

Características das firmas inovadoras no Estado de São Paulo: uma análise empírica a partir da PAEP*

Sérgio Kannebley Júnior[§]

RESUMO

Este artigo reporta um esforço de caracterização das empresas inovadoras do Estado de São Paulo. Esta caracterização foi realizada com base em informações financeiras, patrimoniais e de recursos humanos das empresas constantes na PAEP (Pesquisa de Atividade Econômica Paulista). Para isso foram empregados procedimentos estatísticos não-paramétricos, que informaram, em ordem decrescente, que os três principais fatores distintivos entre empresas inovadoras e não inovadoras são: a orientação exportadora, o tamanho da empresa e a variação interindustrial. De modo menos incisivo, também contribuem para a distinção entre empresas inovadoras e não inovadoras o grau de concentração do mercado e a origem do capital.

Palavras-chave: inovação, economia industrial, metodologia de coleta, estimativa e organização de dados microeconômicos.

ABSTRACT

This article reports an effort to characterize São Paulo's innovative firms. This characterization was done based on PAEP (Paulista Economic Activity Research) information. It was employed statistical non-parametrical procedures, which informs that the three distinctive main factors are the export orientation, the size of the firm, and the industrial sector effect. Besides, it was showed that the market concentration, and the foreign capital origin, are important to distinct between innovators and non innovators firms, but in a less incisive way.

Key words: innovation, industrial economics, methodology for collecting, estimating, and organizing microeconomic data.

JEL classification: O30, L00, C81.

* Agradecimentos à Fundação SEADE, especificamente a Noeli Pereira, pela colaboração e disponibilização das informações para a realização desse trabalho, e à assistência dos auxiliares de pesquisa Harry Richard Hamming e Paulo Rogério Rodrigues Maduro Jr. Qualquer erro, ou omissão, é de responsabilidade exclusiva do autor.

§ Professor FEA-RP/USP.

Recebido em junho de 2003. Aceito em maio de 2004.

Introdução

A literatura empírica sobre os fatores que influenciam, ou que caracterizam, a atividade inovadora das empresas está bastante desenvolvida. As resenhas de Cohen e Levin (1989) e Cohen (1995) apresentam diversos argumentos teóricos e resultados obtidos em testes empíricos para este tema. Usualmente a realização de atividades inovadoras, ou a intensidade com que são realizadas, é explicada pelas características das firmas, das estruturas de mercados, da variação interindustrial, das condições de apropriabilidade e de demanda.

A grande maioria dos trabalhos expostos na literatura refere-se ao exame dos determinantes de atividades de Pesquisa e Desenvolvimento (P&D) das firmas tecnologicamente ativas em países industrializados. Kumar e Siddhartan (1997) apresentam resultados da literatura empírica que demonstram que essa caracterização de atividade inovadora não é representativa para as firmas de países em desenvolvimento. Esses resultados destacam a importância da compra de P&D externo, bem como a realização de progresso técnico por meio da importação de bens de capitais, resultados esses que foram confirmados, recentemente, para o caso brasileiro, por meio das pesquisas de atividade tecnológica contidas na PAEP (Pesquisa de Atividade Econômica Paulista) e na PINTEC (Pesquisa Industrial de Inovação e Tecnologia).

As informações constantes na PAEP estão em concordância com essas evidências, já que demonstram que aproximadamente 78,5% das empresas que declararam não terem realizado inovações de produto e/ou processo no período de 1994 a 1996, sejam elas incrementais ou significativas, também não efetuaram importações de bens de capital estrangeiros. Por outro lado, aproximadamente 57,4% das empresas que afirmaram terem realizado inovações de produto e/ou processo importaram bens de capital no mesmo período. Também de acordo com as informações da PAEP, os gastos com atividades de P&D interno corresponderam por apenas 16,5% do total de gastos em atividades inovadoras das firmas paulistas em 1996.¹

Sabendo dessas características distintas do processo inovador em países em desenvolvimento, e particularmente no caso brasileiro, este trabalho se propõe a analisar as características das firmas inovadoras, e não apenas aquelas realizadoras de atividades de P&D.

