1)$$

<sup>8</sup> Apesar de ser possível representar qualquer número de níveis, as principais características estatísticas de um problema são detectadas num modelo de dois níveis (Raudenbush e Bryk, 2002: 7) e que, como tal, será a “profundidade” adoptada no presente estudo.

<sup>9</sup> Para a apresentação do modelo de regressão multinível, abstrai-se temporariamente do facto de a variável dependente ser uma variável binária que se reflecte, em última instância, numa proporção. Também não se irá considerar a existência de variáveis explicativas no nível superior de análise, *i.e.*, ao nível dos sectores de actividade, embora para efeitos de um entendimento global da modelação multinível tal seja considerado, pelo que se apresenta o volume de negócios médio do sector como exemplo de uma variável explicativa a esse nível.

Como em qualquer equação de regressão,  $\beta_{0j}$  é o termo independente,  $\beta_{1j}$  é o coeficiente associado à variável explicativa  $X$  (ou o declive da regressão) e  $e_{ij}$  é o termo correspondente aos erros residuais. O índice  $j$  refere-se aos sectores de actividade ( $j = 1, \dots, J$ ) e o índice  $i$  refere-se às empresas ( $i = 1, \dots, n_j$ ). A diferença em relação ao modelo normal de regressão é que cada sector de actividade é caracterizado por um termo independente distinto  $\beta_{0j}$  e por um coeficiente associado à variável explicativa  $\beta_{1j}$  também distinto. Os coeficientes  $\beta_j$ , que sendo aleatórios se espera que assumam valores distintos consoante o sector de actividade  $j$ , terão uma distribuição com determinada média e determinada variância, procedendo-se à análise da variação desses coeficientes através da introdução de variáveis explicativas ao nível do sector de actividade<sup>10</sup>:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}W_j + u_{0j} \quad (2)$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}W_j + u_{1j} \quad (3)$$

Nos modelos multinível assume-se que a variância dos erros é única, independentemente dos sectores de actividade, sendo especificada como  $\sigma^2$ .

A equação 2 reflecte o facto de o nível “geral” de incumprimento no sector de actividade  $j$  ( $\beta_{0j}$ ) poder ser previsto pelo volume de negócios médio desse sector de actividade ( $W_j$ )<sup>11</sup>. A equação 3 indica que a relação (expressa por  $\beta_{1j}$ ) entre o incumprimento e o rácio financeiro de cada empresa dependerá do volume de negócios médio do sector de actividade em que se insere (neste caso, do efeito “dimensão” do ambiente em que se insere). É como se o volume de negócios médio do sector de actividade actuasse como uma variável moderadora entre o incumprimento e o valor do rácio financeiro.

Os termos  $u_{0j}$  e  $u_{1j}$  das equações (2) e (3) são termos de erros residuais aleatórios, ao nível do sector de actividade. Assume-se que estes erros  $u_j$  têm média zero e que são

<sup>10</sup> Caso não se introduzam variáveis explicativas ao nível do sector de actividade, os coeficientes são apenas função dos termos independentes  $\gamma_0$  e dos erros residuais  $u_j$ .

<sup>11</sup> Lembra-se que, apesar de se explicar o modelo recorrendo ao exemplo do volume de negócios médio do sector de actividade como variável explicativa de nível superior, na aplicação não se vai avançar com uma variável explicativa a este nível, pretendendo-se apenas perceber a existência, ou não, de uma estrutura hierárquica nos dados (captar a possibilidade de existência um efeito ou de uma variável “latente”).

independentes dos erros residuais  $e_{ij}$  que estão no nível das empresas. A variância dos erros  $u_{0j}$  é especificada por  $\sigma_{00}^2$  e a variância dos erros  $u_{1j}$  é especificada como  $\sigma_{11}^2$ . Geralmente, assume-se que a covariância  $\sigma_{12}$  entre  $u_{0j}$  e  $u_{1j}$  é diferente de zero.

É de notar que nas equações (2) e (3) se assume que os coeficientes de regressão  $\gamma$  não variam entre sectores de actividade, pelo que são referidos como coeficientes fixos. Assume-se que toda a variação entre sectores de actividade não captada na estimação dos coeficientes  $\beta$  a partir da variável explicativa  $W$  ao nível do sector de actividade ( $W_j$ ) é captada pelos termos de erros residuais  $u_j$ , que possuem índice  $j$  para assinalar que dizem respeito ao  $j$ -ésimo sector de actividade. Substituindo as equações (2) e (3) na equação (1), resulta a seguinte equação:

$$Y_{ij} = \gamma_{00} + \gamma_{10}X_{ij} + \gamma_{10}W_j + \gamma_{11}W_jX_{ij} + u_{1j}X_{ij} + u_{0j} + e_{ij} \quad (4)$$

O segmento  $\gamma_{00} + \gamma_{10}X_{ij} + \gamma_{10}W_j + \gamma_{11}W_jX_{ij}$ , que contém os coeficientes fixos, é a chamada parte fixa ou determinística do modelo. O segmento  $u_{1j}X_{ij} + u_{0j} + e_{ij}$  contém os termos aleatórios, pelo que é denominado parte aleatória ou estocástica do modelo. O efeito moderador da variável explicativa do nível superior ( $W_j$ ) na relação entre a variável dependente ( $Y_{ij}$ ) e a variável explicativa do nível mais baixo ( $X_{ij}$ ) é expressa como uma interacção entre níveis ( $W_jX_{ij}$ ). É de notar que o termo de erro  $u_{1j}$  aparece multiplicado pela variável explicativa  $X_{ij}$ , pelo que o erro total será diferente para valores diferentes de  $X_{ij}$ , o que corresponde à situação que na regressão múltipla ordinária é conhecida por heterocedasticidade.

A falta de independência entre observações num mesmo grupo, neste caso, de empresas pertencentes a um mesmo sector de actividade, pode ser expressa num coeficiente de correlação intra-classe,  $\rho$ , o qual pode ser estimado por via do modelo de regressão multinível sem quaisquer variáveis explicativas (apenas com termo independente):

$$Y_{ij} = \beta_{0j} + e_{ij} \quad (5)$$

Por sua vez, sem variáveis explicativas, a equação (2) reduz-se a:

$$\beta_{0j} = \gamma_{00} + u_{0j}. \quad (6)$$

Substituindo a equação (6) na equação (5), obtém-se:

$$Y_{ij} = \gamma_{00} + u_{0j} + e_{ij}. \quad (7)$$

A equação (7) permite fazer a decomposição da variância total da variável dependente em duas componentes independentes, associadas a cada nível:  $\sigma^2$ , que corresponde à variância dos erros no nível inferior  $e_{ij}$ , e  $\sigma_{00}^2$ , que corresponde à variância dos erros do nível superior  $u_{0j}$  (conforme já referido atrás, assume-se que os erros  $u_j$  do nível superior – sectores de actividade – são independentes dos erros residuais  $e_{ij}$  que estão no nível inferior – empresas). A correlação intra-classe  $\rho$  é, então, estimada como o peso da variância dos erros do nível superior na variância total, ou seja, corresponde a uma estimativa, na população, da variância que é explicada pela estrutura de grupo:

$$\rho = \sigma_{00}^2 / (\sigma_{00}^2 + \sigma^2). \quad (8)$$

### 2.3. O modelo de regressão de dois níveis com várias variáveis explicativas nos dois níveis

O modelo atrás descrito corresponde a uma situação simples, em que apenas existe uma variável explicativa no nível inferior (empresas) e uma variável explicativa no nível superior (sector de actividade). Generalizando o modelo para  $P$  variáveis explicativas no nível inferior e  $Q$  variáveis explicativas no nível superior, obtém-se a seguinte expressão.

$$Y_{ij} = \gamma_{00} + \sum_p \gamma_{p0} X_{pij} + \sum_q \gamma_{q0} W_{qj} + \sum_p \sum_q \gamma_{pq} W_{qj} X_{pij} + \sum_p u_{pj} X_{pij} + u_{0j} + e_{ij}. \quad (9)$$

Recorde-se que se assume que os erros do nível inferior  $e_{ij}$  (empresa) seguem uma distribuição normal com média zero e uma variância comum ( $\sigma^2$ ). Os termos  $u_{0j}$  e  $u_{pj}$  são os termos de erro no nível do sector de actividade, que se assume serem independentes dos erros  $e_{ij}$ , observados ao nível da empresas, e segurem uma distribuição normal multivariada com média zero. A variância dos erros  $u_{0j}$  é a variância dos termos independentes entre os

sectores de actividade, sendo especificada como  $\sigma_{00}$ . As variâncias dos erros  $u_{pj}$  são as variâncias dos coeficientes das variáveis explicativas entre sectores de actividade e são especificadas como  $\sigma_{pp}^2$ . Assume-se geralmente que as covariâncias entre os termos de erros residuais,  $\sigma_{pp}$ , são diferentes de zero, e são obtidas na matriz de variância-covariância de nível superior  $\Sigma$ .

Em suma, os parâmetros a estimar num modelo de regressão de dois níveis, com  $P$  variáveis explicativas no nível inferior e  $Q$  variáveis explicativas no nível superior<sup>12</sup> são os seguintes:

- 1 termo independente ( $\gamma_{00}$ );
- 1 variância dos erros ( $e_{ij}$ ) do nível inferior ( $\sigma^2$ );
- $P$  coeficientes das variáveis explicativas do nível inferior ( $\gamma_{p0}$ );
- $P$  variâncias ( $\sigma_{pp}^2$ ) dos erros ( $u_{pj}$ ) do nível superior destes coeficientes;
- $P$  covariâncias de nível superior entre o termo independente e todos os coeficientes das variáveis explicativas ( $\sigma_{0p}^2$ );
- $P(P-1)/2$  covariâncias de nível superior entre todos os coeficientes das variáveis explicativas do nível inferior;
- $Q$  coeficientes das variáveis explicativas do nível superior ( $\gamma_{q0}$ );
- $PQ$  coeficientes para as interacções *cross-level* ( $\gamma_{pq}$ ).

## 2.4. Modelos multinível para dados de resposta binária

Até agora, os modelos apresentados assentam no pressuposto de uma variável dependente contínua e erros normalmente distribuídos. Porém, e como referem Guo e Zhao (2000), existe um considerável interesse, por parte dos investigadores, no estudo de problemas caracterizados por dados com uma estrutura hierárquica e com variáveis dependentes dicotómicas, como no caso do presente trabalho em que a variável dependente é a ocorrência

<sup>12</sup> Ainda que não se considerem variáveis explicativas ao nível superior, ou seja, do sector de actividade económica, apresenta-se aqui a forma mais geral do modelo.



de incumprimento de empresas. Na aplicação feita, existe uma variável dependente dicotômica – a ocorrência ou ausência de incumprimento – que, em última instância, se pretende traduzir numa probabilidade de incumprimento. Tentar-se-á verificar empiricamente a existência de uma estrutura hierárquica nos dados ou, em última análise, verificar se os sectores de actividade em que as empresas se inserem são significativos no entendimento do fenómeno do incumprimento.

Com o tipo de variável dependente em questão (variável de resposta discreta, binária, neste caso) não se verifica a existência de uma relação linear entre essa variável e as variáveis explicativas, assim como não se verificará o pressuposto de homocedasticidade dos erros (do nível inferior) (Hox, 2002). A abordagem corrente consiste em assumir uma distribuição para a variável dependente e modelar o valor esperado dessa mesma variável, que não é necessariamente a distribuição normal. Esta classe de modelos estatísticos denomina-se modelos lineares generalizados (GLM, correspondente à denominação em língua inglesa *Generalized Linear Models*). Os modelos GLM são, então, definidos por três componentes: 1) a componente linear; 2) a distribuição da variável dependente; 3) a função de ligação que é a transformação que relaciona os valores estimados da variável dependente com os valores observados das variáveis explicativas. A função de ligação comumente aplicada para variáveis dependentes dicotômicas é a função logística ( $\text{logit}(p) = \ln(p/(1-p))$ ). No problema em estudo, irá aplicar-se esta função de ligação e define-se a variável dependente como seguindo uma distribuição Bernoulli. De seguida, apresenta-se a especificação do modelo com uma variável dependente dicotômica, começando-se com o caso mais simples de apenas uma variável explicativa no nível inferior (*i.e.*, ao nível das empresas, continuando com o exemplo relativo à aplicação feita no presente trabalho), seguindo a especificação apresentada por Guo e Zhao (2000). Assim,  $Y_{ij}$  é uma variável de resposta binária para a empresa  $i$  do sector de actividade  $j$ , sendo  $X_{ij}$  a variável explicativa ao nível da empresa, e define-se a probabilidade de incumprimento como a probabilidade de a variável  $Y_{ij}$  assumir o valor 1, *i.e.*,  $p_{ij} = P(Y_{ij} = 1)$ , em que a probabilidade de sucesso ( $p_{ij}$ ) é modelada utilizando a função de ligação *logit*. Assume-se ainda que  $Y_{ij}$  segue uma distribuição *Bernoulli*. O modelo com dois níveis e uma variável independente, já na versão reduzida a uma equação, pode ser assim descrito:

$$\text{logit}(p_{ij}) = \gamma_{00} + \gamma_{10}X_{ij} + u_{ij}X_{ij} + u_{0j} \quad (10)$$

Nesta fase, vale a pena fazer duas observações. Por um lado, e uma vez que  $Y_{ij}$  segue uma distribuição *Bernoulli* (caso particular da distribuição Binomial, quando  $n=1$ ),  $\text{logit}(p_{ij})$  terá uma distribuição aproximadamente normal, pelo que, no nível inferior, se utiliza uma componente linear para prever  $\text{logit}(p_{ij})$  (Hox, 2002). Por outro lado, o termo  $e_{ij}$  não aparece na equação (10). No caso de uma variável com distribuição Binomial (ou Bernoulli, o caso particular quando  $n = 1n$ ), a variância depende apenas da probabilidade de sucesso na população: se  $X \sim B(n, p)$  então  $E[X] = np$  e  $\text{Var}[X] = np(1-p)$ , com o caso particular de  $n=1$  a representar uma variável Bernoulli. Consequentemente, na equação (10) a variância do nível das empresas (nível inferior) é determinada pelo valor previsto de  $p_{ij}$ , pelo que não entra no modelo como um termo separado (Hox, 2002).

Em termos práticos, no PROC Glimmix do SAS, que será utilizado na estimação do modelo<sup>13</sup>, a variância é expressa como  $\phi p_{ij}(1-p_{ij})$ . Se, aquando da escolha da distribuição Binomial, a estimativa de  $\phi$  for próxima de 1, então a distribuição dos dados adere, de facto, à distribuição escolhida, no sentido em que a variância observada coincide com a variância esperada de uma distribuição binomial. Contudo, se  $\phi$  for substancialmente diferente de 1, ou seja, se houver dispersão (diferente da assumida sob o modelo binomial), então esse facto deve ser considerado e deve ser adicionado<sup>14</sup> um termo de erro extra na equação (10).

Feita esta introdução simplificada, apresenta-se a especificação do modelo adequado à aplicação apresentada neste trabalho, *i.e.*, com  $P$  variáveis explicativas no nível das empresas (nível inferior) mas sem variáveis explicativas no nível dos sectores de actividade económica (nível superior), já que o que se pretende testar é a existência de um padrão hierárquico nos dados, sem se avançar com efeitos de interacção entre níveis (apenas se pretende detectar a existência de eventuais efeitos latentes). Assim, incluindo várias variáveis explicativas ao nível das empresas, tem-se:

$$\text{logit}(p_{ij}) = \beta_{0j} + \sum_p \beta_{pj} X_{pij} \quad (11)$$

<sup>13</sup> Este procedimento será abordado de forma mais detalhada na Secção 3.3.

<sup>14</sup> No PROC Glimmix, esse termo será adicionado através do comando `random _residual_;`.

Espera-se que os coeficientes das variáveis explicativas variem entre sectores de actividade, sendo esses coeficientes modelados da seguinte forma (não considerando a existência de variáveis explicativas nesse nível):

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad (12)$$

$$\beta_{pj} = \gamma_{p0} + u_{pj} \quad (13)$$

Substituindo as expressões (12) e (13) na expressão (11), obtém-se:

$$\text{logit}(p_{ij}) = \gamma_{00} + \sum_p \gamma_{p0} X_{pij} + u_{0j} + \sum_p u_{pj} X_{pij} \quad (14)$$

A interpretação dos parâmetros da regressão da equação (14) deve ser feita não em termos das probabilidades de incumprimento, mas em termos da variável definida pela função de transformação *logit*. Os parâmetros a estimar serão, então, os seguintes:

- 1 termo independente ( $\gamma_{00}$ );
- $P$  coeficientes das variáveis explicativas do nível inferior ( $\gamma_{p0}$ );
- $P$  variâncias ( $\hat{\sigma}_{pp}^2$ ) dos erros ( $u_{pj}$ ) do nível superior destes coeficientes;
- $P$  covariâncias de nível superior entre o termo independente e todos os coeficientes das variáveis explicativas;
- $P(P-1)/2$  covariâncias de nível superior entre todos os coeficientes das variáveis explicativas.

## 2.5. A centragem das variáveis explicativas

Ainda que a questão da centragem dos dados não seja específica dos modelos multinível, quando aplicada a este caso ganha uma outra complexidade, uma vez que o seu efeito depende da relação entre os dados e do objectivo do estudo (Kreft e De Leeuw, 1998). Por centragem entende-se retirar o mesmo valor a cada variável explicativa. No caso da regressão “clássica”, esse valor será a média da variável; na presença de dois níveis, esse valor poderá ser ou a média global da variável ou a média da variável no grupo (no sector de actividade, no presente caso).

Na regressão “clássica”, a centragem não altera as relações nos dados, uma vez que simples transformações aditivas nos dados não afectam a sua variância nem covariâncias nem correlações com outras variáveis (Aiken e West, 1991). Se as variáveis explicativas forem centradas, o valor do termo independente corresponde ao valor esperado da variável dependente quando todas as variáveis explicativas assumem o seu valor médio; se não forem centradas, corresponde ao caso em que as variáveis explicativas assumem o valor zero. Dito de outra forma, o termo independente corresponde a um parâmetro “livre” que permite garantir a invariância do modelo a alterações na origem das variáveis explicativas. Refira-se que a centragem pode ser um auxiliar importante para o problema da multicolinearidade das variáveis explicativas.

No caso dos modelos multinível, o efeito de centrar as variáveis explicativas considerando a sua média global é mais fácil de perceber do que o efeito de centrar as variáveis considerando a sua média no grupo.

No primeiro caso, o modelo considerando as variáveis originais é estatisticamente equivalente ao modelo com as variáveis centradas, uma vez que se está a subtrair um valor constante a cada variável. Contudo, é de referir que a equivalência estatística entre modelos não significa que as estimativas dos parâmetros sejam idênticas, mas sim que o ajustamento e as previsões o são (e as estimativas dos parâmetros podem ser facilmente transformadas umas nas outras) (Kreft e De Leeuw, 1998).

No segundo caso, é muito provável que as médias sejam diferentes entre grupos e a equivalência entre modelos (centrado e não centrado) não se verifica, situação que, de acordo com Kreft e De Leeuw (1998), tem apenas duas excepções. Uma delas corresponde à situação em que a média nos grupos coincide com a média global, o que acontece em casos de análise de medidas repetidas, mas é uma situação rara quando se procura investigar diferenças entre grupos, como é o caso. A outra excepção corresponde à situação em que se considera apenas um termo independente aleatório e coeficientes fixos associados às variáveis explicativas e em que as médias dos grupos são introduzidas como variáveis explicativas de segundo nível. Neste caso, isso equivale a estimar um modelo com as variáveis explicativas não centradas apenas com um termo independente aleatório, mas sem a média do grupo como variável de segundo nível.

A centragem pode ser interessante em termos técnicos, porque remove correlações elevadas entre o termo independente e os coeficientes aleatórios, assim como correlações entre variáveis de primeiro e segundo níveis e entre interações de níveis. Porém, centrar um modelo pode corresponder a ajustar um modelo diferente do original e a opção de centrar as variáveis não deve assentar somente nestes aspectos técnicos.

A não consideração de variáveis centradas é mais adequada quando existe interesse num modelo que explique o máximo da variação da variável dependente, sem nenhum interesse particular em variáveis de segundo nível, ou caso se esteja mais interessado nos resultados ao nível individual (Kreft e De Leeuw, 1998).

A centragem das variáveis explicativas considerando a média das variáveis no grupo implica uma modelação menos eficaz do que utilizando as variáveis originais, simplesmente porque toda a informação sobre diferenças entre grupos é removida dessa variável. Assim, parece razoável voltar a considerar essa informação introduzindo a média no grupo como uma variável explicativa de segundo nível, mas este procedimento introduz informação adicional sobre a estrutura dos grupos que não está presente quando se consideram as variáveis não centradas, pelo que se obtém assim um modelo mais ajustado do que o obtido não centrando as variáveis. Porém, o procedimento de centragem utilizando a média do grupo aparenta ser uma forma implícita e complicada de alterar o significado dos efeitos individuais e o nível do grupo, incluindo a interpretação das interações entre níveis (Hox, 2002).

Perante a complexidade das questões em torno da centragem das variáveis explicativas, as opções a tomar serão as seguintes:

- na regressão logística, a centragem considerando a média no grupo será utilizada como auxiliar do tratamento da multicolinearidade, verificando-se se as diferenças em relação ao modelo sem consideração das variáveis centradas residem na estimativa do termo independente e dos coeficientes associados às categorias de sectores de actividade, considerada como variáveis *dummy*, que correspondem a “correções” ao termo independente, e que terão consequências nas conclusões dos ensaios de hipóteses sobre a significância de alguns sectores de actividade;
- no modelo multinível, não se procederá a qualquer centragem de variáveis, e a justificação assenta no facto de o objectivo do presente estudo se prender com os resultados ao nível

individual, ou seja, com o incumprimento das empresas, sem que haja um interesse particular em variáveis de segundo nível. Apenas se pretende testar a existência de partilha de características não observadas entre empresas pertencentes ao mesmo sector de actividade.

### 3. Especificação e estimação do modelo

#### 3.1. Seleção das variáveis explicativas e a consideração do sector de actividade

Ainda que a problemática do incumprimento das empresas se tenha mantido como um tópico de interesse ao longo dos anos, não existe propriamente uma formulação teórica desta problemática.

Platt e Platt (1991) referiram que a decisão de incluir determinada variável na especificação de um modelo deve basear-se, por princípio, em considerações teóricas. Contudo, no caso dos modelos de falência, não existe uma teoria genericamente aceite que possa ser utilizada para efeitos de especificação, que se torna, desta forma, um assunto de carácter empírico. Também Greene *et al.* (2007) apontam aos modelos mais simples de incumprimento de empresas a falta de algum conteúdo teórico ou explicativo, uma vez que esses modelos são muitas vezes desenvolvidos com base em processos de iteração estatística a partir de um grande conjunto de rácios financeiros, até que se consiga obter um modelo estatisticamente significativo e com a maior precisão possível. Os autores afirmam que, em muitos casos, o papel explicativo e teórico das variáveis explicativas acaba por ficar em segundo plano.

Não é o objectivo principal desta tese propor uma formulação teórica de um modelo de incumprimento, continuando-se assumidamente na linha da exploração empírica da informação disponível, principalmente de carácter económico-financeiro. A este propósito, Altman (2000) refere que, ainda que, em geral, rácios que medem as dimensões de rentabilidade, liquidez e solvência das empresas tenham prevalecido como os indicadores mais relevantes, a sua ordem de importância não é clara, uma vez que cada estudo parece apontar um rácio diferente como o mais importante para a previsão do incumprimento.

Por outro lado, Hensher e Jones (2004) afirmam que não se tem verificado a aplicação, a esta área de estudo, de desenvolvimentos metodológicos ao nível dos modelos de escolha discreta que têm emergido em outros campos das ciências sociais. A aplicação de modelos *logit* simples (técnica que mantém a sua popularidade) resulta em parâmetros fixos e em estruturas de erros tratadas como “ruído branco” (erros não correlacionados), apesar de a complexidade do problema do incumprimento merecer uma modelação mais robusta e mais apelativa em termos comportamentais, que reconheça a possibilidade de existência de heterogeneidade

observada e não observada entre empresas (Hensher e Jones, 2007). Com vista a colmatar essa insuficiência, alguns artigos onde se aplicam modelos avançados de escolha discreta a esta problemática “clássica” têm sido publicados recentemente (*e.g.*, Hensher e Jones (2004), Barros *et al.* (2007), Greene *et al.* (2007), Hensher e Jones (2007), Hossari (2009)).

No presente estudo irá aplicar-se um modelo multinível de escolha discreta, com função de ligação *logit*, com vista a testar empiricamente a existência de efeitos não observados ao nível do sector de actividade que impliquem diferentes probabilidades de incumprimento de empresas portuguesas. Os resultados assim obtidos serão comparados com os resultados obtidos com um modelo *logit* simples, no qual os sectores de actividade vão ser considerados como variáveis *dummy*.

A consideração explícita do sector de actividade como factor diferenciador da probabilidade de incumprimento das empresas aparece, de alguma forma, na linha do mencionado por alguns autores. Falk e Heintz (1975) referiram que, à data, existia já algum sucesso na identificação da relação entre rácios financeiros das empresas e o seu risco, mas que havia espaço e interesse crescente na consideração dos elementos relativos à pertença a determinado sector de actividade no risco das empresas. Essa ideia levou-os a construir um *ranking* de sectores de actividades de acordo com o seu grau de risco, avaliado com base em rácios financeiros aplicados aos sectores.

Platt e Platt (1991) referiram que, ainda que uma característica comum a vários estudos sobre incumprimento seja a utilização de rácios económico-financeiros como variáveis explicativas que permitem ter em consideração diferenças de dimensão entre empresas, por serem medidas relativas, na verdade os valores dos rácios são distintos consoante os sectores de actividade. E, sendo assim, então os modelos de previsão de incumprimento ou devem considerar ajustamentos aos rácios que tomem esse aspecto em consideração ou só poderão ser comparados se disserem respeito ao mesmo sector de actividade.