1 Essas informações estão em conformidade com aquelas fornecidas pela PINTEC ao relatar que aproximadamente 77% das empresas inovadoras atribuíram alta ou média importância para a atividade inovadora de aquisição de máquinas e equipamentos, correspondendo a mais de 50% dos gastos em atividades inovadoras. Também segundo a PINTEC, as atividades internas de P&D foram classificadas como de média ou alta importância por apenas 34,1% das empresas brasileiras, sendo que os gastos relativos a essa atividade correspondem a 16,7% do total dos gastos em atividades inovadoras das empresas brasileiras no ano de 2000.

Será adotada aqui uma caracterização ampla para a atividade inovadora, ficando aberto o caminho para trabalhos futuros que desejem trabalhar com categorias mais restritas de inovação. A determinação das características das firmas inovadoras em contraposição às firmas não inovadoras deve fornecer subsídios à análise de indicadores de atividade inovadora e tecnológica e também para a formulação de políticas. É essa a proposta final deste trabalho, e para isso são utilizadas informações constantes na PAEP para o período de 1994 a 1996. As variáveis selecionadas para análise dizem respeito às diversas características financeiras, patrimoniais e de recursos humanos das empresas.

O trabalho é composto, além dessa introdução, de mais quatro seções. Na primeira seção são descritas e justificadas as variáveis a serem utilizadas nos modelos estatísticos. Na segunda é realizada uma breve análise das informações constantes na pesquisa. A seguir são apresentados a metodologia de árvores de classificação QUEST e os resultados fornecidos pela estimação das árvores. Por fim, na última seção são tecidas algumas considerações finais sobre a análise realizada.

1 Definição das variáveis

Com base na literatura teórica e empírica sobre o tema, são utilizadas oito variáveis explicativas que buscam retratar as empresas segundo suas características financeiras, patrimoniais e de recursos humanos.

1.1 Tamanho da firma

Esta variável é usualmente associada à hipótese schumpeteriana de relação positiva, mas não necessariamente linear ou contínua, entre atividade inovadora e tamanho da firma. Vários argumentos são tecidos para se justificar o efeito positivo do tamanho da firma sobre a atividade inovadora, dentre as quais se destacam: i) a imperfeição do mercado de crédito (grandes firmas teriam mais facilidade para financiar projetos de P&D de risco visto que tamanho é correlacionado com a disponibilidade e estabilidade de fundos internos para investimento); ii) a existência de economias de escala na função de P&D propriamente dita, considerando que os retornos de P&D são mais altos quando o inovador tem um maior volume de vendas sobre o qual consegue diluir mais facilmente os custos fixos da inovação, particularmente nos casos de inovação de processo; iii) P&D é dito ser mais produtivo nas grandes empresas como resultado da complementaridade entre P&D e outras atividades não-manufatureiras (*marketing* e planejamento) existentes em maior

grau nestas empresas; iv) firmas maiores, diversificadas propiciam economias de escopo ou reduzem o risco associado à inovação.²

Por outro lado, existem contra-argumentos relacionados à perda de eficiência da atividade de P&D. Argumenta-se que quando a firma se torna muito grande, a eficiência em P&D é prejudicada por causa da perda de controle gerencial ou, alternativamente, devido ao aumento excessivo da burocracia que acaba por desviar a atenção dos cientistas. Outro ponto a ser destacado é a diminuição da habilidade dos cientistas e empresários em capturar os benefícios de seus esforços individuais, podendo também diminuir os incentivos destes na busca da inovação.

As evidências empíricas ainda que favoreçam a hipótese schumpeteriana, apresentam resultados ambíguos em razão da definição da variável dependente. Em alguns estudos a variável dependente é o total de gastos em P&D, enquanto que em outros diz respeito à intensidade de P&D, dada pela razão entre o total de gastos de P&D e o total de vendas da empresa. Adicionalmente, Cohen (1995) relata problemas relacionados ao viés de seleção das amostras, compostas em sua maioria por grandes empresas e inovadoras e a desconsideração de fatores explicativos específicos à indústria e à empresa.

1.2 Estrutura de mercado

Também de acordo com a tradição schumpeteriana, postula-se uma relação positiva entre estruturas de mercado mais concentradas e a atividade inovadora, principalmente as atividades de P&D. O argumento básico é que a estrutura de mercado perfeitamente competitiva não incentiva atividades de P&D em razão da baixa apropriabilidade dos resultados dessas atividades, que não favorecem a obtenção de lucros extraordinários pelas empresas inovadoras. Esses lucros extraordinários seriam, por consequência, a fonte de financiamento para novas atividades, gerando um processo de retroalimentação da estrutura de mercado.

A verificação empírica desse postulado é usualmente conduzida por meio da estimação da relação entre um índice de concentração de vendas no mercado e o comportamento inovador da empresa. Conforme antes argumentado, esta relação não está livre da ambigüidade no que diz respeito ao seu sentido causal, reconhecendo-se a possibilidade de simultaneidade entre inovação e estrutura de mercado. Também são ambíguas as evidências empíricas produzidas pela literatura sobre o tema.

2 Chandler (1992) destaca a habilidade das novas grandes empresas intensivas em capital em explorar as vantagens de custos proporcionadas pelas economias de escala e escopo. Por meio dessas vantagens de custos essas firmas encontram condições facilitadas para expandir seus mercados, ou geográficos ou em novos produtos.

1.3 Orientação exportadora

Deve-se esperar que a exposição à competição do mercado internacional incentive a firma a investir em atividades inovadoras. Isso seria feito com o intuito de atender à demanda mais exigente e às pressões competitivas mais intensas enfrentadas no mercado internacional. Argumenta-se também que as exportações, ao ampliarem os mercados das firmas, aumentam os retornos da atividade inovadora na medida em que diluem seus custos. Braga e Willmore (1991) entendem que provavelmente a primeira linha de argumentação é a que melhor se adequa ao caso brasileiro em razão do Brasil ser um país em desenvolvimento. Em seu estudo encontram uma relação positiva e estatisticamente significativa entre a probabilidade de existência de atividade tecnológica e a orientação exportadora das empresas.

Também é importante lembrar que essa relação pode estar sujeita a um problema de simultaneidade, na medida em que a atividade tecnológica também pode aprimorar a capacidade competitiva, permitindo que a empresa enfrente os mercados internacionais mais competitivos.

1.4 Aprendizado acumulado

Considerando os aspectos específicos à firma, uma variável representativa desses aspectos é o tempo de constituição da empresa. É possível argumentar que empresas com maior tempo de constituição acumularam, ao longo do tempo, um aprendizado em atividades inovadoras que reduzem o custo do investimento em novas atividades inovadoras. Desse modo, é postulada uma relação positiva entre o tempo de constituição da empresa e o desempenho de atividades inovadoras, baseada no argumento da acumulação de aprendizado ou conhecimento específico à atividade realizada, além da institucionalização da atividade de pesquisa em empresas grandes já estabelecidas.

Henderson (1993) leva em conta esta hipótese ao distinguir o comportamento das empresas novas (entrantes) e já estabelecidas diante de diferentes formas de inovação. Segundo a autora, a hipótese apresentada acima se aplica mais adequadamente ao caso das inovações incrementais, na medida em que essa forma de inovação baseia-se mais fortemente no conhecimento e nas capacidades previamente adquiridas. Entretanto, ante a possibilidade de inovações radicais, o extensivo conhecimento em uma tecnologia poderia ser uma desvantagem em potencial. Isso ocorreria se essas inovações fossem fundadas em princípios científicos e/ou tecnológicos diversos dos até então utilizados. Nesse caso, as firmas entrantes, em razão das diferenças nas capacidades informacionais, estariam em

vantagem, realizando um esforço tecnológico superior ao das empresas já estabelecidas no mercado.