Hossari (2009) identifica três abordagens no que respeita à forma de considerar o sector de actividade das empresas presentes na literatura. Uma delas é simplesmente contornar a questão, uma vez que os investigadores limitam as suas amostras a empresas pertencentes a um único sector de actividade (o que faz com os resultados obtidos não possam ser extrapolados para empresas pertencentes aos sectores não considerados). Em outra abordagem, as amostras são constituídas por empresas pertencentes a vários sectores de



actividade, embora os investigadores não ajustem as variáveis explicativas (normalmente, rácios económico-financeiros) a potenciais efeitos causados pelos sectores de actividade. Finalmente, numa última abordagem, os investigadores procuram ajustar as variáveis explicativas ao sector de actividade em que as empresas se inserem (considerando, por exemplo, variáveis *dummy* que representam os distintos sectores considerados), sendo nesta última abordagem que se pode inserir o presente trabalho.

Para seleccionar as variáveis com as quais se vai trabalhar, procedeu-se à análise de variáveis utilizadas em vários trabalhos portugueses que incidiram sobre o incumprimento de empresas<sup>15</sup>, o que resultou numa identificação de 43 indicadores/rátios económico-financeiros potencialmente relevantes, aos quais se acrescentará informação sobre a antiguidade das empresas<sup>16</sup> (em meses). Os rácios, apresentados na Tabela 1, podem ser classificados de acordo com a dimensão que estão a medir: actividade, alavancagem e risco, estrutura de custos, estrutura<sup>17</sup>, gestão de tesouraria, liquidez, rentabilidade e solvência. A Tabela 1 resume os indicadores identificados e as dimensões de análise utilizadas em estudos anteriores.

Em relação aos sectores de actividade, o processo que presidiu à selecção dos 23 agrupamentos a considerar é explicitado no Capítulo 4, mais especificamente na Secção 4.3.

**Tabela 1 – Rátios económico-financeiros utilizados em estudos anteriores**

	Dimensão	Variável	Tristão (1997)	Gamelas (2005)	Soares (2006)	Bonfim (2006)	Barros (2008)	Lacerda e Moro (2008)
1	Actividade	Activo total/Volume de negócios (AT/VN)						X
2	Actividade	Volume de negócios/Activo total (VN_AT)	X				X	X
3	Alavancagem e risco	Capital próprio/Activo total (CP_AT)			X	X	X	X
4	Alavancagem e risco	Dívida a terceiros excl. empresas do grupo associados e sócios/Activo			X			

<sup>15</sup> Nomeadamente, Barros (2008), Bonfim (2006), Gamelas (2005), Lacerda e Moro (2008), Soares (2006) e Tristão (1997), mencionados na Secção 1.4. Os autores destes artigos já seleccionaram os indicadores económico-financeiros a partir de revisão bibliográfica, pelo que se considera ser suficiente recorrer ao seu trabalho.

<sup>16</sup> Também utilizada por Lacerda e Moro (2008).

<sup>17</sup> Esta classificação é apresentada por Soares (2006) (e por isso aqui mantida) e os rácios financeiros assim classificados distinguem-se dos rácios de estrutura de custos por apresentarem no numerador montantes a receber, e não montantes devidos.

	<b>Dimensão</b>	<b>Variável</b>	<b>Tristão (1997)</b>	<b>Gamelas (2005)</b>	<b>Soares (2006)</b>	<b>Bonfim (2006)</b>	<b>Barros (2008)</b>	<b>Lacerda e Moro (2008)</b>
		total (DAT_ACC_AT)						
5	Alavancagem e risco	Dívida comercial/Activo total (DCOM_AT)			X			
6	Alavancagem e risco	Dívidas a IC/Dívida a terceiros (DFIN_DAT)			X			
7	Alavancagem e risco	Dívidas a IC e empréstimos por obrigações e tít. de participação/Activo total (DFIN_OB_AT)			X			
8	Alavancagem e risco	Dívidas a IC/Activo total (DFIN_AT)				X		X
9	Alavancagem e risco	Dívidas a terceiros/Activo total (DAT_AT)			X			
10	Alavancagem	Passivo/Activo total (ALAVANCAG)		X		X		
11	Dimensão	Activo total (ln) (ln_AT)	X	X				X
12	Dimensão	Volume de negócios (ln) (ln_VN)						X
13	Estrutura	Dívida de terceiros líquidas de provisões/Activo total menos imobilizado total líquido (DDT_PROV_AT_IM)			X			
14	Estrutura	Dívida de terceiros líquidas de provisões/Dívidas a terceiros total (DDT_PROV_DAT)			X			
15	Estrutura de custos	Juros e custos similares/Dívida a IC (JCS_DFIN)						X
16	Estrutura de custos	Juros e custos similares/Dívidas a terceiros (JCS_DAT)						X
17	Estrutura de custos	Dívidas a IC/Passivo (DFIN_PT)		X				
18	Estrutura de custos	Dívidas a outra entidades/Passivo (DOT_PT)		X				
19	Estrutura de custos	Dívidas aos accionistas/Passivo (DACC_PT)		X				
20	Gestão de tesouraria	Clientes/(Volume de negócios)*365 (PMR)					X	
21	Gestão de tesouraria	Existências/(Custo das mercadorias vendidas e matérias consumidas+Forn. e serviços externos)*365 (PMRE)					X	
22	Gestão de tesouraria	Fornecedores/(Custo das mercadorias vendidas e matérias consumidas+Forn. e serviços externos)*365 (PMP)					X	
23	Liquidez	Activo circulante/Activo total (AC_AT)			X			
24	Liquidez	Activo circulante/dívidas a				X		

	<b>Dimensão</b>	<b>Variável</b>	<b>Tristão (1997)</b>	<b>Gamelas (2005)</b>	<b>Soares (2006)</b>	<b>Bonfim (2006)</b>	<b>Barros (2008)</b>	<b>Lacerda e Moro (2008)</b>
		terceiros (AC_DAT)						
25	Liquidez	Activo circulante/Passivo curto prazo (AC_PCP)	X	X			X	X
26	Liquidez	Activo circulante excl. existências/Passivo curto prazo (AC_EX_PCP)						X
27	Liquidez	Caixa e depósitos/Activo total (CD_AT)			X			X
28	Liquidez	Caixa e depósitos/Passivo de curto prazo (CD_PCP)					X	X
29	Liquidez	Fundo de Maneio/Activo total (FM_AT)	X				X	X
30	Liquidez	Passivo circulante/Activo total (PCP_AT)						X
31	Liquidez	Passivo circulante/Passivo (PCP_PT)						X
32	Liquidez	Passivo circulante/Passivo excl. provisões (PCP_PT_PROV)						X
33	Rentabilidade	Resultado líquido/Activo total (RL_AT)			X	X		X
34	Rentabilidade	Resultado líquido/Capital próprio (RL_CP)					X	
35	Rentabilidade	Resultado líquido/Volume de negócios (RL_VN)					X	X
36	Rentabilidade	Resultado líquido+impostos + juros e custos similares/Volume de negócios (EBIT_VN)						X
37	Rentabilidade	Resultado líquido+impostos + juros e custos similares/Juros e custos similares (EBIT_JCS)	X					X
38	Rentabilidade	Resultado líquido+impostos+encargos financeiros/Activo total (EBIT_AT)	X		X			X
39	Rentabilidade	Resultado operacional/Activo total (RO_AT)	X					X
40	Rentabilidade	Resultado operacional/Volume de negócios (RO_VN)						X
41	Rentabilidade	Resultados transitados/Activo total (RT_AT)	X					
42	Solvência	Activo fixo/Activo total (IM_AT)		X				
43	Solvência	Capital próprio/Passivo total (CP_PT)	X				X	
44	-	Antiguidade da empresa (meses)						X

### 3.2. Estimação e diagnóstico da regressão logística

O modelo de regressão logística a estimar tem a seguinte especificação, considerando  $P$  variáveis explicativas:

$$\text{logit}(p_i) = \beta_0 + \sum_p \beta_p X_{ip}, \quad (15)$$

A regressão logística é estimada pelo método da máxima verosimilhança e, numa lógica de análise exploratória, as variáveis explicativas serão testadas com recurso a selecção *stepwise*, que permite que haja um processo iterativo no qual são excluídas do modelo variáveis não significativas antes da introdução de uma nova variável significativa, com base num critério estatístico. A interpretação dos coeficientes é feita em termos de *odds-ratio*<sup>18</sup>.

A qualidade do ajustamento global do modelo será avaliada recorrendo a várias estatísticas. Será apresentado o coeficiente de determinação generalizado proposto por Cox e Snell (1989) e o ajustamento a esse coeficiente proposto por Nagelkerke (1991):

$$R_{CS}^2 = 1 - e^{(-2/n(LL(\text{mod}) - LL(\text{const})))} \quad (16)$$

$$R_N^2 = \frac{R_{CS}^2}{1 - e^{(2(LL(\text{const}))/n)}} \quad (17)$$

Ambos se baseiam no logaritmo da verosimilhança (*log-likelihood*) do modelo estimado e do modelo apenas com constante e na dimensão da amostra,  $n$ , e a sua interpretação é semelhante à interpretação do coeficiente de determinação da regressão linear, ou seja, dão uma medida da “variância” da variável dependente explicada pelo modelo. O pseudo-R<sup>2</sup> proposto por Nagelkerke tem a vantagem de assumir 1 como valor máximo, sendo por isso mais fácil de interpretar.

O teste à bondade de ajustamento de Hosmer e Lemeshow divide as observações em decis, de acordo com a probabilidade estimada de sucesso (neste caso, de incumprimento), sendo calculada a estatística de Qui-quadrado de Pearson a partir de uma tabela de frequências observadas e esperadas. A hipótese nula subjacente ao teste é a de que não existe diferença entre os valores observados e os valores estimados pelo modelo, o que implica que o modelo se ajusta aos dados a um nível aceitável. Contudo, este teste não pretende verificar se o

<sup>18</sup> Recorde-se que  $\text{logit}(p_i) = \ln(p_i / (1 - p_i))$ . O *odds* corresponde à probabilidade de incumprimento ( $p_i$ ) sobre a probabilidade de não incumprimento ( $1 - p_i$ ).

modelo explica uma larga proporção da variância da variável dependente, mas apenas se a proporção que explica é significativa. As estatísticas *Deviance* e *Pearson* testam o mesmo que o teste de Hosmer e Lemeshow, ou seja, se os valores estimados pelo modelo diferem muito dos valores observados.

A tabela de classificação permite conhecer os casos correctamente classificados, cruzando os casos observados com os casos classificados, constituindo um instrumento interessante de análise da capacidade discriminante do modelo. A partir desta tabela é possível conhecer o erro de tipo I (falsos positivos ou empresas “saudáveis” mas classificadas como estando em incumprimento) e o erro de tipo II (falsos negativos ou empresas em incumprimento classificadas como “saudáveis”), sendo para isso necessário estabelecer a probabilidade a partir da qual uma empresa é classificada como estando em incumprimento<sup>19</sup>.

A curva *Receiver Operating Characteristic* (ROC) constitui um interessante instrumento visual de análise da capacidade discriminante do modelo, representando graficamente a proporção de casos correctamente classificados e a proporção de falsos positivos (erro de tipo I). A diagonal da curva representa uma classificação perfeitamente aleatória, na qual cada empresa teria a mesma probabilidade de ser classificada em cada uma das situações. É possível estabelecer uma relação entre a área abaixo da curva ROC (AUC) e o coeficiente de *Gini* (G), que mede o quão próximo se encontra o modelo estimado do modelo com uma capacidade de discriminação perfeita, dada por:  $G + 1 = 2AUC$ .

Para testar a significância individual dos parâmetros estimados, a análise recai na estatística de Wald, assente no rácio entre a estimativa do coeficiente e o respectivo erro-padrão (S.E.),

$\left( \hat{\beta}_p / S.E.(\hat{\beta}_p) \right)^2$  que segue uma distribuição do Qui-quadrado. Caso se rejeite a hipótese nula

$H_0: \beta_p = 0$ , então o parâmetro estimado é estatisticamente significativo.

<sup>19</sup> Essa probabilidade vai ser fixada em 0,5.

### 3.3. Estimação e diagnóstico do modelo multinível com função de ligação *logit*

A estimação dos modelos multinível será feita recorrendo ao PROC Glimmix (SAS)<sup>20</sup>, um procedimento aplicável a Modelos Lineares Generalizados Mistos (MLGM ou GLMM, correspondente à denominação em língua inglesa *Generalized Linear Mixed Models*), que permite ajustar modelos estatísticos a dados com variância não constante e em que a variável dependente não segue necessariamente uma distribuição normal, *i.e.*, modelos GLM. Este procedimento permite seleccionar a distribuição da variável dependente, mas condicionada a efeitos aleatórios normalmente distribuídos.

Apresenta-se de seguida a notação utilizada para representar o modelo. Supondo que  $\mathbf{Y}$  representa um vector de dados observados de dimensão  $(n \times 1)$  e que  $\mathbf{u}$  representa um vector de efeitos aleatórios de dimensão  $(J \times 1)$ , os modelos ajustados pelo PROC Glimmix assumem que:

$$E[\mathbf{Y}|\mathbf{u}] = g^{-1}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}) \quad (18)$$

em que  $g^{-1}$  corresponde à inversa de uma função de ligação monótona e diferenciável. A matriz  $\mathbf{X}$  é uma matriz de dimensão  $(n \times p)$  e  $\mathbf{Z}$  é uma matriz *design* de dimensão  $(n \times J)$  referente aos efeitos aleatórios. Assume-se que os efeitos aleatórios seguem uma distribuição normal, com média 0 e uma matriz de variância-covariância  $\mathbf{G}$ , ou seja,  $\mathbf{u} \sim N(\mathbf{0}, \mathbf{G})$ . Os GLMM contêm uma componente linear mista na inversa da função de ligação. Esta componente do modelo referida como a componente linear mais não é do que:

$$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}. \quad (19)$$

A variância das observações, condicionada aos efeitos aleatórios, é:

$$\text{var}(\mathbf{Y}|\mathbf{u}) = \mathbf{A}^{1/2} \mathbf{R} \mathbf{A}^{1/2}. \quad (20)$$

A matriz  $\mathbf{A}$  é uma matriz diagonal que contém as funções de variância do modelo. O PROC Glimmix determina a função de variância a partir da distribuição que é referida no código.

<sup>20</sup> Em <http://www.sas.com/apps/demosdownloads/setupcat.jsp?cat=SAS%2FSTAT+Software>. O *software* encontra-se disponível para *download* pelos utilizadores SAS [efectuado o *download* a 15/01/2010]. Na mesma localização é possível encontrar a respectiva documentação de apoio, da qual se retirou a notação do modelo que é apresentada neste capítulo.

Neste caso, a distribuição dos dados é binária, sendo a função de variância estabelecida por defeito como  $p_{ij}(1-p_{ij})$ . A função de ligação é a função *logit*.

A matriz  $\mathbf{R}$  é uma matriz de variâncias especificada pelo utilizador através do comando RANDOM. Se a distribuição dos dados contém um parâmetro adicional, esse parâmetro será parte ou da função de variância ou da matriz  $\mathbf{R}$ . Este parâmetro é o parâmetro de dispersão  $\phi$  já discutido na Secção 2.4 e, no caso de dados binários, esse valor deve ser próximo de 1. O PROC Glimmix denomina os efeitos aleatórios pertencentes à matriz  $\mathbf{u}$  como efeitos aleatórios “*G-side*”.

As colunas de  $\mathbf{X}$  são construídas a partir das variáveis que forem introduzidas como variáveis explicativas (do lado direito do comando MODEL). As colunas de  $\mathbf{Z}$  e as matrizes de variâncias  $\mathbf{G}$  e  $\mathbf{R}$  são construídas a partir do comando RANDOM. A matriz  $\mathbf{R}$  é por defeito definida como  $\mathbf{R}=\phi\mathbf{I}$ , sendo  $\mathbf{I}$  a matriz identidade, e a estimativa do parâmetro  $\phi$  deverá ser próxima de 1 se a distribuição das variáveis adere efectivamente à distribuição Bernoulli. Assim, não será necessário especificar uma matriz  $\mathbf{R}$  distinta da matriz definida por defeito. Refira-se que  $\phi$  é um parâmetro estimado e não fixo, pelo que se a sua estimativa for distinta de 1 poderá existir um problema de dispersão.

Os parâmetros a estimar são o vector de parâmetros fixos  $\beta$  e um vector de parâmetros de variância-covariância  $\theta$  que contém os parâmetros desconhecidos da matriz  $\mathbf{G}$ . É possível obter as soluções para o vector  $\mathbf{u}$ , ou seja, o vector dos efeitos aleatórios.

O método de estimação seleccionado foi RSPL, um acrónimo para: R = *residual (restricted log-likelihood)*, S = *subject-specific expansion*, PL = *pseudo-likelihood*. Sendo aplicada uma linearização às estimativas dos coeficientes a estimar, este processo de estimação aplica uma técnica assente na função de verosimilhança aplicada a um pseudo-modelo e a pseudo-dados<sup>21</sup>, uma vez que a estimação assenta em variáveis latentes (os efeitos aleatórios) e não em variáveis observadas. O prefixo *pseudo* encontra-se, assim, em várias estatísticas de ajustamento do modelo<sup>22</sup>. Neste processo de estimação, o processo de optimização seleccionado corresponde a um processo duplamente iterativo – *dual quasi-Newton* – que não

<sup>21</sup> Esta aproximação foi proposta por Wolfinger e O’Connell (1993).

<sup>22</sup> Cf. Anexo 1, onde se apresenta a transformação do modelo proposta por Wolfinger e O’Connell (1993), tal como consta da documentação de apoio do SAS Glimmix.

requer segundas derivadas, por ser um processo por aproximação, e que impõe um limite inferior de zero para a variância do efeito do grupo considerado. Em cada passo, um modelo linear misto é ajustado e a optimização de cada pseudo-modelo linear pode necessitar de várias iterações. A vantagem de considerar um processo de optimização duplamente iterativo é a possibilidade, quando se atinge a convergência, de obter os instrumentos de inferência estatística que se aplicam aos modelos mais simples<sup>23</sup>.

Em primeiro lugar, irá estimar-se um modelo apenas com termo independente, que prevê a probabilidade de incumprimento, na forma multinível:

$$\text{logit}(p_{ij}) = \gamma_{00} + u_{0j}, \text{ com } u_{0j} \sim N(0, \sigma_{00}^2) \quad (21)$$

Recorde-se que  $\gamma_{00}$  corresponde ao efeito fixo e  $u_{0j}$  ao efeito aleatório, neste caso sujeito aos sectores de actividade económica, pelo que apenas vamos ter um parâmetro de variância. Será a partir deste modelo que se irá calcular a correlação intra-classe, discutida na Secção 2.2, e que é dada por  $\rho = \sigma_{00}^2 / (\sigma_{00}^2 + \sigma^2)$  (equação 8), sendo  $\sigma^2 = \pi^2 / 3$  a variância da distribuição logística.

Depois de se estimar este modelo, tentar-se-á estimar um modelo considerando efeitos aleatórios ao nível das variáveis que resultaram estatisticamente significativas na regressão logística.

A análise do ajustamento do modelo será efectuada com base nas estatísticas de desvio clássicas, como *-2 res Log Likelihood* ou o critério de informação de Akaike (AIC), mas com o prefixo pseudo, porque são construídas sobre o logaritmo de uma pseudo-verosimilhança (*-2 res Log Pseudo-Likelihood* e *pseudo-AIC*). Estas estatísticas são particularmente úteis para comparar modelos multiníveis.

Os ensaios de hipóteses sobre os parâmetros estimados assentam na hipótese nula desses parâmetros serem zero, incluindo-se neste procedimento os efeitos aleatórios  $u_{0j}$ , que não constituem exactamente parâmetros. Recorrendo à especificação do modelo apenas com termo independente, isso corresponderá às hipóteses nulas  $H_0 : \gamma_{00} = 0$  e  $H_0 : u_{0j} = 0$ , as quais podem ser testadas com um teste *t* (Snijders e Bosker, 1999), em que a estatística de

<sup>23</sup> Documentação de apoio ao SAS Glimmix, pp. 151.



teste corresponde ao rácio entre a estimativa do parâmetro e o respectivo erro-padrão. Sob a hipótese nula, a estatística de teste segue uma distribuição aproximada *t de Student*, embora a determinação do número de graus de liberdade a considerar seja mais complicada do que na regressão linear clássica, devido à existência de dois níveis. Optou-se por uma aproximação de *Satterthwaite*<sup>24</sup> para os graus de liberdade a considerar, já que este método inclui graus de liberdade para os efeitos fixos e para os efeitos aleatórios.

Podem igualmente ser efectuados ensaios de hipóteses em relação aos parâmetros de variância. Mais uma vez, partindo das estimativas produzidas e dos respectivos erros-padrão, é possível construir uma estatística de teste correspondente ao rácio entre a estimativa da variância e o respectivo erro-padrão, o chamado teste de Wald, que pode ser avaliado em relação à distribuição Normal-Padrão (Hox, 2002). Contudo, Raudenbush e Bryk (2002) e Hox (2002) referem que esse teste, quando aplicado às variâncias, deve ser feito unilateralmente, já que é uma forma conservadora de considerar que o pressuposto da aproximação à distribuição Normal pode não ser o mais adequado, para além do facto de não fazer sentido considerar, no teste, a hipótese alternativa de a variância assumir um valor negativo.

---

<sup>24</sup> Os detalhes de cálculo podem ser consultados na documentação de apoio ao PROC Glimmix [disponível em <http://www.sas.com/apps/demosdownloads/setupcat.jsp?cat=SAS%2FSTAT+Software>, consultado em 15/01/2010].

## 4. Caracterização dos dados utilizados

### 4.1. Descrição da base de dados

A base de dados utilizada no presente estudo foi disponibilizada pela Coface Serviços Portugal, S.A., dela constando elementos caracterizadores do tecido empresarial português, informações sobre processos de insolvência e informações sobre balanços.

Os campos constantes das várias tabelas fornecidas são apresentados na Tabela 2.

**Tabela 2 – Descrição da base de dados**

<b>Tabela</b>	<b>Campos</b>
Insolvências (16 808 registos)	ID (código identificativo da empresa)
	Atributo (tipo de processo - Apresentada a Insolvência, Declarada Insolvente, Declarada Insolvente (apresentada), Declarada Insolvente (requerida), Declarada a Falência, Requerida a Insolvência)
	Estado (estado do processo)
	Data (data de interposição do processo)
Base (867 440 registos)	ID (código identificativo da empresa)
	Dataconst (data de constituição da empresa)
	Situação actual
	Data da situação actual
	ID_CAE (código de classificação de actividade económica – 5 dígitos)
	CAE (descritivo da actividade económica)
Balanços (1 258 568 registos)	ID (código identificativo da empresa)
	ID_balanço
	Ano
	Output (analítico ou sintético)
Sintéticos	ID_bal (código identificativo do balanço), rubrica, valor
Analíticos	ID_bal (código identificativo do balanço), rubrica, valor

De acordo com o Artigo 1.º do Código de Insolvência e da Recuperação de Empresas, estabelecido pelo Decreto-Lei n.º 53/2004 de 18 de Março de 2004, “o processo de insolvência é um processo de execução universal que tem como finalidade a liquidação do património de um devedor insolvente e a repartição do produto obtido pelos credores, ou a satisfação destes pela forma prevista num Plano de insolvência, que nomeadamente se baseie na recuperação da empresa compreendida na massa insolvente”. No ficheiro de insolvências estão contabilizados os processos desse tipo registados entre 2006 e 2009, independentemente de quem requereu a insolvência, o que pode acontecer por parte do devedor, por parte de qualquer credor, ainda que condicional e qualquer que seja a natureza do seu crédito, ou ainda pelo Ministério Público, em representação das entidades cujos interesses lhe estão legalmente confiados, verificando-se algum dos factos enunciados no número 1 do artigo 20.º do Código de Insolvência e da Recuperação de Empresas. Como uma empresa pode conhecer processos interpostos por diversos agentes, interessa conhecer não tanto o número de processos, mas sim o número de empresas por eles afectadas.

**Tabela 3 – Processos de insolvência**

Ano	N.º processos	N.º empresas
2006	2 254	2 015
2007	3 199	2 828
2008	4 674	4 026
2009	6 681	5 607
Total	16 808	14 476

Não se irá ter em consideração o estado do processo nem o autor, entendendo-se que a existência de acções judiciais deste tipo constitui um indício forte de uma situação de incumprimento, no sentido em que existe uma forte probabilidade de a empresa não respeitar na íntegra as suas responsabilidades perante terceiros.

Como ilustra a Tabela 3, o número de insolvências aumentou consideravelmente em 2008 e 2009, reflectindo o facto de Portugal se situar (de forma não isolada, refira-se) numa fase baixa do ciclo económico, com uma deterioração continuada das condições económicas nos últimos anos. Para efeitos do presente estudo, utilizar-se-á como referência as empresas que registaram processos de insolvência no ano de 2009 e serão considerados os balanços do ano anterior ao ano em que se registou a insolvência, *i.e.*, os balanços de 2008.

A opção por uma abordagem “estática” do problema justifica-se pelo facto de voluntariamente não se querer considerar o factor “posição no ciclo económico” e porque os dados de insolvência do ano 2009 acabam por ser particularmente favoráveis ao estudo do incumprimento, dada a dimensão do fenómeno.

Para cálculo das taxas de incumprimento (total e sectoriais) procede-se à comparação entre o número de empresas em incumprimento com balanço de 2008 disponível e o número de balanços de 2008 de empresas “saudáveis”. O cálculo é feito desta forma uma vez que é o único modo de que se dispõe para contabilizar o número de empresas em actividade, não sendo possível utilizar os dados de balanços de 2009 porque após o processo de insolvência a empresa, em princípio, deixará de depositar as suas demonstrações financeiras.

Tendo-se seleccionado as empresas com balanços referentes a 2008, acrescentaram-se as suas características, presentes na tabela *base*. Ao resultado desta tabela, haverá que acrescentar colunas relativas às empresas em incumprimento em 2009 (incump=1 caso tenha existido um processo de falência, incump=0, caso contrário).

Existem 321 803 balanços de 2008. Contudo, chama-se a atenção para o facto de algumas empresas poderem ter dois balanços, um consolidado e outro não consolidado. Quando existe balanço consolidado, será esse o utilizado. Assim, e porque os 321 803 balanços de 2008 dizem respeito a 321 566 empresas (pelo motivo atrás apresentado), este deve ser o número de balanços a considerar. Neste ficheiro com 321 566 empresas identificaram-se as empresas que registaram processos de insolvência em 2009 e retiraram-se empresas que não registaram processos de insolvência em 2009 mas com um descritivo de situação actual que levantava dúvidas sobre a sua condição.