1.5 Qualificação do trabalho

Ainda que não diretamente relacionado à atividade inovadora, a qualificação do trabalhador é um fator relevante na capacidade de execução de tarefas de maior grau de complexidade existentes em firmas tecnologicamente mais avançadas. A existência de trabalhadores qualificados é complementar à execução de atividades inovadoras. Sem os primeiros, muitos dos esforços empreendidos no desenvolvimento de novos processos e/ou produtos poderiam estar comprometidos pela perda de eficiência e qualidade da produção. Sendo assim, espera-se que exista uma relação positiva entre o grau de qualificação do fator trabalho e a execução de atividades inovadoras.

1.6 Origem do capital controlador

As empresas multinacionais são detentoras de um conjunto de ativos intangíveis, como marcas internacionalmente reconhecidas, acesso cativo à tecnologia e capacitações gerenciais e organizacionais. Em razão disso, é possível que a estratégia competitiva da empresa multinacional em um país local seja uma estratégia de rivalidade não baseada em preços. Essa estratégia de rivalidade não baseada em preços estaria apoiada em gastos de *marketing*, controle de qualidade e desenvolvimento de produtos, além de um conjunto de serviços ao consumidor.

De acordo com esse arcabouço competitivo, é possível conceber-se uma relação positiva entre a execução de atividades tecnológicas e a natureza estrangeira do capital. Esta relação deve ser mais forte em atividades relacionadas ao desenvolvimento e/ou adaptação de produtos para o mercado doméstico. É também importante ressaltar que Braga e Willmore (1991) encontraram evidências de um efeito positivo da propriedade estrangeira do capital sobre a probabilidade das firmas possuírem um programa sistemático de desenvolvimento de produtos na indústria brasileira.

1.7 Apropriabilidade

A que extensão os novos conhecimentos podem ser transmitidos a relativamente baixos custos do seu criador para outros competidores e, particularmente, a que extensão esse conhecimento, incorporado em novos processos e produtos, pode ser copiado ou

imitado a relativamente baixos custos determinam se as recompensas apropriáveis são, ou não, suficientes para justificar o esforço inovador.

Considerando que a efetividade das patentes difere sensivelmente entre as indústrias, a possibilidade de internalização do conhecimento, ao invés do licenciamento, pode ser uma estratégia recompensadora para as empresas que investem em pesquisa ou mudanças tecnológicas.³ Sendo assim, uma empresa com maior proporção do valor adicionado de suas vendas deve ter uma probabilidade maior de investir em atividades de P&D. Este argumento deve ser mais apropriado para o caso de grandes empresas, já que este processo pode ser extremamente custoso ou inviável para pequenas empresas. Ainda assim, deve-se esperar uma relação positiva entre a realização de atividades inovadoras e uma variável dada pela razão entre o valor adicionado e a receita líquida obtida pela empresa.

1.8 Efeitos setoriais

Verifica-se, usualmente, uma diferença interindustrial no grau em que as indústrias se engajam em atividades inovadoras. Os argumentos que tentam explicar essas diferenças estão, na maior parte, relacionados às diferentes oportunidades de progresso tecnológico que as indústrias percebem.

Em termos empíricos, desconsiderar a capacidade explicativa das diferenças interindustriais constitui um problema de viés por omissão de variável relevante. Este problema é freqüentemente encontrado em estudos que procuram analisar as relações entre tamanho, grau de concentração e P&D, levando a estimativas enviesadas do efeito do tamanho ou da estrutura de mercado sobre a atividade inovadora. Nesse trabalho buscar-se-á capturar estes efeitos por meio da categorização dos setores segundo as oportunidades tecnológicas ou de acordo com a intensidade do fator de produção do setor industrial.⁴ A definição dos setores segundo suas oportunidades tecnológicas foi realizada de modo semelhante àquela feita por Quadros *et alii* (2001).⁵

3 Segundo Cohen (1995), a principal razão citada para a limitação da efetividade das patentes é capacidade de os competidores inventarem “legalmente em torno das patentes”

4 A taxonomia de oportunidades tecnológicas foi criada por Pavitt (1984). Esta taxonomia está relacionada ao aparecimento e ao desenvolvimento dos paradigmas tecnológicos entre os diversos setores industriais. Segundo Dosi (1988), o aparecimento de novos paradigmas é desigualmente distribuído entre os setores, sendo também desigualmente distribuídas as dificuldades técnicas para o avanço da eficiência da produção e da performance do produto, bem como a competência tecnológica para inovar. No entanto, argumenta que essa distribuição não é aleatória, mas sim dependente da natureza tecnológica dos setores, da distância que a tecnologia empregada está do “centro revolucionário do paradigma” e da base de conhecimento subjacente à inovação em cada setor.

5 Essa classificação foi realizada com base no esforço tecnológico realizado pelas empresas industriais de São Paulo, relacionado à distribuição do volume de P&D realizado entre os setores. Reclassificados segundo a divisão CNAE, os setores com altas oportunidades tecnológicas são as divisões 11, 30, 32, 33 e 35, com médias oportunidades as divisões 28, 29, 31 e 34, os de média-baixa oportunidade as divisões 21, 23, 24, 25, 26 e 27, e por fim os com baixas oportunidades, as divisões 10, 13, 14, 15, 16, 17, 18, 19, 20 e 22.

2 Análise descritiva das características das empresas inovadoras do Estado de São Paulo – PAEP (1996)

2.1 Base de dados

A PAEP foi realizada pela Fundação Sistema Estadual de Análise de Dados (FSeade). É uma pesquisa de caráter amostral, tendo 1996 como o ano-base para a sua realização e o Estado de São Paulo como o universo de análise. A base expandida contém informações de 41.658 empresas industriais e fornece um banco de dados para a elaboração de pesquisas e análises acerca das transformações técnico-produtivas em curso na economia paulista.⁶ Produz não só dados de mensuração da atividade econômica – como, por exemplo, número de unidades, valor da produção, pessoal ocupado –, mas também indicadores que permitem avaliar a extensão dos recentes processos de reestruturação e seus impactos nos diferentes setores da economia paulista. A disponibilização de microdados de empresas no ramo da indústria permite a mensuração da estrutura econômica paulista, inclusive no que tange às atividades inovadoras das empresas paulistas.

Antes de iniciar a análise propriamente dita é importante destacar algumas limitações e vantagens dessa base de dados. Uma primeira limitação deve-se ao fato da pesquisa ser em um único ponto do tempo, isto é, uma “*cross-section*” Isto não permite que sejam capturadas as relações dinâmicas da atividade inovadora postuladas pela teoria econômica. Também é um fator prejudicial à análise o fato das questões sobre a condução de atividades inovadoras se referirem ao período de 1994 a 1996, enquanto que as demais variáveis que são utilizadas para a caracterização das empresas se referirem ao ano de 1996.⁷ A grande vantagem que o trabalho com a base expandida fornece é a disponibilização de microdados das empresas que combinem as diversas informações de suas características com aquelas sobre suas atividades inovadoras. Adicionalmente, a disponibilidade de informações para todos os tamanhos de empresas evita a produção de resultados **enviesados** em direção às grandes empresas industriais.

6 Para mais informações, ver www.seade.gov.br.

7 Nesse caso, é necessário se admitir a existência de uma relativa inércia na estrutura das empresas e setorial que diminua o poder transformador das atividades inovadoras para aquelas empresas que as realizaram no início do período.

2.2 Análise descritiva

Na Tabela 1 são apresentadas informações sobre o número de empresas que realizaram ou não alguma atividade inovadora para a indústria de transformação.⁸ Esta tabela também contém informações acerca da taxa de inovação do setor, discriminadas segundo a forma de inovação produzida no período, além de um índice de concentração setorial para o Estado de São Paulo para o ano de 1996. As empresas industriais paulistas que produziram inovações de processo ou produto, incrementais ou significativas, no período de 1994 a 1996, apresentaram uma taxa média de inovação, para o Estado, em torno de 31%.⁹ Os setores com as cinco maiores taxas de inovação foram os de Fabricação de Máquinas de Escritório e Equipamentos de Informática (30), Fabricação de Equipamentos de Instrumentação Médico-Hospitalares (33), Fabricação de Material Eletrônico e Aparelhos de Comunicações (32), Fabricação de Outros Equipamentos de Transporte (35), Fabricação de Produtos Químicos (24). Já os setores com as cinco menores taxas de inovação foram os de Fabricação de Celulose, Papel e Produtos de Papel (21), Fabricação de Couros e Calçados (19), Edição, Impressão e Reprodução de Gravações (22), Confecção de Artigos de Vestuário e Acessórios (18) e de Fabricações de Produtos de Madeira (20). Uma análise atenta permite igualmente perceber a existência de uma relação positiva entre a medida de concentração industrial no Estado e a taxa de inovação setorial.¹⁰ É possível verificar que existe, na maioria dos setores, a prevalência de empresas inovadoras em processo e produto, 63% da taxa média de inovação ocorre na forma de inovação de produto e processo, enquanto que apenas 24% correspondem a inovações de produto e 13% a inovações de processo.