Recorde-se que se tinham contabilizado 5 607 empresas com processos de insolvência em 2009. Contudo, destas, apenas 1 563 empresas têm balanços relativos a 2008. Por outro lado, as empresas sobre as quais não existe qualquer dúvida sobre a situação são 291 973, que correspondem a empresas com situação actual “em actividade” e “elementos disponíveis”, ou seja, informação que indica que são empresas em pleno funcionamento.

Se se tivesse acesso aos balanços de 2008 das 5 607 empresas que registaram processos de insolvência em 2009, a taxa de insolvência definida como  $[\text{n.º empresas com processo de insolvência em 2009}]/(\text{n.º empresas com processo de insolvência em 2009 e balanço 2008} + \text{n.º})$

empresas saudáveis em 2008)]\*100% seria  $[5\ 607/(5\ 607+291\ 973)]*100\%$ , ou seja, 1,88%. Assim, é apenas de  $[1\ 563/(1\ 563+219\ 973)]*100\%=0,53\%$ .

Contudo, e ainda que essa informação não vá ser tratada, considera-se ser de salientar que, conforme ilustrado na Tabela 4, das 5 607 empresas que enfrentaram processos de insolvência em 2009, 1 563 têm como último balanço disponível o do ano de 2008, 1 351 têm como último balanço disponível o do ano de 2007, 666 empresas têm como último balanço disponível o do ano de 2006 e 86 empresas têm como último balanço disponível o do ano de 2005. Para 1 941 empresas não se dispõe de informação contabilística.

**Tabela 4 – Disponibilidade de informação financeira das empresas com processos de insolvência interpostos em 2009**

<b>Empresas com processos de insolvência interpostos em 2009</b>	<b>5.607</b>
- Das quais com último balanço de 2008	1 563 (28%)
- Das quais com último balanço de 2007	1 351 (24%)
- Das quais com último balanço de 2006	666 (12%)
- Das quais com último balanço de 2005	86 (2%)
- Das quais sem informação de balanço	1 941 (35%)

Uma justificação que se pode avançar para este fenómeno, ainda que sem grande base de sustentação, é que as empresas com dificuldades são “tradicionalmente” menos transparentes e mais relutantes em depositar as suas demonstrações financeiras.

Para completar a base de empresas, é necessário relacionar cada empresa com os respectivos rácios financeiros, calculados a partir dos dados de balanço. Porém, como existem empresas que só dispõem de balanços sintéticos, os quais colocam limitações ao cálculo de muitos rácios, esses casos não serão considerados.

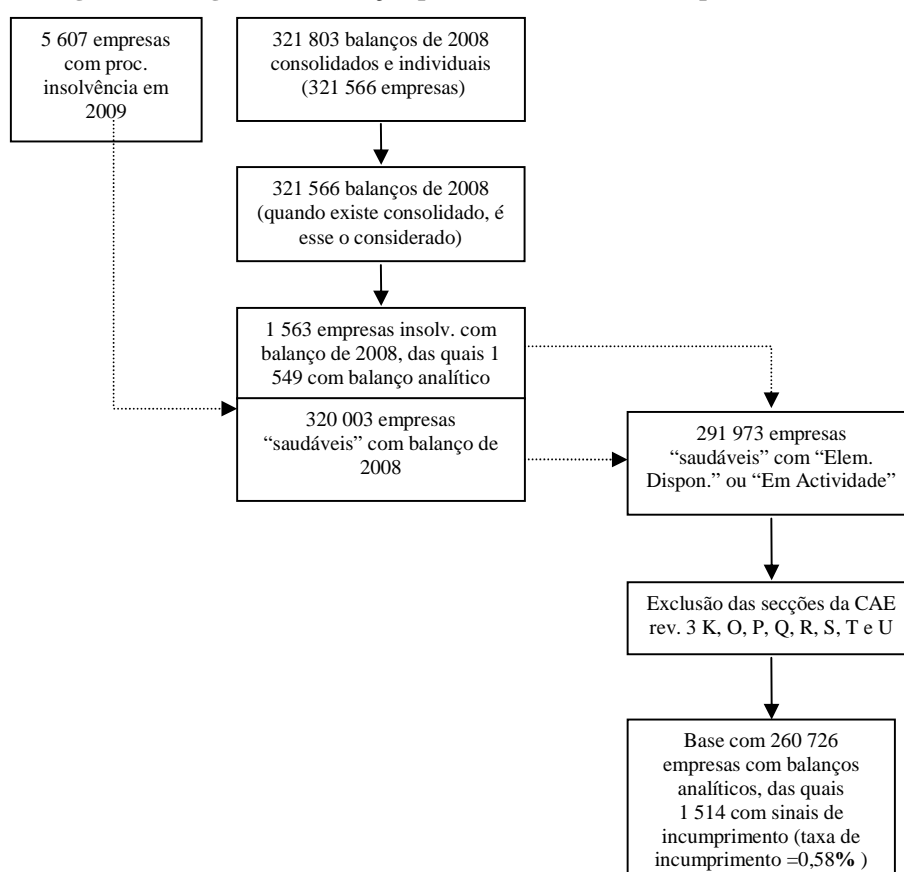
Em termos de sectores de actividade a considerar<sup>25</sup>, irão excluir-se as empresas cuja actividade principal é, de algum modo, similar à prestação de serviços públicos ou de carácter colectivo (Soares, 2006), e que correspondem às seguintes secções: K – Actividades financeiras e de seguros, O – Administração Pública e Defesa; Segurança Social Obrigatória; P – Educação; Q – Actividades de saúde humana e apoio social, R – Actividades artísticas, de espectáculos, desportivas e recreativas, S – Outras actividades de serviços, T –

<sup>25</sup> As empresas estão classificadas de acordo com a Classificação das Actividades Económicas Rev.3 (CAE Rev. 3), considerando-se, no presente caso, a classificação a dois dígitos (designada de Divisão) e a classificação a um dígito (classificação alfabética, designada de Secção).

Actividades das famílias empregadoras de pessoal doméstico e actividades de produção das famílias para uso próprio e U – Actividades dos organismos internacionais e outras instituições extra-territoriais.

Nesta fase, dispõe-se de uma base de estudo que consiste em 260 726 empresas, das quais 1 514 apresentaram sinais de incumprimento em 2009, o que corresponde a uma taxa de incumprimento de 0,58%. Contudo, as análises descritivas, a análise de valores omissos e de multicolinearidade, as quais serão apresentadas nas secções seguintes, irão restringir a amostra de empresas considerada para efeitos de estimação. De qualquer forma, afigura-se útil esquematizar a selecção preliminar de empresas a considerar, ilustrada na Figura 1.

**Figura 1 – Diagrama da selecção preliminar da base de empresas a estudar**



#### 4.2. Análise de valores omissos e de multicolinearidade

Analizando os valores omissos, determinou-se não se considerar as variáveis com mais de 20% de observações nessa situação (Tabela 5).

**Tabela 5 – Análise de valores omissos**

	N	NMiss	% NMISS	P1	P99
AC_EX_PCP	252 246	8 480	3,25%	0,01	119,78
AC_PCP	252 246	8 480	3,25%	0,04	196,72
ALAVANCAG	258 564	2 162	0,83%	0,01	6,55
AT_VN	231 283	29 443	11,29%	0,12	82,37
CD_AT	254 427	6 299	2,42%	0,00	0,96
CD_PCP	246 906	13 820	5,30%	0,00	57,81
CP_AT	260 516	210	0,08%	-5,51	1,00
CP_PT	258 649	2 077	0,80%	-0,85	100,35
DACC_PT	141 775	118 951	45,62%	0	1,00
DAT_ACC_AT	141 698	119 028	45,65%	0	3,4
DAT_AT	258 309	2 417	0,93%	0,01	6,47
DCOM_AT	16 622	244 104	93,62%	0	3,96
DDT_PROV_AT_IM	253 568	7 158	2,75%	-10,55	11,7
DDT_PROV_DAT	251 326	9 400	3,61%	0	30,63
DFIN_AT	121 924	138 802	53,24%	0	1,25
DFIN_DAT	121 931	138 795	53,23%	0	0,98
DFIN_OB_AT	122 290	138 436	53,10%	0	1,26
DFIN_PT	121 931	138 795	53,23%	0	0,97
DOT_PT	181 597	79 129	30,35%	0	1
EBIT_AT	257 520	3 206	1,23%	-1,93	0,52
EBIT_JCS	230 839	29 887	11,46%	-1 464,17	1 586,31
EBIT_VN	231 306	29 420	11,28%	-4,6	0,91
FM_AT	260 539	187	0,07%	-4,54	1
IM_AT	250 703	10 023	3,84%	0	0,98
JCS_DFIN	118 861	141 865	54,41%	0	47,13
JCS_DAT	230 711	30 015	11,51%	0	0,34
ln_AT	260 524	202	0,08%	8,24	16,77
ln_VN	231 307	29 419	11,28%	7,54	16,58
PCP_AT	258 309	2 417	0,93%	0	5,31
PCP_PT	258 418	2 308	0,89%	0	1
PCP_PT_PROV	258 418	2 308	0,89%	0	1
PMP	207 320	53 406	20,48%	0	7 636,83
PMR	154 845	105 881	40,61%	0	1 553,3
PMRE	168 956	91 770	35,20%	0,81	117 907,41
RL_AT	257 520	3 206	1,23%	-2,03	0,49
RL_CP	257 601	3 125	1,20%	-6,91	6,24
RL_VN	231 306	29 420	11,28%	-5,69	0,72
RO_AT	257 425	3 301	1,27%	-1,84	0,57
RO_VN	231 306	29 420	11,28%	-5,29	0,79
RT_AT	206 693	54 033	20,72%	-8,38	0,78
VN	231 314	29 412	11,28%	1864	15 808 778

	N	NMiss	% NMISS	P1	P99
VN_AT	231 285	29 441	11,29%	0,01	8,32
antiguidade	257 287	3 439	1,32%	2,47	774,2

Retirando as variáveis DACC\_PT, DAT\_ACC\_AT, DCOM\_AT, DFIN\_AT, DFIN\_DAT, DFIN\_OB\_AT, DFIN\_PT e DOT\_PT, subsistem observações com valores omissos, as quais, para não se incorrer em potenciais enviesamentos decorrentes de imputação de omissos, não vão ser consideradas. Retirando estes casos, ficamos com 198 095 empresas, sendo que 1 226 empresas tiveram sinais de incumprimento em 2009 (0,62%).

A introdução de variáveis explicativas correlacionadas acarreta vários problemas. Por um lado, o significado do modelo poderá perder-se, uma vez que podem coexistir variáveis das quais se retira a mesma informação, o que faz com que o modelo perca valor em termos de contribuição para a explicação de um determinado fenómeno. Por outro lado, as estimativas dos coeficientes das variáveis explicativas ficam afectadas, com possibilidade de os seus sinais serem estimados erroneamente, e as conclusões dos ensaios de hipóteses podem ficar comprometidas, uma vez que os erros-padrão tendem a ser inflacionados e maiores do que seriam caso as variáveis não estivessem correlacionadas.

A partir da construção de uma matriz de correlações<sup>26</sup>, identificaram-se as correlações superiores a 0,5 ou inferiores a -0,5. Além disso, fizeram-se algumas análises com vista a aprimorar a identificação de variáveis com problemas de colinearidade, calculando-se a tolerância e o seu inverso *Variance Inflation Factor* (VIF), normalmente aplicados nas regressões lineares múltiplas. A tolerância é uma medida da proporção de variância de uma variável explicativa que não é explicada pelas restantes, pelo que interessará reter variáveis com elevados níveis de tolerância. O seu inverso, VIF, representa um factor multiplicativo associado à variância na presença de problemas de colinearidade. Por outras palavras, a variância do coeficiente estimado para determinada variável pode ser várias vezes superior ao que seria de esperar se as variáveis explicativas não fossem correlacionadas. Uma vez identificadas as variáveis correlacionadas, seleccionaram-se as variáveis de acordo com o seu nível de tolerância (e, quando caso disso, pelo interesse da informação que veiculam). A

<sup>26</sup> Cf. Anexo 2, onde se encontra a matriz de correlações, na qual se apresentam as 10 variáveis mais correlacionadas.



Tabela 6 apresenta o diagnóstico de colinearidade das variáveis explicativas com base na tolerância e no VIF.

**Tabela 6 – Diagnóstico de colinearidade das variáveis explicativas**

Diagnóstico de colinearidade	
AC_AT	<b>Altamente correlacionada com IM_AT; tolerância 0,037; seleciona-se AC_AT por ser mais interessante ter uma medida de liquidez</b>
AC_DAT	Correlacionada com DDT_PROV_DAT, AC_PCP, AC_EX_PCP e CD_PCP e tolerância 0,218; seleciona-se CD_PCP pelo nível de tolerância
AC_EX_PCP	Correlacionada com AC_PCP, AC_DAT, DDT_PROV_DAT e CD_PCP e tolerância 0,089; seleciona-se CD_PCP pelo nível de tolerância
AC_PCP	Correlacionada com AC_EX_PCP, AC_DAT, DDT_PROV_DAT e CD_PCP e tolerância 0,086; seleciona-se CD_PCP pelo nível de tolerância
ALAVANCAG	Correlacionada com CP_AT, DAT_AT, PCP_AT, FM_AT, RL_AT e EBIT_AT e tolerância 1.1506E-01; seleciona-se FM_AT pelo nível de tolerância
<b>Antiguidade</b>	<b>Sem problemas de correlação e tolerância 0,929</b>
AT_VN	<b>Sem problemas de correlação e tolerância 0,692</b>
CD_AT	<b>Sem problemas de correlação e tolerância 0,841</b>
CD_PCP	<b>Correlacionada com AC_PCP, AC_EX_PCP e tolerância 0,609; seleciona-se CD_PCP pelo nível de tolerância</b>
CP_AT	Correlacionada com ALAVANCAG, DAT_AT, PCP_AT, FM_AT, RL_AT, EBIT_AT e tolerância 1.1507E-01; seleciona-se FM_AT pelo nível de tolerância
CP_PT	<b>Sem problemas de correlação e tolerância 0,783</b>
DAT_AT	Correlacionada com ALAVANCAG, CP_AT, PCP_AT, FM_AT, RL_AT e EBIT_AT e tolerância 0,005; seleciona-se FM_AT pelo nível de tolerância
<b>DDT_PROV_AT_IM</b>	<b>Sem problemas de correlação e tolerância 0,999</b>
DDT_PROV_DAT	Correlacionada com AC_DAT, AC_EX_PCP, AC_PCP e tolerância 0,273
EBIT_AT	Correlacionada com RL_AT, RO_AT, PCP_AT, ALAVANCAG, CP_AT, FM_AT e DAT_AT e tolerância 0,010; seleciona-se RO_AT pelo nível de tolerância
<b>EBIT_JCS</b>	<b>Sem problemas de correlação e tolerância 0,997</b>
EBIT_VN	Correlacionada com RO_VN e RL_VN e tolerância 0,094; seleciona-se RL_VN por ter uma maior tolerância
FM_AT	<b>Correlacionada com PCP_AT, DAT_AT, CP_AT, ALAVANCAG, RL_AT e EBIT_AT e tolerância 0,116; seleciona-se FM_AT pelo nível de tolerância</b>
IM_AT	Altamente correlacionada com AC_AT, tolerância 0,038
<b>JCS_DAT</b>	<b>Sem problemas de correlação e tolerância 0,972</b>
Ln_AT	Correlacionada com Ln_VN e tolerância 0,282; seleciona-se Ln_VN
<b>Ln_VN</b>	<b>Correlacionada com Ln_AT e tolerância 0,302; seleciona-se Ln_VN</b>
PCP_AT	Correlacionada com FM_AT, DAT_AT, CP_AT, ALAVANCAG, RL_AT e EBIT_AT e tolerância 0
PCP_PT	Altamente correlacionada com PCP_PT_PROV e tolerância 0,004; seleciona-se PCP_PT_PROV por ser uma medida líquida, do lado do passivo
<b>PCP_PT_PROV</b>	<b>Altamente correlacionada com PCP_PT e tolerância 0,004; seleciona-se PCP_PT_PROV por ser uma medida líquida, do lado do passivo</b>
RL_AT	Correlacionada com EBIT_AT, RO_AT, ALAVANCAG, CP_AT, PCP_AT, FM_AT e DAT_AT e tolerância 0,009; seleciona-se RO_AT pelo nível de tolerância
<b>RL_CP</b>	<b>Sem problemas de correlação e tolerância 0,999</b>
<b>RL_VN</b>	<b>Correlacionada com RO_VN e EBIT_VN e tolerância 0,135; seleciona-se RL_VN por ter uma maior tolerância</b>
<b>RO_AT</b>	<b>Correlacionada com RL_AT e EBIT_AT e tolerância 0,389; seleciona-se RO_AT pelo nível de tolerância</b>
RO_VN	Correlacionada com RL_VN e EBIT_VN e tolerância 0,093; seleciona-se

Diagnóstico de colinearidade	
	RL_VN por ter uma maior tolerância
VN_AT	<b>Sem problemas de correlação e tolerância 0,519</b>

Os valores das estatísticas de diagnóstico de colinearidade após selecção de variáveis são apresentados na Tabela 7.

**Tabela 7 – Valores das estatísticas de diagnóstico de colinearidade após selecção de variáveis**

	Tolerância	VIF
Antiguidade	0,944	1,059
AC_AT	0,844	1,185
AT_VN	0,902	1,108
CD_AT	0,880	1,136
CD_PCP	0,948	1,055
CP_PT	0,939	1,065
DDT_PROV_AT_IM	0,999	1,000
EBIT_JCS	0,998	1,002
FM_AT	0,738	1,355
JCS_DAT	0,979	1,021
ln_VN	0,859	1,165
PCP_PT_PROV	0,897	1,115
RL_CP	0,999	1,000
RL_VN	0,902	1,109
RO_AT	0,735	1,360
VN_AT	0,814	1,228

Para confirmar que não existem problemas de multicolinearidade nas variáveis seleccionadas, calculou-se o *Condition Index*, tendo-se verificado que o maior valor desse índice não ultrapassa 30<sup>27</sup>, habitualmente considerado como valor de referência para a existência de problemas graves de multicolinearidade.

### 4.3. Análise descritiva e obtenção da amostra

Tendo-se estabilizado a base de empresas a considerar, procede-se de seguida à sua análise descritiva. Recorde-se que, nesta fase, estão a ser consideradas 198 095 empresas, sendo que 1 226 empresas tiveram sinais de incumprimento em 2009 (0,62%).

<sup>27</sup> Cf. Anexo 3.

Tabela 8 – Distribuição de empresas por sector de actividade e por dimensão

Secção CAE Rev. 3	Distribuição de empresas					Total	
	1	2	3	4			
	Micro empresas	Pequenas empresas	Médias empresas	Grandes empresas			
<b>A</b>	Agricultura, produção animal, caça, floresta e pesca	5 406	188	27	4	5 625	2,84%
		96,11%	3,34%	0,48%	0,07%	100,00%	
<b>B</b>	Indústrias extractivas	539	81	15	3	638	0,32%
		84,48%	12,70%	2,35%	0,47%	100,00%	
<b>C</b>	Indústrias transformadoras	25 266	2 876	767	191	29 100	14,69%
		86,82%	9,88%	2,64%	0,66%	100,00%	
<b>D</b>	Electricidade, gás, vapor, água quente e fria e ar frio	177	89	27	15	308	0,16%
		57,47%	28,90%	8,77%	4,87%	100,00%	
<b>E</b>	Captação, tratamento e distribuição de água; saneamento, gestão de resíduos	342	116	54	4	516	0,26%
		66,28%	22,48%	10,47%	0,78%	100,00%	
<b>F</b>	Construção	24 186	1 584	278	69	26 117	13,18%
		92,61%	6,07%	1,06%	0,26%	100,00%	
<b>G</b>	Comércio por grosso e a retalho; reparação de veículos automóveis e motociclos	59 776	5 755	1 137	224	66 892	33,77%
		89,36%	8,60%	1,70%	0,33%	100,00%	
<b>H</b>	Transportes e armazenagem	10 947	587	183	39	11 756	5,93%
		93,12%	4,99%	1,56%	0,33%	100,00%	
<b>I</b>	Alojamento, restauração e similares	17 370	302	44	7	17 723	8,95%
		98,01%	1,70%	0,25%	0,04%	100,00%	
<b>J</b>	Actividades de informação e de comunicação	3 767	207	74	27	4 075	2,06%
		92,44%	5,08%	1,82%	0,66%	100,00%	
<b>L</b>	Actividades imobiliárias	8 917	307	47	5	9 276	4,68%
		96,13%	3,31%	0,51%	0,05%	100,00%	
<b>M</b>	Actividades de consultoria, científicas, técnicas e similares	18 336	462	85	24	18 907	9,54%
		96,98%	2,44%	0,45%	0,13%	100,00%	
<b>N</b>	Actividades administrativas e dos serviços de apoio	6 558	457	126	21	7 162	3,62%
		91,57%	6,38%	1,76%	0,29%	100,00%	
		181 587	13 011	2 864	633	198 095	
		91,67%	6,57%	1,45%	0,32%		

Em termos de caracterização das empresas em análise (Tabela 8), ressalta que os principais sectores de actividade, em número de empresas, são a secção G – Comércio por grosso e a retalho; reparação de veículos automóveis e motociclos (33,77% das empresas), secção C – Indústria transformadora (14,69% das empresas) e a secção F – Construção. Ressalta

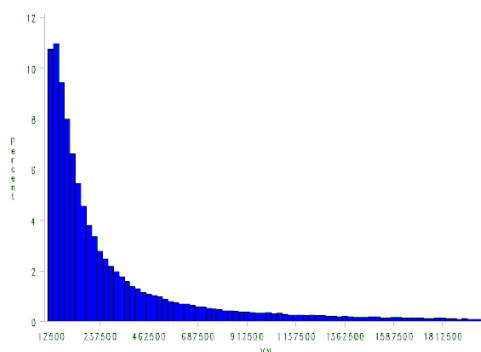
igualmente o peso das empresas de menor dimensão: as micro e pequenas empresas<sup>28</sup> representam 98,23% das empresas, e só as micro-empresas, cujo volume de negócios é inferior a 2 milhões de euros, representam 91,67%. É interessante observar que, na verdade, existe uma concentração de empresas nos volumes de negócios mais baixos em cada escalão de dimensão, o que é particularmente visível nas Figuras 2 a 5. Além disso, as 181 587 empresas incluídas no segmento das micro-empresas apresentam uma distribuição de volume de negócios que interessará captar com maior detalhe, tendo em conta que o limiar de volume de negócios é 2 milhões de euros e a mediana é 144 037 euros (Tabela 9 e Figuras 2 a 5).

**Tabela 9 – Distribuição de empresas por volume de negócios (em Euros)**

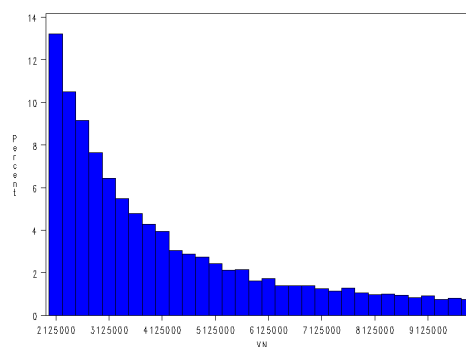
Escalão de VN	N.º empresas	P5	Q1	Mediana	Q3	P95	Média
1 Micro empresas	181 587	12 771	58 331	144 037	365 000	1 204 979	298 520
2 Pequenas empresas	13 011	2 088 080	2 530 000	3 391 757	5 099 391	8 498 830	4 089 449
3 Médias empresas	2 864	10 446 691	12 599 636	16 979 335	25 395 580	41 878 706	20 359 327
4 Grandes empresas	633	52 303 618	64 269 268	89 757 000	167 863 976	717 925 757	237 980 620

<sup>28</sup> A classificação de acordo com a dimensão aqui apresentada baseia-se no disposto pelo artigo 2.º do Anexo à Recomendação da Comissão da Comunidade Europeia 2003/361/CE, de 6 de Maio de 2003, o qual define os efectivos e limiares financeiros que definem as categorias de empresas, dispondo o seguinte: 1. A categoria das micro, pequenas e médias empresas (PME) é constituída por empresas que empregam menos de 250 pessoas e cujo volume de negócios anual não excede 50 milhões de euros ou cujo balanço total anual não excede 43 milhões de euros; 2. Na categoria das PME, uma pequena empresa é definida como uma empresa que emprega menos de 50 pessoas e cujo volume de negócios anual ou balanço total anual não excede 10 milhões de euros; 3. Na categoria das PME, uma microempresa é definida como uma empresa que emprega menos de 10 pessoas e cujo volume de negócios anual ou balanço total anual não excede 2 milhões de euros. A classificação apresentada neste estudo assenta apenas no critério do volume de negócios, porque a base de dados não contém o número de trabalhadores.

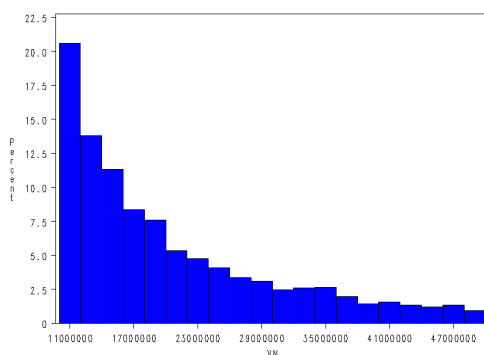
**Figura 2 – Distribuição das micro-empresas por volume de negócios**



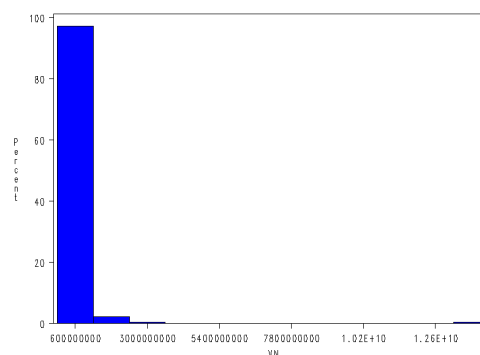
**Figura 3 – Distribuição das pequenas empresas por volume de negócios**



**Figura 4 – Distribuição das médias empresas por volume de negócios**



**Figura 5 – Distribuição das grandes empresas por volume de negócios**



É de alguma forma expectável que seja nos principais sectores de actividade atrás mencionados que se registam as maiores incidências de empresas com sinais de incumprimento em 2009. Os sectores com taxas de incumprimento superiores à taxa média do conjunto analisado (0,619%) são, novamente, a secção C – Indústria transformadora (1,495%) e a secção F – Construção. A taxa de incumprimento da secção G – Comércio por grosso e a retalho; reparação de veículos automóveis e motociclos é 0,565%, ligeiramente inferior à média (Tabela 10).

Tabela 10 – Distribuição do incumprimento por sector de actividade

Secção CAE Rev. 3	Total de empresas		Incumprimento			
	N	%	Sem incump.o	Com incump.	Taxa de incidência	Taxa de incump.
<b>A</b> Agricultura, produção animal, caça, floresta e pesca	5 625	2,84%	5 617	8	0,65%	0,14%
<b>B</b> Indústrias extractivas	638	0,32%	635	3	0,24%	0,47%
<b>C</b> <b>Indústrias transformadoras</b>	<b>29 100</b>	<b>14,69%</b>	<b>28 665</b>	<b>435</b>	<b>35,48%</b>	<b>1,50%</b>
<b>D</b> Electricidade, gás, vapor, água quente e fria e ar frio	308	0,16%	307	1	0,08%	0,33%
<b>E</b> Captação, tratamento e distribuição de água; saneamento, gestão de resíduos	516	0,26%	513	3	0,24%	0,58%
<b>F</b> <b>Construção</b>	<b>26 117</b>	<b>13,18%</b>	<b>25 867</b>	<b>250</b>	<b>20,39%</b>	<b>0,96%</b>
<b>G</b> <b>Comércio por grosso e a retalho; reparação de veículos automóveis e motociclos</b>	<b>66 892</b>	<b>33,77%</b>	<b>66 514</b>	<b>378</b>	<b>30,83%</b>	<b>0,57%</b>
<b>H</b> Transportes e armazenagem	11 756	5,93%	11 716	40	3,26%	0,34%
<b>I</b> Alojamento, restauração e similares	17 723	8,95%	17 692	31	2,53%	0,18%
<b>J</b> Actividades de informação e de comunicação	4 075	2,06%	4 064	11	0,90%	0,27%
<b>L</b> Actividades imobiliárias	9 276	4,68%	9 258	18	1,47%	0,19%
<b>M</b> Actividades de consultoria, científicas, técnicas e similares	18 907	9,54%	18 883	24	1,96%	0,13%
<b>N</b> Actividades administrativas e dos serviços de apoio	7 162	3,62%	7 138	24	1,96%	0,34%
<b>Total</b>	<b>198 095</b>		<b>196 869</b>	<b>1 226</b>		<b>0,62%</b>

Tabela 11 – Distribuição do incumprimento por dimensão da empresa

Escala de VN	Total de empresas		Incumprimento			
	N	%	Sem incump.	Com incump.	Taxa de incidência	Taxa de incumprimento
<b>1</b> Micro empresas	181 587	91,67%	180 597	990	80,75%	0,545%
<b>2</b> Pequenas empresas	13 011	6,57%	12 831	180	14,68%	1,383%
<b>3</b> Médias empresas	2 864	1,45%	2 817	47	3,83%	1,641%
<b>4</b> Grandes empresas	633	0,32%	624	9	0,73%	1,422%
<b>Total</b>	<b>198 095</b>		<b>196 869</b>	<b>1 226</b>		<b>0,619%</b>

Tabela 12 – Distribuição do incumprimento por sector de actividade e dimensão da empresa

Secção CAE Rev. 3	Empresas em incumprimento				Total
	Micro empresas	Pequenas empresas	Médias empresas	Grandes empresas	
A Agricultura, produção animal, caça, floresta e pesca	7 87,50%	1 12,50%	0 0,00%	0 0,00%	8 100,00%
B Indústrias extractivas	1 33,33%	2 66,67%	0 0,00%	0 0,00%	3 100,00%
<b>C Indústrias transformadoras</b>	363 83,45%	60 13,79%	11 2,53%	1 0,23%	435 100,00%
D Electricidade, gás, vapor, água quente e fria e ar frio	0 0	1 100,00%	0 0,00%	0 0,00%	1 100,00%
E Captação, tratamento e distribuição de água; saneamento, gestão de resíduos	2 66,67%	1 33,33%	0 0,00%	0 0,00%	3 100,00%
<b>F Construção</b>	178 71,20%	45 18,00%	20 8,00%	7 2,80%	250 100,00%
<b>G Comércio por grosso e a retalho; reparação de veículos automóveis e motociclos</b>	310 82,01%	54 14,29%	13 3,44%	1 0,26%	378 100,00%
<b>H Transportes e armazenagem</b>	36 90,00%	3 7,50%	1 2,50%	0 0,00%	40 100,00%
<b>I Alojamento, restauração e similares</b>	30 96,77%	1 3,23%	0 0,00%	0 0,00%	31 100,00%
<b>J Actividades de informação e de comunicação</b>	11 100,00%	0 0,00%	0 0,00%	0 0,00%	11 100,00%
<b>L Actividades imobiliárias</b>	12 66,67%	6 33,33%	0 0,00%	0 0,00%	18 100,00%
<b>M Actividades de consultoria, científicas, técnicas e similares</b>	21 87,50%	3 12,50%	0 0,00%	0 0,00%	24 100,00%
<b>N Actividades administrativas e dos serviços de apoio</b>	19 79,17%	3 12,50%	2 8,33%	0 0,00%	24 100,00%
	990 80,75%	180 14,68%	47 3,83%	9 0,73%	

Igualmente expectável é o facto de as empresas com sinais de incumprimento em 2009 serem sobretudo as micro e pequenas empresas, uma vez que representam a maior parte das empresas em análise (Tabelas 11 e 12). Contudo, verifica-se que, em termos de taxas de incumprimento, é nas médias empresas que essa taxa é maior. Ainda que a taxa de incumprimento das grandes empresas seja claramente superior à média, só 9 empresas apresentaram sinais de incumprimento em 2009, pelo que este segmento não será considerado para efeitos da análise que se seguirá.

Na sequência dos gráficos de distribuição do volume de negócios por escalão de dimensão das empresas (Figuras 2 a 5), proceder-se-á a uma categorização da variável volume de negócios, tendo em conta os respectivos quartis (apresentada na Tabela 13).

De forma similar, existem secções da CAE Rev. 3 com poucos casos de incumprimento, pelo que não deverão ser considerados para efeitos da análise que se seguirá. Estão nesta situação a secção A – Agricultura, produção animal, caça, floresta e pesca (8 empresas), B – Indústrias extractivas (3 empresas), D – Electricidade, gás, vapor, água quente e fria e ar frio (1 empresa) E – Captação, tratamento e distribuição de água; saneamento, gestão de resíduos.

Por outro lado, afigura-se interessante detalhar os incumprimentos, em termos de níveis inferiores da CAE Rev. 3, nas secções com um maior número de casos de empresas em incumprimento: secção C – Indústria transformadora (435 empresas), F – Construção (250 empresas) e G – Comércio por grosso e a retalho; reparação de veículos automóveis e motociclos (378 empresas).

Assim, a título de resumo, e perante a análise preliminar dos dados, serão tomadas as seguintes opções antes de se prosseguir na análise (as Tabelas 13 e 14 resumem essa informação):

- serão consideradas apenas as micro, pequenas e médias empresas, procedendo-se à discretização/categorização da variável volume de negócios no segmento das micro-empresas, de acordo com os respectivos quartis;
- as secções da CAE Rev. 3 a considerar são: C – Indústria transformadora, F – Construção, G – Comércio por grosso e a retalho; reparação de veículos automóveis e motociclos, H – Transportes e armazenagem, I – Alojamento, restauração e similares, J – Actividades de informação e comunicação, L – Actividades imobiliárias, M – Actividades de consultoria, científicas, técnicas e similares e N – Actividades administrativas e dos serviços de apoio;
- em relação às secções C, F e G serão considerados níveis inferiores da CAE Rev.3.



**Tabela 13 – Distribuição do incumprimento por novos escalões de volume de negócios (em Euros)**

		Total	Número de empresas		Taxa de incidência	Taxa de incumprimento
			Sem incump.	Com incump.		
1	VN<58 331 euros (Micro-empresas)	43 633	43 501	132	10,98%	0,30%
2	VN entre 58 331 e 144 037 euros (Micro-empresas)	43 900	43 685	215	17,89%	0,49%
3	VN entre 144 037 e 365 000 euros (Micro-empresas)	43 821	43 558	263	21,88%	0,60%
4	VN entre 365 000 e 2 milhões de euros (Micro-empresas)	43 769	43 399	370	30,78%	0,85%
5	VN entre 2 milhões e 10 milhões de euros (Pequenas empresas)	12 537	12 362	175	14,56%	1,40%
6	VN entre 10 milhões e 50 milhões de euros (Médias empresas)	2 741	2 694	47	3,91%	1,71%
		190 401	189 199	1 202		0,63%

**Tabela 14 – Distribuição dos incumprimentos por sector de actividade (agrupamentos considerados)**

CAE Rev. 3			Total	Número de empresas		Taxa de incidência	Taxa de incumprimento
				Sem incump.	Com incump.		
1	C10, C11 e C12	Ind. alimentares, bebidas e tabaco	3 921	3 904	17	1,41%	0,43%
2	C13	Fabricação de têxteis	1 539	1 500	39	3,24%	2,53%
3	C14	Indústria do vestuário	3 367	3 227	140	11,65%	4,16%
4	C15 e C16	Ind. couro, madeira e cortiça	3 609	3 537	72	5,99%	2,00%
5	C17	Fabricação de pasta, de papel, de cartão e seus artigos	323	315	8	0,67%	2,48%
6	C19, C20, C21 e C22	Fabricação de combustíveis e produtos petrolíferos refinados e indústria química	1 343	1 327	16	1,33%	1,19%
7	C23	Fabrico de outros produtos minerais não metálicos	2 083	2 057	26	2,16%	1,25%
8	C24 e C25	Indústrias metalúrgicas	5 246	5 205	41	3,41%	0,78%
9	C26, C27, C28 e C33	Máquinas e equipamentos	2 573	2 555	18	1,50%	0,70%
10	C29 e C30	Fabricação de material de transporte	470	458	12	1,00%	2,55%
11	C18, C31 e C32	Outras indústrias transformadoras	4 435	4 390	45	3,74%	1,01%
12	F41	Promoção imobiliária (desenvolvimento de	14 231	14 106	125	10,40%	0,88%

CAE Rev. 3	Total	Número de empresas		Taxa de incidência	Taxa de incumprimento
		Sem incump.	Com incump.		
13 F42	2 367	2 323	44	3,66%	1,86%
14 F43	9 450	9 376	74	6,16%	0,78%
15 G45	9 701	9 667	34	2,83%	0,35%
16 G46	24 894	24 170	184	15,31%	0,74%
17 G47	32 073	31 914	159	13,23%	0,50%
18 H	11 717	11 677	40	3,33%	0,34%
19 I	17 716	17 685	31	2,58%	0,17%
20 J	4 048	4 037	11	0,92%	0,27%
21 L	9 271	9 253	18	1,50%	0,19%
22 M	18 883	18 859	24	2,00%	0,13%
23 N	7 141	7 117	24	2,00%	0,34%
	190 401	189 199	1 202		0,63%

#### 4.4. Distribuição das potenciais variáveis explicativas na amostra final

A título prévio, refira-se que as questões em torno da amostragem em modelos multinível não parece reunir consenso, pondo-se a questão em termos do número de grupos a considerar e do número de observações por grupo.

Gelman e Hill (2007) afirmam que a ideia de que só se podem utilizar modelos multinível quando o número de grupos é superior a determinado valor ou quando existe um determinado número de observações por grupo podem ser falaciosas, embora seja chamada a atenção para a dificuldade de estimação dos parâmetros de variância em amostras reduzidas. Aliás, com apenas um ou dois grupos, a modelação multinível reduz-se a uma regressão clássica (Gelman e Hill, 2007, pp. 275). Os autores afirmam que duas ou mesmo

uma observação por grupo é suficiente para estimar um modelo multinível, ainda que com menores níveis de precisão. Também Guo e Zhao (2000) referem estudos em que a proporção de grupos com uma única observação era 24% ou 32%.

Raudenbush e Bryk (2002) referem que uma regra comum na análise de regressão clássica é a existência de 10 observações para cada variável explicativa, embora seja difícil encontrar uma regra semelhante aplicável a modelos hierárquicos. Kreft e De Leeuw (1998) concluem que existem tantos factores envolvidos nas análises multinível que é impossível produzir “*rules of thumb*” para a amostragem.

Snijders e Bosker (1999) referem que a dimensão amostral no nível 2 é normalmente o elemento mais restritivo do desenho da amostra e que os requisitos em relação à dimensão da amostra no nível mais elevado, para um modelo hierárquico com um determinado número de variáveis explicativas nesse nível, devem ser pelo menos tão rigorosos como os requisitos para a dimensão da amostra num desenho de apenas um nível com o mesmo número de variáveis explicativas (contudo, no presente estudo não se consideram variáveis explicativas no nível superior).

Segundo Hensher e Jones (2007), a influência da metodologia de amostragem escolhida sobre as estimativas dos parâmetros e as previsões tem merecido particular atenção nos casos de estudo do incumprimento de empresas. Parecem existir duas grandes abordagens: uma que consiste em criar amostras emparelhadas (emparelhar cada empresa em incumprimento com uma empresa “saudável”) e outra em que as amostras são construídas de modo a que as taxas de incumprimento se aproximem das taxas de incumprimento da população. Os autores referem o trabalho de Zmijeswski (1984) que permitiu concluir que a utilização de amostras emparelhadas, ou de amostras que não reflectem, de forma aproximada, a composição da população podem resultar em estimativas de parâmetros e de probabilidades assintoticamente enviesadas.

No presente caso, a constituição da amostra esteve muito dependente de considerações práticas, nomeadamente dos dados disponíveis. Procurou-se aproveitar ao máximo a informação completa disponível, tendo-se constituído uma amostra emparelhada considerando-se as 1 202 empresas em incumprimento em 2009 e seleccionando aleatoriamente o mesmo número de empresas de entre o grupo de empresas saudáveis,

considerando o sector de actividade, variável que define o segundo nível de análise, e a dimensão (medida pelo escalão de volume de negócios).

A não consideração de uma amostra de validação foi uma opção consciente. Pelos motivos apontados como justificativos por se considerar apenas o ano de 2009, não se considerou conveniente criar uma amostra de validação com incumprimentos de anos anteriores. A partição da amostra criada a partir dos incumprimentos de 2009 também levantaria problemas, porque alguns grupos ficariam com um número de observações muito reduzido. Na Tabela 15 incluem-se as variáveis consideradas inicialmente no processo de modelação, para os quais se apresentam estatísticas descritivas (de localização e de dispersão).

**Tabela 15 – Estatísticas descritivas das variáveis potencialmente explicativas na amostra final**

Dimensão de análise	Variável		Média	Desvio-padrão	Q1	Mediana	Q3
Antiguidade	Antiguidade (meses)	Sem incump	183,62	169,19	69,33	137,1	244
		Com incump	194,67	172,79	73,93	151,18	262,7
Dimensão	In_VN	Sem incump	12,85	1,72	11,7	12,78	13,94
		Com incump	12,9	1,71	11,72	12,76	14,04
Actividade	AT_VN	Sem incump	5,94	102,64	0,53	0,86	1,51
		Com incump	18,29	336,39	0,67	1,14	2,09
Liquidez	AC_AT	Sem incump	0,72	0,25	0,57	0,8	0,93
		Com incump	0,74	0,24	0,59	0,81	0,94
Liquidez	CD_AT	Sem incump	0,14	0,19	0,01	0,06	0,19
		Com incump	0,08	0,15	0,01	0,02	0,09
Liquidez	CD_PCP	Sem incump	0,88	6,11	0,02	0,11	0,46
		Com incump	0,18	1,7	0,01	0,03	0,09
Liquidez	FM_AT	Sem incump	0,06	1,15	-0,07	0,18	0,44
		Com incump	-0,58	4,54	-0,48	-0,06	0,17
Liquidez	PCP_PT_PRO V	Sem incump	0,77	0,27	0,64	0,89	0,99
		Com incump	0,79	0,25	0,64	0,88	0,99
Solvência	CP_PT	Sem incump	1,93	17,74	0,08	0,32	0,92
		Com incump	0,06	2,46	-0,28	0,01	0,19
Estrutura	DDT_PROV_ AT_IM	Sem incump	0,87	26,81	0,01	0,50	1,00
		Com incump	0,65	56,73	0,07	0,66	1,07
Estrutura de custos	EBIT_JCS	Sem incump	43,11	934,76	0,42	1,82	6,78
		Com incump	-47,51	568,78	-7,24	-1,31	1,08
Estrutura de	JCS_DAT	Sem incump	0,04	0,07	0,01	0,03	0,05

Dimensão de análise	Variável		Média	Desvio-padrão	Q1	Mediana	Q3
custos		Com incump	0,05	0,07	0,02	0,04	0,07
Rentabilidade	RL_CP	Sem incump	0,06	6,28	0,01	0,07	0,26
		Com incump	1,04	29,00	-0,12	0,12	0,80
Rentabilidade	RL_VN	Sem incump	-0,05	2,40	-0,02	0,01	0,04
		Com incump	-0,61	5,72	-0,40	-0,13	0
Rentabilidade	RO_AT	Sem incump	0,01	0,25	-0,01	0,04	0,08
		Com incump	-0,26	1,23	-0,25	-0,06	0,03
Rentabilidade	VN_AT	Sem incump	1,61	1,92	0,66	1,16	1,90
		Com incump	1,28	1,83	0,48	0,88	1,49

A análise dos valores apresentados<sup>29</sup> permite retirar desde já algumas ideias sobre o que se espera que possa caracterizar o incumprimento.

As empresas em incumprimento aparentam ser ligeiramente mais antigas. Em relação à dimensão, ainda que os valores médios e medianos do logaritmo do volume de negócios sejam muito semelhantes, os valores dos quartis são ligeiramente superiores no caso das empresas em incumprimento.

A dimensão *actividade*, medida pelo rácio activo total/volume de negócios apresenta valores superiores nas empresas em incumprimento, o que pode indiciar alguma capacidade não aproveitada (se se pensar num activo elevado com um reduzido volume de negócios).

A dimensão *liquidez* é medida por cinco rácios financeiros e a informação por eles prestada aponta aproximadamente no mesmo sentido. Em termos médios, o peso do activo circulante no activo total (AC\_AT) e o peso do passivo de curto prazo sobre o passivo total (PCP\_PT\_PROV) não parece muito distinto entre as empresas sem incumprimento e as empresas com incumprimento. Contudo, os valores das variáveis caixa e depósitos/activo total (CD\_AT), caixa e depósitos/passivo de curto prazo (CD\_PCP) e fundo de maneo/activo total (FM\_AT) são superiores nas empresas sem incumprimento, que aparentam assim ter uma situação de liquidez mais robusta do que as empresas em incumprimento.

Em termos de *solvência*, medida pelo rácio capital próprio/passivo total (CP\_PT), esta é, em termos médios, bastante superior nas empresas sem incumprimento.

<sup>29</sup> No anexo 4 apresenta-se uma tabela com os valores médios das variáveis potencialmente explicativas na amostra final (antiguidade, dimensão, actividade e liquidez, estrutura, estrutura de custos, rentabilidade).

Os valores médios e medianos do rácio dívidas de terceiros líquidas de provisões/activo total menos imobilizado total líquido (DDT\_PROV\_AT\_IM) não são muito indicativos, embora observando a distribuição por quartis seja possível verificar que os valores deste rácio são tendencialmente superiores nas empresas com incumprimento.

A dimensão *estrutura de custos* é medida por dois rácios, o resultado líquido antes de juros e impostos/juros e custos similares (EBIT\_JCS) e o peso dos juros e custos similares nas dívidas a terceiros (JCS\_DAT). No caso do primeiro indicador, os valores são bastante superiores nas empresas sem incumprimento, significando que os resultados da actividade suportam devidamente os custos de financiamento necessários para levar a cabo a actividade, enquanto que nas empresas em incumprimento se identificam valores negativos, o que corresponderá a resultados também negativos. Verifica-se também que, em termos médios, o peso dos encargos financeiros é, ainda que ligeiramente, superior nas empresas com incumprimento.

Finalmente, os rácios que materializam a dimensão *rentabilidade* merecem uma análise individualizada. Os valores do rácio resultado líquido/capital próprio (RL\_CP) parecem ser, em média, superiores nas empresas com incumprimento, o que não significa que sejam mais rentáveis. Na verdade, o rácio pode vir inflacionado por um denominador (capital próprio) que pode ser inferior nestas empresas. Aliás, os valores do rácio que materializa a solvência (capital próprio/passivo total) parecem indicar isso mesmo. Em relação ao rácio resultado líquido/volume de negócios (RL\_VN), as empresas com incumprimento apresentam valores inferiores, verificando-se que tanto a média como a mediana indicia, nesses casos, resultados líquidos negativos. A informação prestada pelo rácio resultado operacional sobre activo total (RO\_AT) vai no mesmo sentido, desta feita com os problemas a situarem-se nos resultados operacionais, ou seja, com os proveitos directamente provenientes das actividades principais das empresas. O rácio volume de negócios sobre activo total (VN\_AT) indicia uma maior rentabilidade nas empresas sem incumprimento.

## 5. Resultados empíricos

### 5.1. Regressão logística

Estimaram-se, de acordo com os mesmos procedimentos, dois modelos de regressão logística: um não centrando as variáveis explicativas e o outro considerando as variáveis centradas, tendo-se retirado a cada uma delas a média do respectivo agrupamento de sector de actividade em que a empresa se insere<sup>30</sup>. Em termos de ajustamento global do modelo, os resultados obtidos foram idênticos, identificando-se, como esperado, diferenças nas estimativas do termo independente e também dos coeficientes associados aos sectores de actividade, os quais, quando considerados, resultam na alteração do termo independente<sup>31</sup>. A centragem teve, assim, consequências ao nível dos ensaios de hipóteses relativos à significância de alguns sectores de actividade.

Recorde-se que as variáveis consideradas constam na Tabela 15 e os 23 agrupamentos de sectores de actividade apresentados constam na Tabela 14. Começa-se por apresentar os resultados do ajustamento global do modelo. Recorde-se que as variáveis explicativas potenciais foram testadas com recurso a selecção *stepwise*, conforme referido na Secção 3.2. O sector de actividade considerado como referência (Grupo 23 - Actividades administrativas e dos serviços de apoio) foi seleccionado por defeito.

O coeficiente de determinação generalizado de Cox e Snell é  $R_{CS}^2 = 0,29$  e o coeficiente ajustado de Nagelkerke é  $R_N^2 = 0,39$ . No teste de Hosmer e Lemeshow (Tabela 16) rejeita-se a hipótese de um bom ajustamento dos modelos.

**Tabela 16 – Ajustamento global do modelo (Teste de Hosmer e Lemeshow)**

Teste de Hosmer e Lemeshow		
Qui-quadrado	Graus de liberdade	Pr > ChiSq
25,398	8	0,001

Estes resultados levantam algumas reservas à qualidade global dos modelos. Contudo, os valores da tabela de classificação são bastante razoáveis (Tabela 17), com 75,5% das

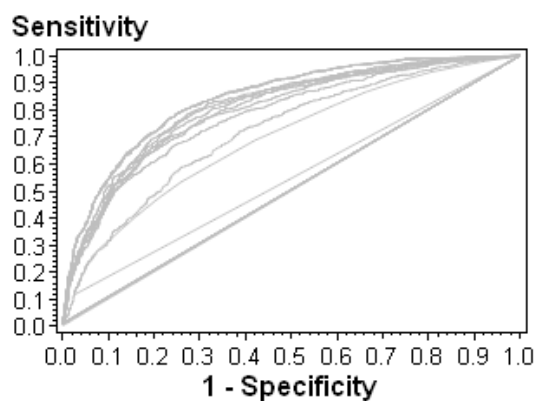
<sup>30</sup> Cf. discutido na Secção 2.5.

<sup>31</sup> Nos Anexos 5 e 6 encontram-se os *outputs* relativos à estimação dos modelos.

empresas a serem correctamente classificadas quanto à situação de incumprimento, com os erros de tipo I (falsos positivos) e de tipo II (falsos negativos) bastante próximos, o que se reflecte numa configuração da curva *Receiver Operating Characteristic* (ROC) (Figura 6) e na área entre a curva e a diagonal (0,84)<sup>32</sup> que indicam uma satisfatória capacidade discriminante do modelo estimado.

Tabela 17 – Matriz de classificação

		Estimado	
		Com incumprimento	Sem incumprimento
Observado	Com incumprimento	924 (76,9%)	278 (23,8%)
	Sem incumprimento	310 (25,1%)	892 (74,2%)

Figura 6 – Curva ROC<sup>33</sup>

<sup>32</sup> Aplicando a relação  $G + 1 = 2AUC$ , então o coeficiente de Gini é 0,68, o que indica uma razoável capacidade de distinção das empresas em incumprimento das empresas sem incumprimento.

<sup>33</sup> A área abaixo da curva é dada pelo valor da estatística  $c$  do quadro *Association of Predicted Probabilities and Observed Responses* (0,84) (cf. Anexo 5).