8 As informações referentes à indústria extrativa não foram apresentadas aqui em razão de sigilo estatístico.

9 Estritamente, na PAEP define-se a realização de INOVAÇÃO DE PRODUTO (NATUREZA INCREMENTAL) como aquela correspondente a um substancial aperfeiçoamento de um produto previamente existente; realização de INOVAÇÃO DE PRODUTO (NATUREZA SIGNIFICATIVA) aquela correspondente à introdução de um produto inteiramente novo, com características tecnológicas ou de uso e finalidade que o distinguem daqueles produzidos até então e a realização de INOVAÇÃO DE PROCESSO como correspondente à incorporação de um novo processo de produção ou às modificações tecnológicas em processos já adotados.

10 O coeficiente de correlação apurado para essa relação é de 0,11. É importante também lembrar que essa medida não é a mais apropriada para refletir a relação teórica estabelecida acima entre estrutura de mercado e atividade inovadora, já que não corresponde à estrutura nacional do setor, nem captura a contestabilidade do setor realizada pela possibilidade de comércio internacional. Porém, considerando a representatividade da indústria paulista sobre a indústria nacional, não deixa de ser uma medida referencial importante.

Tabela 1
Empresas Não Inovadoras e Inovadoras Segundo a Divisão CNAE

DIVISÃO CNAE	NÃO INOVA	INOVA	PROCESSO (%)	PRODUTO (%)	PROD. e PROC. (%)	TAXA de INOVAÇÃO	CR4
15	1060	383	16	21	62	27	0.2
17	1489	458	11	27	62	24	
18	4623	765	11	25	63	14	0.2
19	1219	327	13	31	56	21	0.1
20	802	82	6	64	30	9	0.3
21	691	215	17	29	54	24	0.4
22	1910	519	20	8	72	21	0.3
23	48	19	41	0	59	29	0.3
24	940	651	11	19	70	41	0.3
25	1654	827	13	31	56	33	0.2
26	1099	421	9	37	54	28	0.3
27	784	335	16	12	72	30	0.2
28	2569	1027	23	20	57	29	0.4
29	1841	955	11	21	69	34	0.1
30	20	52	6	27	67	72	0.2
31	823	327	10	20	69	28	0.6
32	261	201	10	19	71	43	0.3
33	317	248	4	20	76	44	0.6
34	599	341	14	19	68	36	0.5
35	64	44	9	30	61	41	0.7
36	2374	884	9	24	67	27	0.6

CR4 é a medida proporcional das receitas das quatro maiores empresas industriais da divisão.

Na Tabela 2 são apresentadas informações relativas ao número e ao percentual de empresas inovadoras e não inovadoras segundo suas faixas de tamanho e origem do capi-

tal.¹¹ Como é possível observar, o percentual de empresas não inovadoras é decrescente à medida que aumenta a faixa de tamanho, independentemente da origem do capital da empresa. Ainda que o inverso não ocorra de modo exato para as empresas inovadoras, é possível perceber que o percentual de empresas tende a ser maior à medida que aumenta o tamanho das empresas. A exceção a essa tipificação é a categoria de empresas nacionais, em que, de modo similar às empresas não inovadoras nacionais, o percentual de empresas inovadoras é decrescente à medida que aumenta a faixa de tamanho. No caso das empresas estrangeiras e mistas, este percentual é menor na faixa de tamanho de microempresas, atingindo um máximo na faixa de médias empresas e declinando na faixa de grandes empresas. Ainda assim, é importante notar que o número de empresas inovadoras nacionais é bastante superior ao número de empresas estrangeiras e mistas inovadoras.