Tabela 18 – Resumo dos modelos estimados (regressão logística)

Variável	Variáveis não centradas			Variáveis centradas		
	Coefficiente estimado	Erro-padrão	P-value	Coefficiente estimado	Erro-padrão	P-value
<b>Intercept</b>	-1,354	0,474	0,0043	-0,120	0,103	0,2410
<b>AC_AT</b> Activo circulante/Activo total (Liquidez)	0,948	0,227	<0,0001	0,948	0,227	<0,0001
<b>CD_AT</b> Caixa e depósitos/Activo total (Liquidez)	-0,861	0,349	0,0137	-0,861	0,349	0,0137
<b>FM_AT</b> Fundo de Maneio/Activo total (Liquidez)	-0,233	0,069	0,0008	-0,232	0,069	0,0008
<b>CP_PT</b> Capital próprio/Passivo total (Solvência)	-0,629	0,082	<0,0001	-0,629	0,082	<0,0001
<b>JCS_DAT</b> Juros e custos similares/Dívidas a terceiros (Estrutura de custos)	8,368	1,066	<0,0001	8,368	1,066	<0,0001
<b>RO_AT</b> Resultados operacionais/Activo total (Rentabilidade)	-2,413	0,254	<0,0001	-2,413	0,254	<0,0001
<b>VN_AT</b> Volume de negócios/Activo total (Rentabilidade)	-0,425	0,049	<0,0001	-0,425	0,049	<0,0001
<b>ln_VN</b> Volume de negócios (ln) (Dimensão)	0,080	0,033	0,0155	0,080	0,033	0,0155
<b>Grupo</b> 1 - Indústrias alimentares, bebidas e tabaco	-0,441	0,336	0,1891	-1,030	0,334	0,0020
<b>Grupo</b> 2 - Fabricação de têxteis	1,083	0,404	0,0073	2,059	0,408	<0,0001
<b>Grupo</b> 3 - Indústria do vestuário	1,655	0,232	<0,0001	2,160	0,240	<0,0001
<b>Grupo</b> 4 - Indústria de couro, madeira e cortiça	1,129	0,282	<0,0001	2,102	0,298	<0,0001
<b>Grupo</b> 5 - Fabricação de pasta, de papel, de cartão e seus artigos	0,847	0,799	0,2888	3,587	0,819	<0,0001
<b>Grupo</b> 6 - Fabricação de combustíveis e produtos petrolíferos refinados e Indústria Química	-0,001	0,419	0,9979	0,569	0,420	0,1760
<b>Grupo</b> 7 - Fabrico de outros produtos minerais não metálicos	0,285	0,341	0,4028	0,841	0,348	0,0157
<b>Grupo</b> 8 - Indústrias metalúrgicas	0,175	0,261	0,5039	0,536	0,270	0,0471
<b>Grupo</b> 9 - Máquinas e equipamentos	-0,416	0,330	0,2073	-0,388	0,333	0,2442
<b>Grupo</b> 10 - Fabricação de material de transporte	0,707	0,623	0,2566	1,321	0,624	0,0342
<b>Grupo</b> 11 - Outras Indústrias Transformadoras	0,043	0,250	0,8643	0,504	0,256	0,0487
<b>Grupo</b> 12 - Promoção imobiliária (desenvolvimento de projectos de edifícios); construção de edifícios	0,274	0,171	0,1100	0,511	0,170	0,0026

Variável	Variáveis não centradas			Variáveis centradas			
		Coefficiente estimado	Erro-padrão	P-value	Coefficiente estimado	Erro-padrão	P-value
<b>Grupo</b> 13- Engenharia Civil		0,807	0,313	0,0099	1,142	0,315	0,0003
<b>Grupo</b> 14 - Actividades especializadas de construção		0,153	0,201	0,4469	0,667	0,206	0,0012
<b>Grupo</b> 15 - Comércio, manutenção e reparação, de veículos automóveis e motociclos		-0,778	0,231	0,0008	-0,708	0,233	0,0023
<b>Grupo</b> 16 - Comércio por grosso (inclui agentes), excepto de veículos automóveis e motociclos		-0,243	0,136	0,0730	0,339	0,142	0,0172
<b>Grupo</b> 17 - Comércio a retalho, excepto de veículos automóveis e motociclos		-0,286	0,137	0,0363	-0,837	0,140	<0,0001
<b>Grupo</b> 18 - Transportes e armazenagem		-0,119	0,242	0,6219	-2,141	0,325	<0,0001
<b>Grupo</b> 19 - Alojamento, restauração e similares		-1,145	0,263	<0,0001	-1,750	0,254	<0,0001
<b>Grupo</b> 20 - Actividades de informação e de comunicação		-0,762	0,413	0,0650	-0,797	0,414	0,0544
<b>Grupo</b> 21 - Actividades imobiliárias		-1,379	0,300	<0,0001	-6,100	0,656	<0,0001
<b>Grupo</b> 22 - Actividades de consultoria, científicas, técnicas e similares		-1,126	0,265	<0,0001	-1,844	0,267	<0,0001

Partindo da Tabela 18, verifica-se que as variáveis cuja variação tem um impacto no mesmo sentido na probabilidade de incumprimento, ou seja, quanto maiores (menores), maior (menor) a probabilidade de incumprimento (para um nível de significância de 0,05), são as seguintes:

- A variável de liquidez *Activo circulante/Activo Total* (AC\_AT);
- A variável de estrutura de custos *Juros e custos similares/Dívidas a terceiros* (JCS\_DAT);
- A variável de dimensão *Logaritmo natural do volume de negócios* (ln\_VN).

As variáveis cuja variação tem um impacto no sentido oposto na probabilidade de incumprimento, ou seja, quanto maiores (menores), menor (maior) a probabilidade de incumprimento, são as seguintes:

- As variáveis de liquidez *Caixa e depósitos/Activo total* (CD\_AT) e *Fundo de maneo/Activo total* (FM\_AT);

- A variável de solvência *Capital próprio/Passivo total* (CP\_PT);
- As variáveis de rentabilidade *Resultados operacionais/Activo total* (RO\_AT) e *Volume de negócios/Activo total* (VN\_AT).

Considerando os resultados obtidos com as variáveis centradas, os agrupamentos de sectores de actividade cuja pertença apresenta evidência de ser mais problemática em termos de incumprimento são, fundamentalmente, sectores da indústria transformadora e construção: Grupo 2 - Fabricação de têxteis, Grupo 3 - Indústria do vestuário, Grupo 4 - Indústria de couro, madeira e cortiça, Grupo 5 - Fabricação de pasta, de papel, de cartão e seus artigos, Grupo 7 - Fabrico de outros produtos minerais não metálicos, Grupo 8 - Indústrias metalúrgicas, Grupo 10 - Fabricação de material de transporte, Grupo 11 - Outras Indústrias transformadoras, Grupo 12 - Promoção imobiliária (desenvolvimento de projectos de edifícios); construção de edifícios, Grupo 13- Engenharia civil, Grupo 14 - Actividades especializadas de construção e Grupo 16 - Comércio por grosso (inclui agentes), excepto de veículos automóveis e motociclos.

Considerando as variáveis não centradas, os agrupamentos de sectores de actividade cuja pertença apresenta evidências de ser mais problemática em termos de incumprimento são: Grupo 2 – Fabricação de têxteis, Grupo 3 - Indústria do vestuário, Grupo 4 - Indústria de couro, madeira e cortiça e Grupo 13- Engenharia civil.

## **5.2. Modelo multinível com função de ligação *logit***

Recorde-se que, após a discussão da questão da centragem das variáveis, na Secção 2.5, se apresentou a opção simplificada de não centrar as variáveis com a justificação de que o objectivo do presente estudo se prende com os resultados ao nível individual, ou seja, com o incumprimento das empresas, sem que haja um interesse particular em variáveis de segundo nível. Apenas se pretende testar a existência de partilha de características não observadas entre empresas pertencentes ao mesmo sector de actividade.

Em primeiro lugar, estimou-se<sup>34</sup> um modelo apenas com termo independente que prevê a probabilidade de incumprimento, na forma multinível (cf. equação 21),  $\text{logit}(p_{ij}) = \gamma_{00} + u_{0j}$ , com  $u_{0j} \sim N(0, \sigma_{00}^2)$ .

Recorde-se que  $\gamma_{00}$  corresponde ao efeito fixo e  $u_{0j}$  ao efeito aleatório, neste caso sujeito aos sectores de actividade económica definidos pela variável *Grupo*, pelo que apenas vamos ter um parâmetro de variância.

**Tabela 19 – Estatísticas de ajustamento (modelo multinível apenas com termo independente)**

<b>-2 Res Log Pseudo-Likelihood</b>	10470,28
<b>Pseudo-AIC</b>	10472,28
<b>Qui-quadrado generalizado</b>	2376,26
<b>Qui-quadrado gener. / Graus de liberdade</b>	0,99

O rácio *Generalized Chi-square/DF* apresentado na Tabela 19 dá a estimativa do parâmetro de dispersão  $\phi$ , discutido nas Secções 2.4 e 3.3. Sendo este parâmetro praticamente 1, a variabilidade dos dados foi propriamente modelada (dados binários).

**Tabela 20 – Estimativa da variância do termo independente aleatório (modelo multinível apenas com termo independente)**

	Sujeito a	Estimativa	Erro-padrão
<b>Termo independente</b>	<b>Grupo</b>	0,657	0,224

A Tabela 20 contém a estimativa da variância do termo independente aleatório  $u_{0j}$  (medida na escala *logit*). A hipótese alternativa é a variância ser maior do que zero, *i.e.*,  $H_1 : \sigma_{00}^2 > 0$ . Conforme discutido na Secção 3.3, efectuando um teste de Wald, a estatística de teste é dada por:

$$\frac{\hat{\sigma}_{00}^2 - 0}{S.E.(\hat{\sigma}_{00}^2)} = \frac{0,657 - 0}{0,224} = 2,934$$

Comparando o valor da estatística de teste (unilateral) com a distribuição Normal-Padrão, rejeita-se a hipótese nula, ou seja, existe evidência de que a variância é significativamente positiva (com um nível de significância de 0,05). Existem evidências estatísticas de que se deve considerar a existência de efeitos aleatórios ao nível dos sectores de actividade.

<sup>34</sup> No Anexo 7 encontram-se os *outputs* relativos à estimação.

Recordando a equação (8), a correlação intra-classe é dada por  $\rho = \sigma_{00}^2 / (\sigma_{00}^2 + \sigma^2)$ , sendo  $\sigma^2 = \pi^2 / 3$  a variância da distribuição logística. Assim, a estimativa dessa correlação (na escala *logit*) é  $\rho = 0,657 / (0,657 + \pi^2 / 3) = 0,166$ .

As Tabelas 21 e 22 apresentam as estimativas do efeito fixo e as soluções para os efeitos aleatórios.

**Tabela 21 – Efeitos fixos (modelo multinível apenas com termo independente)**

	Estimativa	Erro-padrão	G.L.	Valor t	Pr >  t
<b>Termo independente</b>	0,08471	0,1793	21,35	0,47	0,6413

**Tabela 22 – Efeitos aleatórios (modelo apenas com termo independente)**

	Sujeito a		Estimativa	Erro-Padrão	G.L.	Valor t	Pr >  t
Termo independente	Grupo 1	Indústrias alimentares, bebidas e tabaco	-0,603	0,321	129,4	-1,88	0,0630
Termo independente	Grupo 2	Fabricação de têxteis	1,157	0,353	141,5	3,28	0,0013
Termo independente	Grupo 3	Indústria do vestuário	1,373	0,255	75,96	5,39	<0,0001
Termo independente	Grupo 4	Indústria de couro, madeira e cortiça	1,136	0,290	106,70	3,92	0,0002
Termo independente	Grupo 5	Fabricação de pasta, de papel, de cartão e seus artigos	0,718	0,533	79,03	1,35	0,1817
Termo independente	Grupo 6	Fabricação de combustíveis e produtos petrolíferos refinados e Indústria Química	0,235	0,380	142,90	0,62	0,5371
Termo independente	Grupo 7	Fabrico de outros produtos minerais não metálicos	0,459	0,341	138,20	1,35	0,1808
Termo independente	Grupo 8	Indústrias metalúrgicas	0,209	0,283	101,20	0,74	0,4613
Termo independente	Grupo 9	Máquinas e equipamentos	-0,396	0,326	131,80	-1,22	0,2265
Termo independente	Grupo 10	Fabricação de material de transporte	0,689	0,465	112,00	1,48	0,1411
Termo independente	Grupo 11	Outras Indústrias Transformadoras	0,209	0,276	94,88	0,76	0,4510
Termo independente	Grupo 12	Promoção imobiliária (desenvolvimento de projectos de edifícios); construção de edifícios	0,327	0,223	48,91	1,46	0,1495
Termo independente	Grupo 13	Engenharia civil	0,772	0,309	121,80	2,50	0,0139
Termo independente	Grupo 14	Actividades especializadas de construção	0,104	0,241	63,46	0,43	0,6681
Termo independente	Grupo 15	Comércio, manutenção e reparação, de veículos	-0,771	0,262	82,09	-2,94	0,0042

Sujeito a		Estimativa	Erro-Padrão	G.L.	Valor t	Pr >  t	
Termo independente	Grupo 16	automóveis e motocicletas Comércio por grosso (inclui agentes), excepto de veículos automóveis e motocicletas	-0,130	0,204	35,12	-0,64	0,5267
Termo independente	Grupo 17	Comércio a retalho, excepto de veículos automóveis e motocicletas	-0,371	0,205	35,62	-1,81	0,0784
Termo independente	Grupo 18	Transportes e armazenagem	-0,395	0,262	81,99	-1,51	0,1344
Termo independente	Grupo 19	Alojamento, restauração e similares	-1,059	0,261	81,37	-4,06	0,0001
Termo independente	Grupo 20	Actividades de informação e de comunicação	-0,646	0,363	143,00	-1,78	0,0769
Termo independente	Grupo 21	Actividades imobiliárias	-1,084	0,299	114,10	-3,63	0,0004
Termo independente	Grupo 22	Actividades de consultoria, científicas, técnicas e similares	-1,326	0,272	91,24	-4,88	<0,0001
Termo independente	Grupo 23	Actividades administrativas e dos serviços de apoio	-0,608	0,291	108,30	-2,09	0,0392

Tendo-se rejeitado o termo independente fixo como estatisticamente significativo, o resultado da estimação é apenas, para cada sector de actividade, o termo independente aleatório, quando estatisticamente significativo. A título de exemplo, para o Grupo 2 – Fabricação de Têxteis, tem-se que  $\text{logit}(\hat{p}_{iG2}) = 1,157$ , ou seja, uma empresa que se dedique à Fabricação de Têxteis terá como probabilidade de incumprimento  $\frac{e^{1,157}}{1 + e^{1,157}} = 0,761$ . Para o Grupo 22 – Actividade de consultoria, científicas, técnicas e similares, tem-se que  $\text{logit}(\hat{p}_{iG22}) = -1,326$ , ou seja, uma empresa pertencente a esse sector de actividade terá como probabilidade de incumprimento  $\frac{e^{-1,326}}{1 + e^{-1,326}} = 0,210$ . Recordemos que estamos a usar amostras com igual dimensão de não incumprimento e incumprimento. Na Tabela 23 apresenta-se um resumo das soluções dos efeitos aleatórios estatisticamente significativos em termos de probabilidades estimadas.

**Tabela 23 – Resumo das soluções dos efeitos aleatórios estatisticamente significativos (modelo multinível apenas com termo independente)**

Grupo	Secção/divisão da CAE Rev.3	Descrição	Logit ( $\hat{p}_i$ )	$\hat{p}_i$
Grupo 2	C13	Fabricação de têxteis	1,1567	0,7607
Grupo 3	C14	Indústria do vestuário	1,3733	0,7979
Grupo 4	C15 e C16	Indústria de couro, madeira e cortiça	1,1361	0,7570
Grupo 13	F42	Engenharia Civil	0,7723	0,6840
Grupo 15	G45	Comércio, manutenção e reparação, de veículos automóveis e motociclos	-0,7705	0,3164
Grupo 19	I	Alojamento, restauração e similares	-1,0589	0,2575
Grupo 21	L	Actividades imobiliárias	-1,0836	0,2528
Grupo 22	M	Actividades de consultoria, científicas, técnicas e similares	-1,3256	0,2099
Grupo 23	N	Actividades administrativas e dos serviços de apoio	-0,6082	0,3525

Depois da estimação do modelo apenas com termo independente, testou-se a presença de efeitos aleatórios nas variáveis explicativas que resultaram estatisticamente significativas na regressão logística, o que corresponde à equação (14):

$$\text{logit}(p_{ij}) = \gamma_{00} + \sum_p \gamma_{p0} X_{pij} + u_{0j} + \sum_p u_{pj} X_{pij} .$$

Contudo, o modelo não convergiu<sup>35</sup>, pelo que se passou a testar as variáveis uma a uma. Apenas se obteve resultados considerando a variável *juros e custos similares/dívidas a terceiros* (JCS\_DAT)<sup>36</sup>, relativa à dimensão estrutura de custos, tendo este modelo valores mais elevados das estatísticas *-2 Res Log Pseudo-Likelihood* e *Pseudo-AIC* do que o modelo apenas com termo independente, indiciando um ajustamento mais fraco. Assim, para cada sector de actividade *j* tem-se:

$$\text{logit}(p_i) = \gamma_{00} + \gamma_{10} X_{1ij} + u_{0j} + u_{1j} X_{1ij} \quad (22)$$

Os efeitos fixos são  $\gamma_{00}$  e  $\gamma_{01}$  e os efeitos aleatórios são  $u_{j0}$  e  $u_{ij}$ . Os principais resultados são apresentados de seguida.

**Tabela 24 – Estimativas das variâncias dos efeitos aleatórios (modelo multinível com variável JCS\_DAT)**

	Sujeitos a	Estimativa	Erro-padrão
<b>Termo independente</b>	<b>Grupo</b>	0,639	0,228
<b>JCS_DAT</b>	<b>Grupo</b>	28,269	19,219

<sup>35</sup> Cf. Anexo 8, onde se encontra o *output* resultante desta estimação.

<sup>36</sup> Cf. Anexo 9.

Da mesma forma como se fez para o modelo apenas com termo independente, podem realizar-se testes estatísticos para verificar se as variâncias estimadas são significativamente diferentes de zero.

Assim, para a variância do efeito aleatório  $u_{0j}$ , o ensaio de hipóteses é o seguinte:

$$H_0 : \sigma_{00}^2 = 0 \text{ vs. } H_1 : \sigma_{00}^2 > 0$$

Efectuando um teste de Wald, a estatística de teste é dada por:

$$\frac{\hat{\sigma}_{00}^2 - 0}{S.E.(\hat{\sigma}_{00}^2)} = \frac{0,639 - 0}{0,228} = 2,801$$

Comparando o valor da estatística de teste com a distribuição Normal-Padrão, rejeita-se a hipótese nula, ou seja, existem evidências de que a variância é significativamente positiva (com um nível de significância de 0,05).

Para a variância do efeito aleatório  $u_{1j}$ , o ensaio de hipóteses é o seguinte:

$$H_0 : \sigma_{u1}^2 = 0 \text{ vs. } H_1 : \sigma_{u1}^2 > 0$$

A estatística de teste é dada por:

$$\frac{\hat{\sigma}_{00}^2 - 0}{S.E.(\hat{\sigma}_{00}^2)} = \frac{28,269 - 0}{19,219} = 1,471$$

Comparando o valor da estatística de teste (unilateral) com a distribuição Normal-Padrão, não se rejeita a hipótese nula com um nível de significância de 0,05, embora seja possível rejeitar com um nível de significância superior (0,10). Nas Tabelas 25 e 26 apresentam-se as estimativas dos efeitos fixos e as soluções dos efeitos aleatórios.

**Tabela 25 – Efeitos fixos (modelo multinível com variável JCS\_DAT)**

	Estimativa	Erro-padrão	G.L.	Valor t	Pr >  t
<b>Termo independente</b>	-0,00317	0,1803	20,91	-0,02	0,9861

**Tabela 26 – Efeitos aleatórios (modelo multinível com variável JCS\_DAT)**

Sujeito a		Estimativa	Erro-padrão	G.L.	Valor t	Pr >  t	
			Pred				
<b>Termo independente</b>	<b>Grupo 1</b>	Indústrias alimentares, bebidas e tabaco	-0,627	0,349	116,20	-1,80	0,0751
<b>JCS_DAT</b>	<b>Grupo 1</b>	Indústrias alimentares, bebidas e tabaco	2,402	3,258	30,63	0,74	0,4666
<b>Termo</b>	<b>Grupo 2</b>	Fabricação de têxteis	1,255	0,389	82,49	3,23	0,0018



Sujeito a		Estimativa	Erro-padrão Pred	G.L.	Valor t	Pr >  t	
<b>independente</b>							
JCS_DAT	Grupo 2	Fabricação de têxteis	-0,760	4,505	8,386	-0,17	0,8701
Termo independente	Grupo 3	Indústria do vestuário	1,265	0,280	81,02	4,52	<0,0001
JCS_DAT	Grupo 3	Indústria do vestuário	4,992	3,566	21,38	1,40	0,1758
Termo independente	Grupo 4	Indústria de couro, madeira e cortiça	1,072	0,337	75,91	3,18	0,0021
JCS_DAT	Grupo 4	Indústria de couro, madeira e cortiça	3,079	3,997	13,52	0,77	0,4543
Termo independente	Grupo 5	Fabricação de pasta, de papel, de cartão e seus artigos	0,736	0,536	67,36	1,37	0,1739
JCS_DAT	Grupo 5	Fabricação de pasta, de papel, de cartão e seus artigos	1,414	5,158	4,88	0,27	0,7952
Termo independente	Grupo 6	Fabricação de combustíveis e produtos petrolíferos refinados e Indústria Química	0,296	0,423	85,34	0,70	0,4856
JCS_DAT	Grupo 6	Fabricação de combustíveis e produtos petrolíferos refinados e Indústria Química	0,193	4,098	12,22	0,05	0,9632
Termo independente	Grupo 7	Fabrico de outros produtos minerais não metálicos	0,415	0,368	91,90	1,13	0,2623
JCS_DAT	Grupo 7	Fabrico de outros produtos minerais não metálicos	3,757	4,750	6,79	0,79	0,4557
Termo independente	Grupo 8	Indústrias metalúrgicas	0,018	0,326	75,85	0,06	0,9561
JCS_DAT	Grupo 8	Indústrias metalúrgicas	6,432	4,075	12,51	1,58	0,1394
Termo independente	Grupo 9	Máquinas e equipamentos	-0,478	0,356	85,78	-1,34	0,1827
JCS_DAT	Grupo 9	Máquinas e equipamentos	5,048	4,703	7,067	1,07	0,3184
Termo independente	Grupo 10	Fabricação de material de transporte	0,690	0,488	75,50	1,41	0,1618
JCS_DAT	Grupo 10	Fabricação de material de transporte	1,704	5,028	5,41	0,34	0,7475
Termo independente	Grupo 11	Outras Indústrias Transformadoras	0,355	0,314	74,65	1,13	0,2608
JCS_DAT	Grupo 11	Outras Indústrias Transformadoras	-1,749	4,150	11,64	-0,42	0,6812
Termo independente	Grupo 12	Promoção imobiliária (desenvolvimento de projectos de edifícios); construção de edifícios	0,468	0,251	65,17	1,87	0,0667
JCS_DAT	Grupo 12	Promoção imobiliária (desenvolvimento de projectos de edifícios); construção de edifícios	-1,218	2,576	78,31	-0,47	0,6376
Termo independente	Grupo 13	Engenharia Civil	0,622	0,358	71,29	1,74	0,0861
JCS_DAT	Grupo 13	Engenharia Civil	5,085	4,307	10,03	1,18	0,2649
Termo independente	Grupo 14	Actividades especializadas de	-0,081	0,275	74,48	-0,29	0,7696

Sujeito a		Estimativa	Erro-padrão Pred	G.L.	Valor t	Pr >  t	
		construção					
JCS_DAT	Grupo 14	Actividades especializadas de construção	6,353	3,331	28,05	1,91	0,0667
Termo independente	Grupo 15	Comércio, manutenção e reparação, de veículos automóveis e motociclos	-0,802	0,289	74,26	-2,77	0,0070
JCS_DAT	Grupo 15	Comércio, manutenção e reparação, de veículos automóveis e motociclos	3,880	4,259	10,51	0,91	0,3826
Termo independente	Grupo 16	Comércio por grosso (inclui agentes), excepto de veículos automóveis e motociclos	-0,396	0,228	48,63	-1,74	0,0884
JCS_DAT	Grupo 16	Comércio por grosso (inclui agentes), excepto de veículos automóveis e motociclos	7,927	2,425	99,73	3,27	0,0015
Termo independente	Grupo 17	Comércio a retalho, excepto de veículos automóveis e motociclos	-0,231	0,215	40,86	-1,07	0,2887
JCS_DAT	Grupo 17	Comércio a retalho, excepto de veículos automóveis e motociclos	-1,203	1,561	580,60	-0,77	0,4413
Termo independente	Grupo 18	Transportes e armazenagem	-0,275	0,284	93,66	-0,97	0,3351
JCS_DAT	Grupo 18	Transportes e armazenagem	-0,693	2,183	151,50	-0,32	0,7514
Termo independente	Grupo 19	Alojamento, restauração e similares	-0,885	0,277	87,34	-3,19	0,0020
JCS_DAT	Grupo 19	Alojamento, restauração e similares	-2,355	2,936	46,52	-0,80	0,4266
Termo independente	Grupo 20	Actividades de informação e de comunicação	-0,606	0,373	119,50	-1,63	0,1067
JCS_DAT	Grupo 20	Actividades de informação e de comunicação	1,963	5,058	5,28	0,39	0,7131
Termo independente	Grupo 21	Actividades imobiliárias	-1,080	0,335	91,14	-3,22	0,0018
JCS_DAT	Grupo 21	Actividades imobiliárias	1,937	3,871	15,38	0,50	0,6239
Termo independente	Grupo 22	Actividades de consultoria, científicas, técnicas e similares	-1,257	0,296	85,59	-4,25	<0,0001
JCS_DAT	Grupo 22	Actividades de consultoria, científicas, técnicas e similares	0,519	4,028	13,13	0,13	0,8995
Termo independente	Grupo 23	Actividades administrativas e dos serviços de apoio	-0,473	0,327	82,04	-1,45	0,1508
JCS_DAT	Grupo 23	Actividades administrativas e dos serviços de apoio	-1,455	4,211	10,99	-0,35	0,7363

Tendo-se novamente rejeitado a componente fixa do modelo, a Tabela 27 resume as soluções para os efeitos aleatórios que resultaram estatisticamente significativos.

**Tabela 27 – Resumo das estimativas dos efeitos aleatórios (modelo multinível com variável JCS\_DAT)**

	Grupo	Descrição	Logit ( $\hat{p}_i$ )	$\hat{p}_i$
Termo independente	Grupo 2	Fabricação de têxteis	1,255	0,778
Termo independente	Grupo 3	Indústria do vestuário	1,265	0,780
Termo independente	Grupo 4	Indústria de couro, madeira e cortiça	1,072	0,745
Termo independente	Grupo 15	Comércio, manutenção e reparação, de veículos automóveis e motociclos	-0,802	0,310
JCS_DAT	Grupo 16	Comércio por grosso (inclui agentes), excepto de veículos automóveis e motociclos	7,927	$\frac{e^{7,927 JCS\_DAT_i}}{1 + e^{7,927 JCS\_DAT_i}}$
Termo independente	Grupo 19	Alojamento, restauração e similares	-0,885	0,292
Termo independente	Grupo 21	Actividades imobiliárias	-1,080	0,254
Termo independente	Grupo 22	Actividades de consultoria, científicas, técnicas e similares	-1,257	0,222

### 5.3. Discussão comparativa dos resultados

Tendo-se estimado modelos de previsão de incumprimento de empresas através da regressão logística clássica e através de um modelo multinível com função de ligação *logit*, encontram-se evidências de que o sector de actividade em que a empresa se insere constitui um factor diferenciador da probabilidade de incumprimento.

Ainda que subsistam dúvidas sobre a qualidade do ajustamento global dos modelos estimados, é possível detectar a significância individual das variáveis que reflectem o sector de actividade.

No que respeita aos resultados da regressão logística, resultaram como variáveis explicativas, por um lado, um conjunto de variáveis de índole económico-financeira que representam a liquidez, estrutura de custos, solvência, rentabilidade e dimensão; por outro lado, a pertença a alguns sectores de actividade, que foi integrada no modelo como variável *dummy*.

A aplicação do modelo multinível com função de ligação *logit* permite detectar a existência de efeitos não observados ao nível do sector de actividade que implicam diferentes probabilidades de incumprimento das empresas portuguesas, possibilitando o cálculo da correlação intra-grupo, no modelo estimado apenas com termo independente. Conclui-se pela não significância estatística do efeito fixo, mas foi possível obter as soluções dos efeitos aleatórios correspondentes aos sectores de actividade, sem que nenhum desses sectores tivesse de servir como sector de referência, como acontece no caso da regressão logística.

A tentativa de estimar um modelo multinível com função de ligação *logit* considerando como potenciais variáveis explicativas as variáveis que resultaram estatisticamente significativas na regressão logística não foi muito bem sucedida. Apenas resultaram significativos os efeitos aleatórios correspondentes aos termos independentes de alguns sectores de actividade, excepto para o sector Comércio por grosso (inclui agentes), excepto de veículos automóveis e motociclos (grupo 16), para o qual resultou estatisticamente significativo o efeito aleatório associado à variável juros e custos similares/dívidas a terceiros (JCS\_DAT), embora este modelo apresente um pior ajustamento do que o modelo apenas com termo independente. Na regressão logística, este sector revelou-se significativo quando se consideram as variáveis centradas, estando associado a uma maior probabilidade de incumprimento. Apresenta-se de seguida um quadro resumo com as estimativas dos coeficientes para os três modelos obtidos, encontrando-se dentro de parêntesis o respectivo erro-padrão. Para promover a comparabilidade, apresentam-se na Tabela 28<sup>37</sup> os resultados da regressão logística sem variáveis centradas, já que os resultados de ajustamento dos modelos com e sem variáveis centradas são idênticos para as variáveis financeiras. A Tabela 29 resume as probabilidades estimadas associadas aos sectores de actividade no modelo estimado com a regressão logística e no modelo multinível apenas com termo independente, sendo apresentados apenas os coeficientes que resultaram estatisticamente significativos.

---

<sup>37</sup> Os valores entre parêntesis correspondem aos erros-padrão.

Tabela 28 – Quadro-resumo das estimativas dos coeficientes das soluções obtidas

Variável		Regressão logística (sem. cent.)	Logit Multinível apenas com termo independente	Logit multinível com termo independente e JCS_DAT
<b>Intercept</b>		-1,354 (0,474)	-	-
<b>AC_AT</b>	Activo circulante/Activo total (Liquidez)	0,948 (0,227)	-	-
<b>CD_AT</b>	Caixa e depósitos/Activo total (Liquidez)	-0,861 (0,349)	-	-
<b>FM_AT</b>	Fundo de Maneio/Activo total (Liquidez)	-0,232 (0,069)	-	-
<b>CP_PT</b>	Capital próprio/Passivo total (Solvência)	-0,629 (0,082)	-	-
<b>JCS_DAT</b>	Juros e custos similares/Dívidas a terceiros (Estrutura de custos)	8,368 (1,066)	-	-
<b>RO_AT</b>	Resultados operacionais/Activo total (Rentabilidade)	-2,413 (0,254)	-	-
<b>VN_AT</b>	Volume de negócios/Activo total (Rentabilidade)	-0,425 (0,049)	-	-
<b>ln_VN</b>	Volume de negócios (ln) (Dimensão)	0,080 (0,033)	-	-
<b>Grupo</b>	1 - Indústrias alimentares, bebidas e tabaco	-0,441 (0,336)	-	-
<b>Grupo</b>	2 - Fabricação de têxteis	1,083 (0,404)	1,157 (0,353)	1,255 (0,389)
<b>Grupo</b>	3 - Indústria do vestuário	1,655 (0,232)	1,373 (0,255)	1,265 (0,280)
<b>Grupo</b>	4 - Indústria de couro, madeira e cortiça	1,129 (0,282)	1,136 (0,290)	1,072 (0,337)
<b>Grupo</b>	13- Engenharia Civil	0,807 (0,313)	0,772 (0,309)	-
<b>Grupo</b>	15 - Comércio, manutenção e reparação, de veículos automóveis e motociclos	-0,778 (0,231)	-0,771 (0,262)	-0,802 (0,289)
<b>Grupo</b>	16 - Comércio por grosso (inclui agentes), excepto de veículos automóveis e motociclos	-	-	7,927 JCS_DAT (2,425)
<b>Grupo</b>	17 - Comércio a retalho, excepto de veículos automóveis e motociclos	-0,286 (0,137)	-	-
<b>Grupo</b>	19 - Alojamento, restauração e similares	-1,145 (0,263)	-1,059 (0,261)	-0,885 (0,277)
<b>Grupo</b>	21 - Actividades imobiliárias	-1,379 (0,300)	-1,084 (0,299)	-1,080 (0,335)
<b>Grupo</b>	22 - Actividades de consultoria, científicas, técnicas e similares	-1,126 (0,265)	-1,326 (0,272)	-1,257 (0,296)
<b>Grupo</b>	23 - Actividades administrativas e dos serviços de apoio	0 (Referência)	-0,608 (0,291)	-
<b>Correlação intra-grupo</b>		-	0,167	-

Variável	Regressão logística (sem. cent.)	Logit Multinível apenas com termo independente	Logit multinível com termo independente e JCS_DAT
-2logL	2500,454	-	-
AIC	2562,454	-	-
-2 Res Log Pseudo-Likelihood	-	10470,28	10502,44
Pseudo-AIC	-	10472,28	10506,44

Tabela 29 – Resumo das probabilidades de incumprimento associadas aos sectores de actividade

Grupo	Secção/divisão da CAE Rev.3	Descrição	Regressão logística		Logit multinível com termo independente	
			Logit ( $\hat{p}_i$ )	$\hat{p}_i$	Logit ( $\hat{p}_i$ )	$\hat{p}_i$
Grupo 1	C10, C11 e C12	Ind. alimentares, bebidas e tabaco	-0,441	0,392		
Grupo 2	C13	Fabricação de têxteis	1,083	0,747	1,1567	0,7601
Grupo 3	C14	Indústria do vestuário	1,655	0,840	1,3733	0,798
Grupo 4	C15 e C16	Indústria de couro, madeira e cortiça	1,129	0,756	1,1361	0,757
Grupo 13	F42	Engenharia Civil	0,807	0,691	0,7723	0,684
Grupo 15	G45	Comércio, manutenção e reparação, de veículos automóveis e motociclos	-0,778	0,315	-0,7705	0,316
Grupo 17	G47	Comércio a retalho, excepto de veículos automóveis e motociclos	-0,286	0,429		
Grupo 19	I	Alojamento, restauração e similares	-1,145	0,241	-1,0589	0,258
Grupo 21	L	Actividades imobiliárias	-1,379	0,201	-1,0836	0,253
Grupo 22	M	Actividades de consultoria, científicas, técnicas e similares	-1,126	0,245	-1,3256	0,210
Grupo 23	N	Actividades administrativas e dos serviços de apoio	0 (referência)	0,50	-0,6082	0,353

## Conclusão

O estudo do incumprimento das empresas é um tópico que continua a despertar o interesse de investigadores e académicos, estando rodeado de um inegável interesse prático. São vários os motivos apontados para que tal aconteça, que vão desde a vertente económico-social que caracteriza o assunto, passando pela necessidade de conhecimento do perfil de risco das empresas que necessitam de se financiar junto da banca para o normal desenvolvimento da sua actividade, pela aplicação a testes de resiliência do sistema financeiro a choques que podem passar por dificuldades generalizadas das empresas mutuárias em cumprir as obrigações assumidas.

Porém, não é fácil identificar uma formulação teórica única do incumprimento, o qual pode até ser definido de diferentes formas. Certos autores fazem coincidir o incumprimento com o acontecimento falência/insolvência, identificando-se igualmente abordagens mais conservadoras, nas quais o incumprimento coincide com um atraso de 90 dias no pagamento dos seus compromissos. Na verdade, a modelação do incumprimento parece ser um problema especialmente marcado por uma grande componente empírica, estando muito dependente dos dados disponíveis.

Ao longo dos anos, as variáveis de índole económico-financeira adquiriram um estatuto privilegiado de melhores variáveis explicativas do incumprimento das empresas, sendo que a sua utilização remonta a sistemas de análise assentes em julgamento humano, até que o desenvolvimento de técnicas estatísticas e computacionais permitiram a utilização generalizada de formas mais objectivas de análise. A análise discriminante multivariada e os modelos de probabilidade condicionada, nomeadamente os modelos *logit*, também parecem ter adquirido um estatuto privilegiado como auxiliares de análise do incumprimento das empresas. Em relação a estes dois aspectos, alguns autores chamam a atenção para as seguintes questões: por um lado, ainda que os rácios financeiros, como medidas relativas que são, tenham vindo a revelar-se particularmente úteis para a compreensão do fenómeno do incumprimento, que ocorre em empresas de diferente dimensão, a verdade é que os rácios podem ser distintos de acordo com o sector de actividade a que as empresas pertencem; por outro lado, alguns autores consideram uma lacuna o facto de não se ter verificado a

aplicação, a esta área de estudo, de desenvolvimentos metodológicos ao nível dos modelos de escolha discreta que têm emergido em outros campos das ciências sociais.

No presente trabalho são acomodadas as duas questões, tendo-se procurado testar a relevância do sector de actividade como factor diferenciador da probabilidade de incumprimento das empresas portuguesas, para o que se aplicou um modelo multinível com função de ligação *logit*, para além de uma regressão logística “tradicional”. A modelação foi em grande parte conduzida pelos dados disponíveis, tendo-se privilegiado uma lógica exploratória das variáveis identificadas em estudos recentes portugueses sobre o incumprimento de empresas. A base de dados utilizada no presente estudo foi disponibilizada pela Coface Serviços Portugal, S.A., dela constando elementos caracterizadores do tecido empresarial português, informações sobre processos de insolvência e dados de balanço. São consideradas como empresas em incumprimento as que registaram processos de insolvência em 2009, independentemente do estado do processo ou de quem o interpõe, entendendo-se que a existência de acções judiciais deste tipo constitui um indício forte de que existirá uma considerável probabilidade de a empresa não respeitar na íntegra as suas responsabilidades perante terceiros.

Tendo-se estimado modelos de previsão de incumprimento de empresas através da regressão logística clássica e através de um modelo multinível com função de ligação *logit*, encontraram-se evidências de que o sector de actividade em que a empresa se insere constitui um factor diferenciador da probabilidade de incumprimento. Ainda que subsistam dúvidas sobre a qualidade do ajustamento global dos modelos estimados, é possível detectar a significância individual das variáveis que reflectem o sector de actividade. Os sectores que aparecem como mais problemáticos, independentemente da técnica estatística utilizada, são a Fabricação de têxteis, a Indústria do vestuário e Indústria de couro, madeira e cortiça.

Na regressão logística resultaram ainda como variáveis significativas rácios económico-financeiros de liquidez, estrutura de custos, solvência, rentabilidade e dimensão.

A aplicação do modelo multinível com função de ligação *logit* permitiu detectar a existência de efeitos não observados ao nível do sector de actividade que implicam diferentes probabilidades de incumprimento das empresas portuguesas, possibilitando o cálculo da correlação intra-grupo, no modelo estimado apenas com termo independente. Conclui-se pela



não significância estatística do efeito fixo, mas foi possível estimar os efeitos aleatórios correspondentes aos sectores de actividade, sem que nenhum desses sectores tivesse de servir como sector de referência, como acontece no caso da regressão logística. A tentativa de estimar um modelo multinível com função de ligação *logit* considerando como potenciais variáveis explicativas as variáveis que resultaram estatisticamente significativas na regressão logística acabou por não ser bem sucedida. Apenas se conseguiram resultados considerando um rácio referente à estrutura de custos (juros e custos similares/dívidas a terceiros), e apenas se estimou um coeficiente aleatório estatisticamente significativo associado a essa variável no sector Comércio por grosso (inclui agentes), excepto de veículos automóveis e motociclos.

Como principal limitação ao desenvolvimento deste trabalho identifica-se a disponibilidade de dados. A modelação multinível implica uma particular exigência a este nível, dado que a vantagem da sua aplicação residirá na existência de um número interessante de grupos com significado, ou seja, com observações suficientes por categoria de estudo. Neste trabalho, optou-se por excluir alguns sectores de actividade precisamente devido ao facto de neles existirem poucos casos de empresas em incumprimento. A construção da amostra também se pautou bastante por considerações em torno dos dados disponíveis e a não consideração de uma amostra de validação foi uma opção consciente, embora se reconheça a sua utilização como desejável. Contudo, porque a desaceleração da economia se acentuou em 2009, efeito que não se pretendia captar neste trabalho, não se considerou conveniente criar uma amostra de validação com incumprimentos de anos anteriores. Por outro lado, a partição da amostra criada a partir dos incumprimentos de 2009 também levantaria problemas, porque alguns grupos ficariam com um número de observações muito reduzido.

Em aberto fica a possibilidade de estudar a adequabilidade de um modelo multinível considerando no nível superior variáveis relativas ao sector de actividade distintas da mera média das variáveis consideradas ao nível das empresas, mas outras que constituam, por exemplo, indicadores de risco sectorial.

## Bibliografia

- Aiken, L. S. e West, S. G. (1991). *Multiple Regression: Testing and Interpreting Interactions*, Newbury Park, CA: Sage
- Aitkin M., Anderson D. e Hinde J. (1981), Statistical modelling in school effectiveness studies (with discussion), *Journal of the Royal Statistical Society A*, 149, pp. 1-43
- Almeida, V. e Félix, R. (2006), Cálculo do produto potencial e do hiato do produto para a economia portuguesa, *Boletim Económico*, Banco de Portugal, Outono.
- Altman, E. I. e Saunders, A. (1998), Credit risk measurement: Developments over the last 20 years, *Journal of Banking and Finance*, 21, pp. 1721-1742
- Altman, E.I., Haldeman, R., Narayanan, P. (1977), Zeta analysis: A new model to identify bankruptcy risk of corporations, *Journal of Banking and Finance*, pp. 29-54.
- Altman, E.I. (1968), Financial ratios, discriminant analysis and the prediction of corporate bankruptcy, *Journal of Finance*, 23(4), pp. 589-609
- Altman, E.I. (2000), Predicting financial distress of companies: revisiting the z-score and zeta models [disponível em <http://pages.stern.nyu.edu/~ealtman/Zscores.pdf>, consultado em 3/12/2009]
- Amrhein D.G. e Katz, J.A. (1998), Cash flow as a factor in the mortality risk of a business, International Council for Small Business, 1998 Singapore Conference Proceedings, Paper nr. 18 [citado em Balcaen e Ooghe (2006)]
- Antunes, A., Ribeiro, N. e Antão, P. (2005), Estimativas de probabilidades de incumprimento em contexto macroeconómico, *Relatório de Estabilidade Financeira*, Banco de Portugal
- Balcaen, S. e Ooghe, H. (2006), 35 years of studies on business failure: An overview of the classical statistical methodologies and their related problems, *British Accounting Review*, 38(1), pp. 63-93
- Banco de Portugal (2007), Aviso do Banco de Portugal n.º 5/2007 [disponível em [http://www.bportugal.pt/sibap/sibap\\_p.htm](http://www.bportugal.pt/sibap/sibap_p.htm); consultado em 3/12/2009]

- Barros, C., Ferreira, C. e Williams, J. (2007), Analysing the determinants of performance of best and worst European Banks: A mixed logit approach, *Journal of Banking and Finance*, 31, pp. 2189-2203
- Barros, G. (2008), *Modelos de previsão da falência de empresas – aplicação empírica ao caso das pequenas e médias empresas portuguesas*, Tese de mestrado em Economia e Políticas Públicas, ISCTE
- Basel Committee on Banking Supervision (2006), International Convergence of Capital Measurement and Capital Standards, A revised framework comprehensive version, *Bank for International Settlements*
- Beaver, W. H. (1966), Financial ratios as predictors of failure, *Journal of Accounting Research*, 4, pp. 71-111
- Bennet, N. (1976), *Teaching Styles and Pupil Progress*, London: Open Books [citado em Guo e Zhao (2000)]
- Bickerdyke, I., Lattimore, R. e Madge, A. (2000), Business failure and change: an Australian perspective, Productivity Commission Staff Research Paper, Ausinfo, Camberra, pp. 1-191 [citado em Balcaen e Ooghe (2006)]
- Bonfim, D. (2006), Factores determinantes do risco de crédito: O contributo de características das empresas e da envolvente macroeconómica, *Relatório de Estabilidade Financeira*, Banco de Portugal
- Breiman, L., Friedman, J. H., Olshen, R. A. e Stone, C. J. (1984), *Classification and Regression Trees*, Wadsworth International Group, Belmont, CA
- Chan-Lau, J.A (2006), Fundamentals-based estimation of default probabilities: A survey, *IMF Working Paper* 06/149, pp. 1-20 [disponível em <http://ssrn.com/abstract=920241>; consultado em 4/12/2009]
- Código da Insolvência e da Recuperação de Empresas, estabelecido pelo Decreto-Lei n.º 53/2004 de 18 de Março de 2004 [disponível <http://www.iapmei.pt/iapmei-leg-03.php?lei=2785>, consultado em 17/03/2010]
- Cox, D.R. e Snell, E.J., (1989), *Analysis of Binary Data*, 2nd Ed. London: Chapman and Hall

- Deakin, E. D. (1976), Distributions of financial accounting ratios: Some empirical evidence, *The Accounting Review*, 51(1), pp. 90-96
- Doumpos M. e Zopoudinis C. (1999), A multicriteria discrimination method for the prediction of financial distress: the case of Greece, *Multinational Finance Journal*, 3(2), pp. 71-101 [citado em Balcaen e Ooghe (2006)]
- Falk, H. e Heintz, J. A. (1975), Assessing industry risk by ratio analysis, *The Accounting Review*, 50(4), pp. 758-779
- Fitzpatrick, P. (1932), A comparison of ratios of successful industrial enterprises with those of failed firms, *Certified Public Accountant* 12, pp. 598-605 [citado em Chan-Lau (2006)]
- Gamelas, E. (2005), *Las quiebras de empresas en Portugal: análisis económico y evidencia empírica*, Tese de doutoramento em Economia Financeira e Contabilidade (Ciências Empresariais), Universidad de Extremadura
- Gelman, A. e Hill, J. (2007), *Data Analysis using Regression and Multilevel/Hierarchical Models*, Cambridge University Press
- Goldstein, H. (1999), *Multilevel Statistical Models*, London: Institute of Education, Multilevel Models Project [disponível em <http://www.soziologie.uni-halle.de/langer/multilevel/books/goldstein.pdf>; consultado em 3/12/2009]
- Greene, W. (2003), *Econometric Analysis*, 5.<sup>a</sup> Ed., New Jersey, EUA: Prentice Hall
- Greene, W., Hensher, D. e Jones, S. (2007), An error component logit analysis of corporate bankruptcy and insolvency risk in Australia, *The Economic Record*, 83(260), pp. 86-103
- Guo, G. e Zhao, H. (2000), Multilevel modeling for binary data, *Annual Review of Sociology*, 26, pp. 441-462
- Hagan, M.T., Demuth, H. B. e Beale, M. (1996), *Neural Network Design*, PWS Publishing Company [citados em Pereira *et al.* (2007)]
- Hensher, D. e Jones, S. (2004), Predicting firm financial distress: A mixed logit model, *The Accounting Review*, 79(4), pp. 1011-1038
- Hensher, D. e Jones, S. (2007), Forecasting corporate bankruptcy: Optimizing the performance of the mixed logit model, *ABACUS*, 43(3), pp. 241-264

- Hossari, G. (2009), Absence of industry effect in modelling corporate collapse in Australia, *Global Economy and Finance Journal*, 2(1), pp. 20-30
- Hox, J.J. (2002), *Applied Multilevel Analysis*, Amsterdam: TT - Publikaties
- Izan, H. Y. (1984), Corporate distress in Australia, *Journal of Banking and Finance*, 8, pp. 303-320 [citado em Altman e Saunders (1998)]
- Kiang, M.Y. e Tam, K. Y. (1992), Managerial applications of neural networks: The case of bank failure predictions, *Management Science*, 38(7), pp. 926-947
- Kreft, I. e De Leeuw, J. (1998), *Introducing Multilevel Modelling*, London: Sage Publications
- Kumar, P. e Ravi, V. (2007), Bankruptcy prediction in banks and firms via statistical and intelligent techniques – A review, *European Journal of Operational Research*, 180, pp. 1–28
- Lacerda, A. e Moro, R. (2008), Analysis of the predictors of default for portuguese firms, *Working Papers*, 22, Banco de Portugal
- McCulloch, W.S. e Pitts, W. (1943), A logical calculus of the ideas immanent in nervous activity, *Bulletin of Mathematical Biophysics*, 5, pp. 115-133
- Merton, R. C. (1974), On the pricing of corporate debt: the risk structure of interest rate, *Journal of Finance*, 29(2), pp. 449-470
- Nagelkerke, N. J. D. (1991), A note on a general definition of the coefficient of determination, *Biometrika*, 78(3), pp. 691-692
- Ohlson, J. A. (1980), Financial ratios and the probabilistic prediction of bankruptcy, *Journal of Accounting Research*, 18(1), pp. 109-131
- Pereira, J. M., Dominguez, M. A. e Ocejó, J.L.S. (2007), Modelos de previsão do fracasso empresarial: Aspectos a considerar, *Revista de Estudos Politécnicos*, IV(7), pp. 111-148
- Platt, H.D e Platt, M.B. (1991), A note on the use of industry-relative ratios in bankruptcy prediction, *Journal of Banking and Finance*, 15, pp 1183-1194
- Platt, H.D. e Platt, M.B. (2006), Comparing financial distress and bankruptcy, *Working Paper* [disponível em <http://ssrn.com/abstract=876470>; consultado em 4/12/2009]

- Raudenbush, S. e Bryk, A. (2002), *Hierarchical Linear Models*, CA: Sage Publications
- Santos, P. (2000), *Falência empresarial - modelo discriminante e logístico de previsão aplicados às PME do sector têxtil e do vestuário*, Tese de mestrado em Contabilidade e Auditoria, Universidade Aberta, ISCAC
- SAS (2006), *SAS for Mixed Models*, 2<sup>nd</sup> Edition, SAS Institute Inc., Cary: NC
- Shepanski, A. (1983), Tests of theories of information processing behaviour in credit judgement, *Accounting Review*, 58, pp. 581-599 [citado em Kiang e Tam (1992)]
- Shumway, T. (2001), Forecasting bankruptcy more accurately: A simple hazard model, *Journal of Business*, 74(1), pp. 101-124
- Singer, J. (1998), Using SAS Proc Mixed to fit multilevel models, hierarchical models, and individual growth models, *Journal of Educational and Behavioral Statistics*, 24(4), pp. 323-355
- Snijders, T. e Bosker, R. (1999), *Multilevel Analysis*, London: Sage Publications
- Soares, M. C. (2006), Modelação de um indicador de notação da qualidade de crédito de sociedades não financeiras – Um estudo preliminar fundado na análise discriminante, *Relatório de Estabilidade Financeira*, Banco de Portugal
- Tam, K. Y. e Kiang, M. Y. (1992), Managerial applications of neural networks: The case of bank failure predictions, *Management Science*, 38(7), pp. 926-947
- Tamari, M. (1966), Financial ratios as a means of forecasting bankruptcy, *Management International Review*, 4, pp. 15-21 [citado em Balcaen e Ooghe (2006)]
- Tristão, M. (1997), *Risco de falência das empresas portuguesas*, Tese de mestrado em Ciências Empresariais, ISCTE
- Vermeulen, R. (2008), Modelling industry-level corporate credit risk for the Netherlands, *DNB Working paper*, n. ° 190
- Wolfinger, R. e O'Connell, M. (1993), Generalized linear mixed models: a pseudo-likelihood approach, *Journal of Statistical Computation and Simulation*, 4, pp. 233-243 [citado em SAS (2006) ]

Zhang, J.L. e Härdle, W. (2008), The Bayesian additive classification tree applied to credit risk modelling, *SFB 649, Discussion Paper*, Universidade Humboldt, Berlim, Alemanha [disponível em: <http://sfb649.wiwi.hu-berlin.de/papers/pdf/SFB649DP2008-003.pdf>; consultado em 3/12/2009]

Zmijewski, M. E. (1984), Methodological issues related to the estimation of financial distress prediction models, *Journal of Accounting Research*, 22, pp. 59-82

## **Anexos**



**Anexo 1 - Transformação proposta por Wolfinger e O'Connell (1993)<sup>38</sup>**

Retomando as equações (18) e (19),  $E[\mathbf{Y}|\mathbf{u}] = \mathbf{g}^{-1}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}) = \mathbf{g}^{-1}(\boldsymbol{\eta}) = \boldsymbol{\varphi}$ , em que  $\mathbf{u} \sim N(\mathbf{0}, \mathbf{G})$  e  $\text{var}(\mathbf{Y}|\mathbf{u}) = \mathbf{A}^{1/2} \mathbf{R} \mathbf{A}^{1/2}$ .

Seguindo Wolfinger e O'Connell (1993), é possível linearizar  $\mathbf{g}^{-1}(\boldsymbol{\eta})$  aplicando uma série de Taylor de primeira-ordem<sup>39</sup>:

$$\mathbf{g}^{-1}(\boldsymbol{\eta}) \cong \mathbf{g}^{-1}(\tilde{\boldsymbol{\eta}}) + \tilde{\Delta} \mathbf{X} (\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}) + \tilde{\Delta} \mathbf{Z} (\mathbf{u} - \tilde{\mathbf{u}}), \quad (\text{i})$$

em que  $\tilde{\Delta} = \frac{\partial \mathbf{g}^{-1}(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}$  é uma matriz de derivadas da média condicionada avaliada no ponto de expansão em série.

Arranjando os termos obtém-se a expressão:

$$\tilde{\Delta}^{-1} (\boldsymbol{\mu} - \mathbf{g}^{-1}(\tilde{\boldsymbol{\eta}})) + \mathbf{X} \tilde{\boldsymbol{\beta}} + \mathbf{Z} \tilde{\mathbf{u}} \cong \mathbf{X} \boldsymbol{\beta} + \mathbf{Z} \boldsymbol{\mu} \quad (\text{ii})$$

Sendo o lado esquerdo da equação o valor esperado, condicionado a  $\mathbf{u}$ , de  $\mathbf{P}$ , e

$$\text{var}[\mathbf{P}|\mathbf{u}] = \tilde{\Delta}^{-1} \mathbf{A}^{1/2} \mathbf{R} \mathbf{A}^{1/2} \tilde{\Delta}^{-1} \quad (\text{iii})$$

Pode, assim, considerar-se o modelo

$$\mathbf{P} = \mathbf{X} \boldsymbol{\beta} + \mathbf{Z} \mathbf{u} + \boldsymbol{\varepsilon}, \quad (\text{iv})$$

que é um modelo linear misto com uma pseudo-resposta  $\mathbf{P}$ , efeitos fixos  $\boldsymbol{\beta}$ , efeitos aleatórios  $\mathbf{u}$  e  $\text{var}(\boldsymbol{\varepsilon}) = \text{var}[\mathbf{P}|\mathbf{u}]$ . Este modelo pode ser estimado com métodos *standard* e as estimativas de  $\boldsymbol{\beta}$  e as soluções de  $\mathbf{u}$  obtidas são utilizadas para actualizar os pseudo-dados, num processo iterativo.

<sup>38</sup> Apresentada na documentação de apoio do SAS Glimmix.

<sup>39</sup> Uma série de Taylor é uma representação de uma função como uma soma infinita de termos calculados a partir dos valores das suas derivadas num único ponto. Sendo de primeira ordem, consideram-se apenas as primeiras derivadas.

## Anexo 2 – Matriz de correlações

antiguidade	Antiguidade	ln_AT	ln_VN	AC_AT	IM_AT	FM_AT	PCP_PT	PCP_PT_PROV	PCP_AT	ALAVANCAG
	1	0,23578	0,20725	0,07608	-0,075	0,06223	-0,0607	-0,05715	-0,04631	-0,03855
		<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001
AC_AT	AC_AT	IM_AT	CD_AT	FM_AT	PCP_PT_PROV	PCP_PT	ln_AT	antiguidade	VN_AT	ln_VN
	1	-0,98065	0,26048	0,21148	0,18025	0,17889	-0,0844	0,07608	0,06979	0,04232
		<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001
AC_DAT	AC_DAT	DDT_PROV_DAT	AC_PCP	AC_EX_PC_P	CD_PCP	CP_PT	JDS_DAT	CD_AT	ln_VN	AC_AT
	1	0,81931	0,76135	0,75317	0,49687	0,19089	0,06087	0,01101	-0,00898	0,00875
		<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001
AC_PCP	AC_PCP	AC_EX_PC_P	AC_DAT	DDT_PROV_DAT	CD_PCP	CP_PT	JDS_DAT	PCP_PT	PCP_PT_PROV	ln_AT
	1	0,95244	0,76135	0,628	0,60349	0,1452	0,04637	-0,03143	-0,03116	0,01444
		<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001
AC_EX_PCP	AC_EX_PCP	AC_PCP	AC_DAT	DDT_PROV_DAT	CD_PCP	CP_PT	JDS_DAT	PCP_PT	PCP_PT_PROV	ln_AT
	1	0,95244	0,75317	0,62994	0,59499	0,13463	0,0435	-0,02211	-0,02179	0,01191
		<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001
IM_AT	IM_AT	AC_AT	CD_AT	FM_AT	PCP_PT_PROV	PCP_PT	ln_AT	antiguidade	VN_AT	ln_VN
	1	-0,98065	-0,25364	-0,20089	-0,1788	-0,1773	0,08518	-0,07501	-0,0744	-0,04681
		<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001
ln_AT	ln_AT	ln_VN	CD_AT	antiguidade	PCP_AT	VN_AT	PCP_PT	PCP_PT_PROV	ALAVANCAG	CP_AT
	1	0,75925	-0,26496	0,23578	-0,1854	-0,18211	-0,177	-0,17312	-0,16677	0,16675
		<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001
AT_VN	AT_VN	RL_VN	RO_VN	EBIT_VN	ln_VN	ln_AT	AC_AT	IM_AT	CD_PCP	CD_AT
	1	-0,31151	-0,12964	-0,05687	-0,0296	0,01553	-0,007	0,00634	0,00528	-0,00405
		<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	0,0019	0,0047	0,0188	0,0717
CD_AT	CD_AT	ln_AT	AC_AT	IM_AT	ln_VN	FM_AT	VN_AT	DAT_AT	CP_AT	ALAVANCAG
	1	-0,26496	0,26048	-0,25364	-0,1303	0,12099	0,1134	-0,09202	0,08818	-0,08818
		<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001
CD_PCP	CD_PCP	AC_PCP	AC_EX_PC_P	AC_DAT	DDT_PROV_DAT	CP_PT	JDS_DAT	CD_AT	PCP_PT	PCP_PT_PROV
	1	0,60349	0,59499	0,49687	0,43236	0,21785	0,08181	0,03876	-0,02063	-0,02056
		<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001

CP_AT	CP_AT	ALAVANCAG	DAT_AT	PCP_AT	FM_AT	RL_AT	EBIT_AT	RO_AT	VN_AT	ln_AT
	1	-1	-0,99732	-0,90727	0,8927	0,53633	0,51925	0,40947	-0,17989	0,16675
		<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001
CP_PT	CP_PT	DDT_PROV_DAT	CD_PCP	AC_DAT	AC_PCP	AC_EX_PC P	JDS_DAT	ln_VN	CD_AT	ALAVANCAG
	1	0,38027	0,21785	0,19089	0,1452	0,13463	0,13142	-0,02018	0,0184	-0,01409
		<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001
DDT_PROV_A T_IM	DDT_PROV_AT_I M	IM_AT	ln_VN	AC_AT	ln_AT	antiguidade	FM_AT	VN_AT	RO_AT	PCP_PT
	1	-0,00264	0,00237	0,00229	0,00093	0,00066	0,00062	0,00062	0,00034	-0,00031
		0,2397	0,2907	0,3074	0,6797	0,7678	0,7833	0,7842	0,8789	0,8912
DDT_PROV_D AT	DDT_PROV_DAT	AC_DAT	AC_EX_PC P	AC_PCP	CD_PCP	CP_PT	JDS_DAT	ln_VN	FM_AT	AC_AT
	1	0,81931	0,62994	0,628	0,43236	0,38027	0,10831	-0,01555	0,01439	0,01424
		<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001
DAT_A T	DAT_AT	ALAVANCAG	CP_AT	PCP_AT	FM_AT	RL_AT	EBIT_AT	RO_AT	VN_AT	ln_AT
	1	0,99732	-0,99732	0,9087	-0,8935	-0,52821	-0,5107	-0,40111	0,1699	-0,16563
		<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001
FM_AT	FM_AT	PCP_AT	DAT_AT	CP_AT	ALAVANCAG	RL_AT	EBIT_AT	RO_AT	AC_AT	IM_AT
	1	-0,97485	-0,89348	0,8927	-0,8927	0,53094	0,51631	0,40189	0,21148	-0,20089
		<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001
JDS_D AT	JDS_DAT	CP_PT	DDT_PROV_DAT	CD_PCP	AC_DAT	AC_PCP	AC_EX_PC P	VN_AT	CD_AT	DAT_AT
	1	0,13142	0,10831	0,08181	0,06087	0,04637	0,0435	0,01793	0,01086	-0,01075
		<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001
PCP_A T	PCP_AT	FM_AT	DAT_AT	CP_AT	ALAVANCAG	RL_AT	EBIT_AT	RO_AT	PCP_PT	PCP_PT PROV
	1	-0,97485	0,9087	-0,90727	0,90727	-0,53464	-0,5204	-0,40154	0,23906	0,23801
		<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001
PCP_P T	PCP_PT	PCP_PT_P ROV	PCP_AT	FM_AT	AC_AT	IM_AT	ln_AT	ln_VN	CD_AT	antiguidade
	1	0,99787	0,23906	-0,19379	0,17889	-0,1773	-0,177	-0,07372	0,06757	-0,06069
		<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001
PCP_P T_PROV	PCP_PT_PROV	PCP_PT	PCP_AT	FM_AT	AC_AT	IM_AT	ln_AT	ln_VN	CD_AT	VN_AT
	1	0,99787	0,23801	-0,19247	0,18025	-0,17876	-0,1731	-0,07045	0,06807	0,05811
		<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001
ALAVANCAG	ALAVANCAG	CP_AT	DAT_AT	PCP_AT	FM_AT	RL_AT	EBIT_AT	RO_AT	VN_AT	ln_AT
	1	-100.000	0,99732	0,90727	-0,8927	-0,53633	-0,5193	-0,40947	0,17988	-0,16677
		<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001

RL_AT	RL_AT	EBIT_AT	RO_AT	ALAVANC AG	CP_AT	PCP_AT	FM_AT	DAT_AT	VN_AT	ln_AT
	1	0,99336	0,7513	-0,53633	0,53633	-0,53464	0,53094	-0,52821	-0,2139	0,14605
		<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001
RL_CP	RL_CP	ln_VN	VN_AT	ln_AT	EBIT_AT	RO_AT	RL_AT	CD_AT	antiguidade	PCP_PT_ PROV
	1	0,00338	0,0027	0,00223	0,00214	0,00192	0,00182	0,00151	0,00142	0,00109
		0,1322	0,2291	0,3219	0,3419	0,3928	0,417	0,5026	0,5288	0,6287
RL_VN	RL_VN	RO_VN	EBIT_VN	AT_VN	ln_VN	DAT_AT	ALAVANC AG	CP_AT	antiguidade	RL_AT
	1	0,89681	0,88014	-0,31151	0,03322	-0,01186	-0,0118	0,01181	-0,00924	0,00792
		<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	0,0004
EBIT_VN	EBIT_VN	RO_VN	RL_VN	AT_VN	ln_VN	CD_PCP	DAT_AT	ALAVAN CAG	CP_AT	EBIT_A T
	1	0,94153	0,88014	-0,05687	0,02068	0,01146	-0,0105	-0,01039	0,01039	0,00844
		<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	0,0002
EBIT_J CS	EBIT_JCS	EBIT_AT	RL_AT	RO_AT	ALAVANC AG	CP_AT	ln_AT	DAT_AT	FM_AT	PCP_AT
	1	0,03962	0,03822	0,03797	-0,0236	0,02364	0,02343	-0,02313	0,02258	-0,02146
		<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001
EBIT_AT	EBIT_AT	RL_AT	RO_AT	PCP_AT	ALAVANC AG	CP_AT	FM_AT	DAT_AT	VN_AT	ln_AT
	1	0,99336	0,73597	-0,52036	-0,5193	0,51925	0,51631	-0,51072	-0,16233	0,14914
		<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001
RO_AT	RO_AT	RL_AT	EBIT_AT	CP_AT	ALAVANC AG	FM_AT	PCP_AT	DAT_AT	VN_AT	ln_AT
	1	0,7513	0,73597	0,40947	-0,4095	0,40189	-0,4015	-0,40111	-0,33623	0,14068
		<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001
RO_VN	RO_VN	EBIT_VN	RL_VN	AT_VN	ln_VN	ALAVANC AG	CP_AT	DAT_AT	RO_AT	antiguidade
	1	0,94153	0,89681	-0,12964	0,02355	-0,00935	0,00935	-0,00928	0,00713	-0,00685
		<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	0,0015	0,0023
ln_VN	ln_VN	ln_AT	antiguidade	VN_AT	CD_AT	EBIT_AT	FM_AT	RO_AT	DAT_AT	PCP_AT
	1	0,75925	0,20725	0,14838	-0,1303	0,12982	0,12691	0,12566	-0,1244	-0,12019
		<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001
VN_AT	VN_AT	RO_AT	RL_AT	ln_AT	PCP_AT	CP_AT	ALAVANC AG	DAT_AT	EBIT_AT	FM_AT
	1	-0,33623	-0,2139	-0,18211	0,17991	-0,17989	0,17988	0,1699	-0,16233	-0,1603
		<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001

### Anexo 3 – Estudo da multicolinearidade

Collinearity Diagnostics																			
Number	Eigenvalue	Condition Index	Proportion of Variation																
			Intercept	antiguidade	AC_AT	AT_VN	CD_AT	CD_PCP	CP_PT	DDT_PR OV_AT_I M	EBIT_JCS	FM_AT	JDS_DAT	ln_VN	PCP_PT_ PROV	RL_CP	RL_VN	RO_AT	VN_AT
1	5,14	1,00	0,00	0,01	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,36	0,00	0,00	0,01
2	1,53	1,83	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,71	0,00	0,19	0,00	0,00	0,00	0,04	0,00	0,22	0,04
3	1,31	1,98	0,01	0,00	0,06	0,34	0,38	0,01	0,01	0,00	0,00	0,00	0,00	0,46	0,08	0,00	0,34	0,00	0,00
4	1,29	1,99	0,00	0,00	0,00	0,01	0,00	0,26	0,30	0,01	0,00	0,00	0,15	0,00	0,00	0,03	0,01	0,00	0,00
5	1,00	2,27	0,09	0,00	0,62	0,34	0,00	0,00	0,00	0,23	0,01	0,00	0,22	0,06	0,25	0,76	0,03	0,33	0,00
6	1,00	2,27	0,01	0,00	0,35	0,15	0,00	0,02	0,00	0,77	0,00	0,00	0,22	0,01	0,00	0,23	0,01	0,04	0,00
7	1,00	2,27	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,98	0,00	0,00	0,00	0,00	0,01	0,24	0,00	0,00
8	0,93	2,35	0,28	0,00	0,00	0,00	0,00	0,22	0,03	0,00	0,00	0,00	0,79	0,05	0,00	0,00	0,00	0,00	0,00
9	0,78	2,57	0,00	0,00	0,00	0,00	0,00	0,49	0,66	0,00	0,00	0,00	0,05	0,00	0,00	0,01	0,00	0,00	0,00
10	0,71	2,68	0,00	0,04	0,00	0,01	0,10	0,01	0,01	0,00	0,00	0,35	0,00	0,00	0,00	0,00	0,00	0,09	0,25
11	0,69	2,73	0,00	0,00	0,00	0,65	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,65	0,00	0,00
12	0,57	3,01	0,00	0,16	0,00	0,00	0,46	0,00	0,00	0,88	0,00	0,24	0,00	0,00	0,00	0,00	0,00	0,17	0,00
13	0,50	3,21	0,00	0,02	0,00	0,00	0,22	0,00	0,00	0,75	0,00	0,03	0,00	0,00	0,00	0,00	0,00	0,42	0,56
14	0,39	3,63	0,00	0,69	0,01	0,00	0,12	0,00	0,00	0,50	0,00	0,05	0,00	0,00	0,02	0,00	0,00	0,06	0,08
15	0,08	7,84	0,00	0,01	0,67	0,00	0,02	0,00	0,00	0,00	0,00	0,12	0,00	0,00	0,54	0,57	0,00	0,01	0,00
16	0,07	8,33	0,04	0,04	0,30	0,00	0,03	0,00	0,00	0,60	0,00	0,00	0,00	0,07	0,38	0,09	0,00	0,00	0,01
17	0,01	25,14	0,96	0,02	0,01	0,00	0,03	0,00	0,00	0,00	0,00	0,01	0,00	0,93	0,05	0,00	0,00	0,03	0,05

#### Anexo 4 – Valores médios das variáveis potencialmente explicativas na amostra final

Valores médios das variáveis potencialmente explicativas na amostra final (antiguidade, dimensão, actividade e liquidez)

		Antiguidade	Dimensão		Actividade		Liquidez		
			In_VN	AT_VN	AC_AT	CD_AT	CD_PCP	FM_AT	PCP_PT_PROV
1	Ind. alimentares, bebidas e tabaco	216,71	13,28	3,59	0,51	0,12	0,36	-0,10	0,66
2	Fabricação de têxteis	273,32	13,35	1,62	0,60	0,06	0,21	-0,47	0,63
3	Indústria do vestuário	154,04	12,45	1,50	0,72	0,16	0,29	-0,39	0,76
4	Ind. couro, madeira e cortiça	228,25	13,12	3,54	0,75	0,12	0,17	-1,66	0,81
5	Fab. pasta, de papel, de cartão e seus artigos	123,02	12,32	1,55	0,54	0,09	0,11	-1,77	0,85
6	Fab. combustíveis e prod. petrolíferos refinados e Ind. Química	205,49	14,22	1,43	0,67	0,09	0,69	-0,16	0,80
7	Fabrico de outros produtos minerais não metálicos	252,93	13,05	7,10	0,67	0,05	0,06	-0,09	0,77
8	Indústrias metalúrgicas	206,10	13,14	1,49	0,72	0,07	0,2	0	0,76
9	Máquinas e equipamentos	250,43	13,63	3,95	0,74	0,07	0,37	0	0,78
10	Fabricação de material de transporte	294,45	14,30	1,74	0,67	0,07	0,17	-0,11	0,74
11	Outras Indústrias Transformadoras	260,42	12,81	4,13	0,69	0,08	0,23	-0,23	0,79
12	Prom. imobiliária (desenvolvimento de projectos de edifícios); construção de edifícios	164,78	13,16	70,35	0,85	0,08	0,48	0,09	0,72
13	Engenharia Civil	227,36	14,09	5,14	0,79	0,08	1,02	0,10	0,72
14	Actividades especializadas de construção	147,62	12,77	1,48	0,81	0,12	0,48	-0,25	0,79
15	Comércio, manutenção e reparação, de veíc. automóveis e motociclos	207,60	12,99	4,01	0,79	0,08	0,16	-0,21	0,79
16	Comércio por grosso (inclui agentes), excepto de veíc. automóveis e	190,62	13,40	2,12	0,80	0,10	0,33	-0,36	0,82

		<b>Antiguidade</b>	<b>Dimensão</b>	<b>Actividade</b>	<b>Liquidez</b>				
	motociclos								
17	Comércio a retalho, excepto de veículos automóveis e motociclos	205,26	12,62	2,02	0,74	0,12	0,86	-0,12	0,82
18	Transportes e armazenagem	177,37	12,48	1,56	0,71	0,13	1,43	-0,07	0,83
19	Aloj., restauração e similares	188,45	11,81	1,95	0,52	0,21	1,27	-0,53	0,77
20	Actividades de informação e de comunicação	100,44	12,12	3,99	0,76	0,12	0,41	-0,17	0,81
21	Act. imobiliárias	156,21	12,16	130,02	0,60	0,09	0,16	-0,25	0,75
22	Act. consultoria, científicas, técnicas e similares	129,68	12,18	1,61	0,70	0,16	0,68	-0,10	0,76
23	Act. administrativas e dos serviços de apoio	141,88	12,79	1,36	0,65	0,14	0,54	-0,20	0,78
	<b>Média geral</b>	189,14	12,88	12,12	0,73	0,11	0,53	-0,26	0,78

**Valores médios das variáveis potencialmente explicativas na amostra final (estrutura, estrutura de custos, rentabilidade)**

		<b>Solvência</b>	<b>Estrutura</b>	<b>Estrutura de custos</b>			<b>Rentabilidade</b>		
		<b>CP_PT</b>	<b>DDT_PROV_ AT_IM</b>	<b>EBIT_JCS</b>	<b>JCS_DAT</b>	<b>RL_CP</b>	<b>RL_VN</b>	<b>RO_AT</b>	<b>VN_AT</b>
<b>1</b>	Ind. alimentares, bebidas e tabaco	1,20	-1,56	76,49	0,05	-0,19	-3,14	0	1,15
<b>2</b>	Fabricação de têxteis	0,17	-1,54	-51,35	0,05	14,22	-0,43	-0,28	1,12
<b>3</b>	Indústria do vestuário	0,07	0,10	-80,43	0,05	3,58	-0,30	-0,27	2,17
<b>4</b>	Ind. couro, madeira e cortiça	0,17	0,30	-7,17	0,05	-0,25	-0,30	-0,17	1,32
<b>5</b>	Fab. pasta, de papel, de cartão e seus artigos	-0,21	0,62	-31,69	0,03	0,66	-0,90	-0,89	0,92
<b>6</b>	Fab. combust. e prod. petrolíferos refinados e Ind. Química	0,42	-1,98	42,60	0,06	0,73	-0,20	-0,10	0,97
<b>7</b>	Fabrico de outros produtos minerais não metálicos	0,20	1,10	-20,75	0,04	0,09	-0,42	-0,09	0,70
<b>8</b>	Indústrias metalúrgicas	0,33	0,41	17,94	0,05	0,22	-0,13	-0,06	1,03
<b>9</b>	Máquinas e equipamentos	0,65	2,31	-17,04	0,04	-0,71	-0,31	-0,03	1,11
<b>10</b>	Fabricação de material de	0,29	-0,75	-3,84	0,05	-0,39	-0,53	-0,13	1,06

		<b>Solvência</b>	<b>Estrutura</b>	<b>Estrutura de custos</b>		<b>Rentabilidade</b>			
		<b>CP_PT</b>	<b>DDT_PROV_ AT_IM</b>	<b>EBIT_JCS</b>	<b>JCS_DAT</b>	<b>RL_CP</b>	<b>RL_VN</b>	<b>RO_AT</b>	<b>VN_AT</b>
	transporte								
<b>11</b>	Outras Indústrias Transformadoras	0,29	-18,21	-261,53	0,04	0,65	-0,39	-0,13	1,03
<b>12</b>	Prom. imobiliária (desenvolvimento de projectos de edifícios); construção de edifícios	0,69	0,39	150,86	0,05	0,15	-0,34	-0,07	1,07
<b>13</b>	Engenharia Civil	0,52	2,46	1,57	0,05	0,15	-0,19	-0,04	0,96
<b>14</b>	Actividades especializadas de construção	0,44	0,57	60,47	0,05	0,75	-0,12	-0,17	1,30
<b>15</b>	Comércio, manutenção e reparação, de veículos automóveis e motociclos	0,44	0,32	-35,33	0,03	0,98	-0,18	-0,13	1,78
<b>16</b>	Comércio por grosso (inclui agentes), excepto de veículos automóveis e motociclos	0,30	0,71	-2,18	0,05	0,01	-0,19	-0,18	1,60
<b>17</b>	Comércio a retalho, excepto de veículos automóveis e motociclos	1,35	6,69	-31,07	0,05	-0,38	-0,45	-0,08	1,69
<b>18</b>	Transportes e armazenagem	3,95	-0,78	21,20	0,06	-4,42	-0,09	-0,08	1,36
<b>19</b>	Alojamento, restauração e similares	1,28	-0,37	-15,55	0,05	0,28	-0,22	-0,20	2,00
<b>20</b>	Act. informação e de comunicação	0,38	1,28	-16,14	0,02	-0,26	-0,87	-0,08	1,30
<b>21</b>	Actividades imobiliárias	8,05	0,32	16,84	0,04	3,85	0,01	-0,02	0,78
<b>22</b>	Act. consultoria, científicas, técnicas e similares	1,20	1,15	28,66	0,03	0,27	-0,13	-0,01	1,25
<b>23</b>	Act. administrativas e dos serviços de apoio	0,57	0,92	2,37	0,04	-0,34	-0,18	-0,06	1,74
	<b>Média geral</b>	<b>0,99</b>	<b>0,76</b>	<b>-2,20</b>	<b>0,05</b>	<b>0,55</b>	<b>-0,33</b>	<b>-0,13</b>	<b>1,45</b>



## Anexo 5 – Resultados da regressão logística com variáveis explicativas não centradas

Model Fit Statistics			
Criterion	Intercept Only	Intercept and Covariates	
AIC	3334,652	2562,454	
SC	3340,437	2741,786	
-2 Log L	3332,652	2500,454	
R-Square	0,2926	Max-rescaled R-Square	0,3901

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	832,1973	30	<0,0001
Score	450,8974	30	<0,0001
Wald	436,5404	30	<0,0001

Deviance and Pearson Goodness-of-Fit Statistics				
Criterion	Value	DF	Value/DF	Pr > ChiSq
Deviance	2500,4543	2373	1,0537	0,0338
Pearson	2,20493E21	2373	9,292E17	<0,0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-1,3540	0,4744	8,1458	0,0043
AC_AT	1	0,9482	0,2271	17,4346	<0,0001
CD_AT	1	-0,8606	0,3491	6,0769	0,0137
CP_PT	1	-0,6291	0,0820	58,8198	<0,0001
FM_AT	1	-0,2318	0,0694	11,1429	0,0008
JDS_DAT	1	8,3681	1,0662	61,5964	<0,0001
RO_AT	1	-2,4127	0,2538	90,3572	<0,0001
ln_VN	1	0,0803	0,0332	5,8609	0,0155
VN_AT	1	-0,4250	0,0494	74,0099	<0,0001
Grupo 1	1	-0,4412	0,3360	1,7246	0,1891
Grupo 2	1	1,0831	0,4039	7,1900	0,0073
Grupo 3	1	1,6551	0,2316	51,0673	<0,0001
Grupo 4	1	1,1289	0,2816	16,0752	<0,0001
Grupo 5	1	0,8473	0,7987	1,1254	0,2888
Grupo 6	1	-0,00110	0,4189	0,0000	0,9979
Grupo 7	1	0,2854	0,3411	0,7000	0,4028
Grupo 8	1	0,1747	0,2614	0,4467	0,5039
Grupo 9	1	-0,4160	0,3299	1,5905	0,2073
Grupo 10	1	0,7070	0,6232	1,2872	0,2566
Grupo 11	1	0,0427	0,2497	0,0292	0,8643
Grupo 12	1	0,2737	0,1712	2,5541	0,1100
Grupo 13	1	0,8073	0,3129	6,6575	0,0099
Grupo 14	1	0,1531	0,2014	0,5784	0,4469
Grupo 15	1	-0,7775	0,2308	11,3485	0,0008
Grupo 16	1	-0,2433	0,1357	3,2139	0,0730
Grupo 17	1	-0,2864	0,1368	4,3813	0,0363
Grupo 18	1	-0,1192	0,2416	0,2433	0,6219

Analysis of Maximum Likelihood Estimates						
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	
<b>Grupo 19</b>	1	-1,1452	0,2627	19,0056	<0,0001	
<b>Grupo 20</b>	1	-0,7616	0,4128	3,4044	0,0650	
<b>Grupo 21</b>	1	-1,3793	0,3004	21,0867	<0,0001	
<b>Grupo 22</b>	1	-1,1261	0,2649	18,0665	<0,0001	

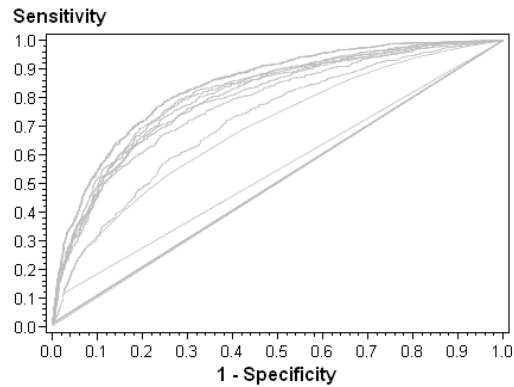
Association of Predicted Probabilities and Observed Responses			
<b>Percent Concordant</b>	83,9	<b>Somers' D</b>	0,680
<b>Percent Discordant</b>	15,8	<b>Gamma</b>	0,682
<b>Percent Tied</b>	0,3	<b>Tau-a</b>	0,340
<b>Pairs</b>	1444804	<b>c</b>	0,840

Wald Confidence Interval for Parameters			
Parameter	Estimate	95% Confidence Limits	
<b>Intercept</b>	-1,3540	-2,2839	-0,4242
<b>AC_AT</b>	0,9482	0,5031	1,3933
<b>CD_AT</b>	-0,8606	-1,5449	-0,1764
<b>CP_PT</b>	-0,6291	-0,7899	-0,4684
<b>FM_AT</b>	-0,2318	-0,3678	-0,0957
<b>JDS_DAT</b>	8,3681	6,2784	10,4579
<b>RO_AT</b>	-2,4127	-2,9102	-1,9152
<b>ln_VN</b>	0,0803	0,0153	0,1453
<b>VN_AT</b>	-0,4250	-0,5218	-0,3281
<b>Grupo 1</b>	-0,4412	-1,0997	0,2173
<b>Grupo 2</b>	1,0831	0,2914	1,8748
<b>Grupo 3</b>	1,6551	1,2011	2,1090
<b>Grupo 4</b>	1,1289	0,5770	1,6808
<b>Grupo 5</b>	0,8473	-0,7181	2,4127
<b>Grupo 6</b>	-0,00110	-0,8222	0,8200
<b>Grupo 7</b>	0,2854	-0,3831	0,9538
<b>Grupo 8</b>	0,1747	-0,3376	0,6870
<b>Grupo 9</b>	-0,4160	-1,0626	0,2305
<b>Grupo 10</b>	0,7070	-0,5144	1,9285
<b>Grupo 11</b>	0,0427	-0,4467	0,5320
<b>Grupo 12</b>	0,2737	-0,0620	0,6093
<b>Grupo 13</b>	0,8073	0,1941	1,4206
<b>Grupo 14</b>	0,1531	-0,2415	0,5478
<b>Grupo 15</b>	-0,7775	-1,2299	-0,3252
<b>Grupo 16</b>	-0,2433	-0,5092	0,0227
<b>Grupo 17</b>	-0,2864	-0,5545	-0,0182
<b>Grupo 18</b>	-0,1192	-0,5928	0,3544
<b>Grupo 19</b>	-1,1452	-1,6600	-0,6303
<b>Grupo 20</b>	-0,7616	-1,5706	0,0474
<b>Grupo 21</b>	-1,3793	-1,9680	-0,7906
<b>Grupo 22</b>	-1,1261	-1,6453	-0,6068

Partition for the Hosmer and Lemeshow Test					
Group	Total	incump = 1		incump = 0	
		Observed	Expected	Observed	Expected
1	240	11	9,97	229	230,03
2	240	31	41,23	209	198,77
3	240	60	68,82	180	171,18
4	240	74	92,49	166	147,51
5	240	103	112,52	137	127,48
6	240	141	131,52	99	108,48
7	240	162	151,02	78	88,98
8	240	191	171,27	49	68,73
9	240	202	194,86	38	45,14
10	244	227	228,30	17	15,70

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
25,3983	8	0,0013

Classification Table									
Prob Level	Correct		Incorrect		Percentages				
	Event	Non-Event	Event	Non-Event	Correct	Sensitivity	Specificity	False POS	False NEG
0.500	924	892	310	278	75,5	76,9	74,2	25,1	23,8



## Anexo 6 – Resultados da regressão logística com variáveis explicativas centradas considerando a média do grupo

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	3334,652	2562,454
SC	3340,437	2741,786
-2 Log L	3332,652	2500,454

R-Square	0,2926	Max-rescaled R-Square	0,3901
----------	--------	-----------------------	--------

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	832,1973	30	<0,0001
Score	450,8974	30	<0,0001
Wald	436,5404	30	<0,0001

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
8,6582	8	0,3719

Deviance and Pearson Goodness-of-Fit Statistics				
Criterion	Value	DF	Value/DF	Pr > ChiSq
Deviance	2500,4543	2373	1,0537	0,0338
Pearson	2,20493E21	2373	9,292E17	<0,0001

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
ln_VN_c	1	5,8609	0,0155
AC_AT_c	1	17,4346	<0,0001
CD_AT_c	1	6,0769	0,0137
FM_AT_c	1	11,1429	0,0008
CP_PT_c	1	58,8198	<0,0001
JDS_DAT_c	1	61,5964	<0,0001
RO_AT_c	1	90,3572	<0,0001
VN_AT_c	1	74,0099	<0,0001
grupo3	22	328,5547	<0,0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-0,1203	0,1026	1,3747	0,2410
ln_VN_c	1	0,0803	0,0332	5,8609	0,0155
AC_AT_c	1	0,9482	0,2271	17,4346	<0,0001
CD_AT_c	1	-0,8606	0,3491	6,0769	0,0137
FM_AT_c	1	-0,2318	0,0694	11,1429	0,0008
CP_PT_c	1	-0,6291	0,0820	58,8198	<0,0001
JCS_DAT_c	1	8,3681	1,0662	61,5964	<0,0001
RO_AT_c	1	-2,4127	0,2538	90,3572	<0,0001
VN_AT_c	1	-0,4250	0,0494	74,0099	<0,0001
Grupo 1	1	-1,0304	0,3337	9,5335	0,0020
Grupo 2	1	2,0586	0,4083	25,4192	<0,0001
Grupo 3	1	2,1600	0,2395	81,3710	<0,0001

Analysis of Maximum Likelihood Estimates						
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	
Grupo 4	1	2,1019	0,2984	49,6276	<0,0001	
Grupo 5	1	3,5871	0,8186	19,2024	<0,0001	
Grupo 6	1	0,5688	0,4203	1,8311	0,1760	
Grupo 7	1	0,8412	0,3483	5,8310	0,0157	
Grupo 8	1	0,5363	0,2701	3,9426	0,0471	
Grupo 9	1	-0,3875	0,3327	1,3562	0,2442	
Grupo 10	1	1,3212	0,6238	4,4865	0,0342	
Grupo 11	1	0,5044	0,2559	3,8870	0,0487	
Grupo 12	1	0,5114	0,1701	9,0356	0,0026	
Grupo 13	1	1,1418	0,3146	13,1749	0,0003	
Grupo 14	1	0,6668	0,2058	10,4968	0,0012	
Grupo 15	1	-0,7079	0,2327	9,2551	0,0023	
Grupo 16	1	0,3389	0,1422	5,6788	0,0172	
Grupo 17	1	-0,8367	0,1395	35,9691	<0,0001	
Grupo 18	1	-2,1412	0,3248	43,4545	<0,0001	
Grupo 19	1	-1,7497	0,2541	47,4235	<0,0001	
Grupo 20	1	-0,7966	0,4141	3,7000	0,0544	
Grupo 21	1	-6,1002	0,6562	86,4141	<0,0001	
Grupo 22	1	-1,8436	0,2666	47,8215	<0,0001	

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	83,9	Somers' D	0,680
Percent Discordant	15,8	Gamma	0,682
Percent Tied	0,3	Tau-a	0,340
Pairs	1444804	c	0,840

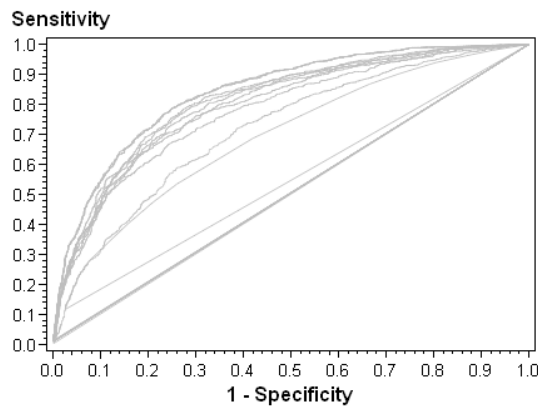
Wald Confidence Interval for Parameters			
Parameter	Estimate	95% Confidence Limits	
Intercept	-0,1203	-0,3213	0,0808
ln_VN_c	0,0803	0,0153	0,1453
AC_AT_c	0,9482	0,5031	1,3933
CD_AT_c	-0,8606	-1,5449	-0,1764
FM_AT_c	-0,2318	-0,3678	-0,0957
CP_PT_c	-0,6291	-0,7899	-0,4684
JCS_DAT_c	8,3681	6,2784	10,4579
RO_AT_c	-2,4127	-2,9102	-1,9152
VN_AT_c	-0,4250	-0,5218	-0,3281
Grupo 1	-1,0304	-1,6845	-0,3763
Grupo 2	2,0586	1,2583	2,8588
Grupo 3	2,1600	1,6907	2,6294
Grupo 4	2,1019	1,5171	2,6867
Grupo 5	3,5871	1,9827	5,1914
Grupo 6	0,5688	-0,2551	1,3926
Grupo 7	0,8412	0,1584	1,5239
Grupo 8	0,5363	0,00693	1,0658
Grupo 9	-0,3875	-1,0396	0,2647
Grupo 10	1,3212	0,0987	2,5438
Grupo 11	0,5044	0,00296	1,0059
Grupo 12	0,5114	0,1779	0,8448
Grupo 13	1,1418	0,5252	1,7583
Grupo 14	0,6668	0,2634	1,0701
Grupo 15	-0,7079	-1,1639	-0,2518
Grupo 16	0,3389	0,0602	0,6176

Wald Confidence Interval for Parameters				
Parameter	Estimate	95% Confidence Limits		
Grupo 17	-0,8367	-1,1101	-0,5632	
Grupo 18	-2,1412	-2,7778	-1,5046	
Grupo 19	-1,7497	-2,2477	-1,2517	
Grupo 20	-0,7966	-1,6082	0,0151	
Grupo 21	-6,1002	-7,3864	-4,8140	
Grupo 22	-1,8436	-2,3661	-1,3211	

Partition for the Hosmer and Lemeshow Test					
Group	Total	incump = 1		incump = 0	
		Observed	Expected	Observed	Expected
1	240	11	9,97	229	230,03
2	240	31	41,23	209	198,77
3	240	60	68,82	180	171,18
4	240	74	92,49	166	147,51
5	240	103	112,52	137	127,48
6	240	141	131,52	99	108,48
7	240	162	151,02	78	88,98
8	240	191	171,27	49	68,73
9	240	202	194,86	38	45,14
10	244	227	228,30	17	15,70

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
25.3983	8	0,0013

Classification Table									
Prob Level	Correct		Incorrect		Percentages				
	Event	Non-Event	Event	Non-Event	Correct	Sensitivity	Specificity	False POS	False NEG
0.500	924	892	310	278	75,5	76,9	74,2	25,1	23,8



Anexo 7 – Resultados do modelo *logit* multinível apenas com termo independente

Model Information	
Data Set	DADOS.AMOSTRA_VN_ALTERNATIVA
Response Variable	incump
Response Distribution	Binary
Link Function	Logit
Variance Function	Default
Variance Matrix Blocked By	Grupo
Estimation Technique	Residual PL
Degrees of Freedom Method	Satterthwaite

Class Level Information		
Class	Levels	Values
Grupo	23	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

Number of Observations Read	2404
Number of Observations Used	2404

Response Profile		
Ordered Value	incump	Total Frequency
1	0	1202
2	1	1202

The GLIMMIX procedure is modeling the probability that incump='1'.

Dimensions	
G-side Cov. Parameters	1
Columns in X	1
Columns in Z per Subject	1
Subjects (Blocks in V)	23
Max Obs per Subject	377

Optimization Information	
Optimization Technique	Dual Quasi-Newton
Parameters in Optimization	1
Lower Boundaries	1
Upper Boundaries	0
Fixed Effects	Profiled
Starting From	Data

Iteration History					
Iteration	Restarts	Subiterations	Objective Function	Change	Max Gradient
0	0	4	10584,120633	0,20834541	6,55E-6
1	0	2	10468,015445	0,00100138	1,404E-7
2	0	1	10470,270624	0,00008034	4,982E-7
3	0	1	10470,278502	0,00000171	0,000034
4	0	1	10470,278642	0,00000507	0,0001
5	0	1	10470,278229	0,00000377	4,639E-8
6	0	0	10470,278536	0,00000000	5,497E-7

Convergence criterion (PCONV=1.11022E-8) satisfied.

Fit Statistics	
-2 Res Log Pseudo-Likelihood	10470,28
Pseudo-AIC	10472,28
Pseudo-AICC	10472,28
Pseudo-BIC	10473,41
Pseudo-CAIC	10474,41
Pseudo-HQIC	10472,56
Generalized Chi-Square	2376,26
Gener. Chi-Square / DF	0,99

Fit statistics based on pseudo-likelihoods are not useful for comparing models that differ in their pseudo-data.

Covariance Parameter Estimates			
Cov Parm	Subject	Estimate	Standard Error
Intercept	Grupo	0,6570	0,2239

Solutions for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr >  t
Intercept	0,08471	0,1793	21,35	0,47	0,6413

Solution for Random Effects						
Effect	Subject	Estimate	Std Err Pred	DF	t Value	Pr >  t
Intercept	Grupo 1	-0,6029	0,3215	129,4	-1,88	0,0630
Intercept	Grupo 2	1,1567	0,3526	141,5	3,28	0,0013
Intercept	Grupo 3	1,3733	0,2549	75,96	5,39	<,0001
Intercept	Grupo 4	1,1361	0,2896	106,7	3,92	0,0002
Intercept	Grupo 5	0,7184	0,5332	79,03	1,35	0,1817
Intercept	Grupo 6	0,2353	0,3803	142,9	0,62	0,5371
Intercept	Grupo 7	0,4585	0,3408	138,2	1,35	0,1808
Intercept	Grupo 8	0,2093	0,2830	101,2	0,74	0,4613
Intercept	Grupo 9	-0,3960	0,3259	131,8	-1,22	0,2265
Intercept	Grupo 10	0,6894	0,4652	112	1,48	0,1411
Intercept	Grupo 11	0,2088	0,2758	94,88	0,76	0,4510
Intercept	Grupo 12	0,3268	0,2232	48,91	1,46	0,1495
Intercept	Grupo 13	0,7723	0,3093	121,8	2,50	0,0139
Intercept	Grupo 14	0,1038	0,2408	63,46	0,43	0,6681
Intercept	Grupo 15	-0,7705	0,2617	82,09	-2,94	0,0042
Intercept	Grupo 16	-0,1304	0,2039	35,12	-0,64	0,5267
Intercept	Grupo 17	-0,3709	0,2047	35,62	-1,81	0,0784
Intercept	Grupo 18	-0,3954	0,2616	81,99	-1,51	0,1344
Intercept	Grupo 19	-1,0589	0,2609	81,37	-4,06	0,0001
Intercept	Grupo 20	-0,6464	0,3627	143	-1,78	0,0769
Intercept	Grupo 21	-1,0836	0,2987	114,1	-3,63	0,0004
Intercept	Grupo 22	-1,3256	0,2718	91,24	-4,88	<,0001
Intercept	grupo 23	-0,6082	0,2914	108,3	-2,09	0,0392



## Anexo 8 – Resultados do modelo *logit* multinível com as variáveis que resultaram estatisticamente significativas na regressão logística

Model Information	
Data Set	DADOS.AMOSTRA_VN_ALTERNATIVA
Response Variable	incump
Response Distribution	Binary
Link Function	Logit
Variance Function	Default
Variance Matrix Blocked By	Grupo
Estimation Technique	Residual PL

Class Level Information	
Class	Levels
Grupo	23
	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

Number of Observations Read	2404
Number of Observations Used	2404

Response Profile		
Ordered Value	incump	Total Frequency
1	0	1202
2	1	1202

The GLIMMIX procedure is modeling the probability that incump='1'.

Dimensions	
G-side Cov. Parameters	9
Columns in X	1
Columns in Z per Subject	9
Subjects (Blocks in V)	23
Max Obs per Subject	377

Optimization Information	
Optimization Technique	Dual Quasi-Newton
Parameters in Optimization	9
Lower Boundaries	9
Upper Boundaries	0
Fixed Effects	Profiled
Starting From	Data

Iteration History					
Iteration	Restarts	Subiterations	Objective Function	Change	Max Gradient
0	0	29	10604,567327	2,00000000	530,5509
1	0	29	1,340781E154	0,00000000	,

Did not converge.

---

<b>Covariance Parameter Estimates</b>			
<b>Cov Parm</b>	<b>Subject</b>	<b>Estimate</b>	<b>Standard Error</b>
<b>Intercept</b>	<b>grupo</b>	0,2599	,
<b>AC_AT</b>	<b>grupo</b>	3,42E-19	,
<b>CD_AT</b>	<b>grupo</b>	2,6086	,
<b>FM_AT</b>	<b>grupo</b>	0,02485	,
<b>CP_PT</b>	<b>grupo</b>	2,16E-21	,
<b>JDS_DAT</b>	<b>grupo</b>	12,4094	,
<b>RO_AT</b>	<b>grupo</b>	1,4953	,
<b>VN_AT</b>	<b>grupo</b>	0,04274	,
<b>ln_VN</b>	<b>grupo</b>	0,000806	,

Anexo 9 – Resultados do modelo *logit* multinível com a variável JCS\_DAT

Model Information	
Data Set	DADOS.AMOSTRA_VN_ALTERNATIVA
Response Variable	incump
Response Distribution	Binary
Link Function	Logit
Variance Function	Default
Variance Matrix Blocked By	Grupo
Estimation Technique	Residual PL

Class Level Information	
Class	Levels
Grupo	23

Number of Observations Read	2404
Number of Observations Used	2404

Response Profile		
Ordered Value	incump	Total Frequency
1	0	1202
2	1	1202

The GLIMMIX procedure is modeling the probability that incump='1'.

Dimensions	
G-side Cov. Parameters	2
Columns in X	1
Columns in Z per Subject	2
Subjects (Blocks in V)	23
Max Obs per Subject	377

Optimization Information	
Optimization Technique	Dual Quasi-Newton
Parameters in Optimization	2
Lower Boundaries	2
Upper Boundaries	0
Fixed Effects	Profiled
Starting From	Data

Iteration History					
Iteration	Restarts	Subiterations	Objective Function	Change	Max Gradient
0	0	6	10594,410319	0,94714266	0,002899
1	0	4	10488,548426	2,00000000	0,00026
2	0	4	10499,941029	1,37556411	2,375E-6
3	0	1	10502,260398	0,00106862	0,002535
4	0	2	10502,289419	0,09694337	0,000145
5	0	1	10502,458761	0,00025372	0,000797
6	0	1	10502,456315	0,00030442	0,0001
7	0	1	10502,458692	0,00024362	0,001216
8	0	2	10502,454654	0,01129228	8,719E-6
9	0	0	10502,437416	0,00000000	3,868E-6

Convergence criterion (PCONV=1,11022E-8) satisfied.

Fit Statistics	
-2 Res Log Pseudo-Likelihood	10502,44
Pseudo-AIC	10506,44
Pseudo-AICC	10506,44
Pseudo-BIC	10508,71
Pseudo-CAIC	10510,71
Pseudo-HQIC	10507,01
Generalized Chi-Square	2358,58
Gener. Chi-Square / DF	0,98

Fit statistics based on pseudo-likelihoods are not useful for comparing models that differ in their pseudo-data.

Covariance Parameter Estimates			
Cov Parm	Subject	Estimate	Standard Error
Intercept	grupo3	0,6392	0,2282
JCS_DAT	grupo3	28,2688	19,2186

Solutions for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr >  t
Intercept	-0,00317	0,1803	20,91	-0,02	0,9861

Solution for Random Effects						
Effect	Subject	Estimate	Std Err Pred	DF	t Value	Pr >  t
Intercept	Grupo 1	-0,6267	0,3489	116,2	-1,80	0,0751
JCS_DAT	Grupo 1	2,4016	3,2579	30,63	0,74	0,4666
Intercept	Grupo 2	1,2553	0,3886	82,49	3,23	0,0018
JCS_DAT	Grupo 2	-0,7596	4,5047	8,386	-0,17	0,8701
Intercept	Grupo 3	1,2650	0,2797	81,02	4,52	<,0001
JCS_DAT	Grupo 3	4,9923	3,5655	21,38	1,40	0,1758
Intercept	Grupo 4	1,0715	0,3365	75,91	3,18	0,0021
JCS_DAT	Grupo 4	3,0786	3,9964	13,52	0,77	0,4543
Intercept	Grupo 5	0,7364	0,5358	67,36	1,37	0,1739
JCS_DAT	Grupo 5	1,4138	5,1584	4,883	0,27	0,7952
Intercept	Grupo 6	0,2959	0,4226	85,34	0,70	0,4856
JCS_DAT	Grupo 6	0,1929	4,0980	12,22	0,05	0,9632
Intercept	Grupo 7	0,4153	0,3682	91,9	1,13	0,2623
JCS_DAT	Grupo 7	3,7569	4,7497	6,79	0,79	0,4557
Intercept	Grupo 8	0,01803	0,3264	75,85	0,06	0,9561
JCS_DAT	Grupo 8	6,4320	4,0753	12,51	1,58	0,1394
Intercept	Grupo 9	-0,4783	0,3560	85,78	-1,34	0,1827
JCS_DAT	Grupo 9	5,0476	4,7027	7,065	1,07	0,3184
Intercept	Grupo 10	0,6895	0,4881	75,5	1,41	0,1618
JCS_DAT	Grupo 10	1,7038	5,0280	5,407	0,34	0,7475
Intercept	Grupo 11	0,3554	0,3136	74,65	1,13	0,2608
JCS_DAT	Grupo 11	-1,7485	4,1504	11,64	-0,42	0,6812
Intercept	Grupo 12	0,4682	0,2510	65,17	1,87	0,0667
JCS_DAT	Grupo 12	-1,2181	2,5761	78,31	-0,47	0,6376
Intercept	Grupo 13	0,6221	0,3575	71,29	1,74	0,0861
JCS_DAT	Grupo 13	5,0853	4,3066	10,03	1,18	0,2649
Intercept	Grupo 14	-0,08083	0,2750	74,48	-0,29	0,7696
JCS_DAT	Grupo 14	6,3531	3,3306	28,05	1,91	0,0667
Intercept	Grupo 15	-0,8024	0,2892	74,26	-2,77	0,0070
JCS_DAT	Grupo 15	3,8801	4,2585	10,51	0,91	0,3826
Intercept	Grupo 16	-0,3962	0,2279	48,63	-1,74	0,0884
JCS_DAT	Grupo 16	7,9272	2,4254	99,73	3,27	0,0015
Intercept	Grupo 17	-0,2314	0,2153	40,86	-1,07	0,2887

<b>Solution for Random Effects</b>							
<b>Effect</b>	<b>Subject</b>	<b>Estimate</b>	<b>Std Err</b>	<b>Pred</b>	<b>DF</b>	<b>t Value</b>	<b>Pr &gt;  t </b>
<b>JCS_DAT</b>	<b>Grupo 17</b>	-1,2032	1,5614	580,6	-0,77	0,4413	
<b>Intercept</b>	<b>Grupo 18</b>	-0,2753	0,2841	93,66	-0,97	0,3351	
<b>JCS_DAT</b>	<b>Grupo 18</b>	-0,6929	2,1833	151,5	-0,32	0,7514	
<b>Intercept</b>	<b>Grupo 19</b>	-0,8849	0,2771	87,34	-3,19	0,0020	
<b>JCS_DAT</b>	<b>Grupo 19</b>	-2,3545	2,9357	46,52	-0,80	0,4266	
<b>Intercept</b>	<b>Grupo 20</b>	-0,6062	0,3730	119,5	-1,63	0,1067	
<b>JCS_DAT</b>	<b>Grupo 20</b>	1,9626	5,0580	5,283	0,39	0,7131	
<b>Intercept</b>	<b>Grupo 21</b>	-1,0798	0,3352	91,14	-3,22	0,0018	
<b>JCS_DAT</b>	<b>Grupo 21</b>	1,9368	3,8705	15,38	0,50	0,6239	
<b>Intercept</b>	<b>Grupo 22</b>	-1,2571	0,2961	85,59	-4,25	<,0001	
<b>JCS_DAT</b>	<b>Grupo 22</b>	0,5186	4,0282	13,13	0,13	0,8995	
<b>Intercept</b>	<b>Grupo 23</b>	-0,4734	0,3265	82,04	-1,45	0,1508	
<b>JCS_DAT</b>	<b>Grupo 23</b>	-1,4546	4,2108	10,99	-0,35	0,7363	