Tabela 2
Empresas Inovadoras Segundo a Faixa de Tamanho e Origem do Capital

	NACIONAL		ESTRANGEIRA		MISTO	
	número	%	número	%	número	%
NÃO INOVA						
MICRO	16822	67,7	58	24,9	37	40,0
PEQUENA	6759	27,2	76	32,3	33	36,3
MÉDIA	1133	4,6	75	32,1	20	21,9
GRANDE	138	0,6	25	10,7	2	1,8
INOVA						
MICRO	3900	45,2	29	12,3	3	2,8
PEQUENA	3166	36,7	70	30,0	46	50,7
MÉDIA	1303	15,1	113	48,2	56	61,8
GRANDE	264	3,1	94	40,0	35	37,9

11 As faixas de tamanho são definidas segundo o número total de pessoas ocupadas. É considerada uma microempresa aquela que possui de 0 a 19 empregados, pequena empresa, aquela que possui de 20 a 99 empregados, média empresa, aquela com 100 a 499 empregados, e grande empresa, aquelas com mais de 500 empregados.

Na Tabela 3 é apresentada uma extensão da tabulação realizada na Tabela 2. Nessa tabulação as empresas são classificadas primeiramente de acordo com a origem do seu capital em contraposição ao fato de terem realizado ou não atividade inovadora. Posteriormente, as firmas são reclassificadas de acordo com suas faixas de tamanho, em contraposição ao fato de exportarem ou não. Essas informações demonstram, como de conhecimento geral, que independentemente do fato dessas empresas inovarem ou não, bem como da origem de seu capital, existe uma relação positiva entre a orientação exportadora e o tamanho da empresa. No entanto, é importante destacar que no caso das empresas exportadoras estrangeiras inovadoras essa relação é mais tênue do que nas exportadoras nacionais. É possível também perceber a redução do percentual de empresas não exportadoras inovadoras, em contraposição às empresas não inovadoras e não exportadoras. Assim, quando se compara a proporção entre empresas exportadoras e não exportadoras, considerando o fato de serem inovadoras ou não, esta proporção é favorável ao conjunto de empresas inovadoras, sendo que as maiores razões são obtidas no caso das empresas de capital estrangeiro ou misto.¹² A única exceção a essa evidência é verificada na faixa das grandes empresas de capital estrangeiro.

Tabela 3
Percentual de Empresas Inovadoras Segundo a Orientação Exportadora

	NÃO INOVA			INOVA		
	Não-Exporta (A)	Exporta (B)	(B/A)*100	Não-Exporta (C)	Exporta (D)	(D/C)*100
NACIONAL						
MICRO	98.39	1.61	1.64	96.26	3.74	3.89
PEQUENA	93.18	6.82	7.32	86.25	13.75	15.94
MÉDIA	72.08	27.92	38.74	57.04	42.96	75.32
GRANDE	41.57	58.43	140.58	29.10	70.90	243.68
ESTRANGEIRO e MISTO						
MICRO	97.70	2.30	2.35	35.48	64.52	181.82
PEQUENA	48.33	51.67	106.90	30.16	69.84	231.58
MÉDIA	27.10	72.90	268.97	24.73	75.27	304.35
GRANDE	6.06	93.94	1550.0	15.86	84.14	530.43
			0			

¹² Ou seja, a diferença entre a coluna (D) e a coluna (B).

Na Tabela 4 são apresentadas informações relativas ao ano de constituição da empresa, classificadas segundo suas faixas de tamanho para empresas inovadoras ou não. Como era de se esperar, as empresas maiores são as mais antigas, independentemente do fato de terem promovido atividades inovadoras ao longo do período de 1994 a 1996. Em termos gerais, as estatísticas descritivas fornecidas não mostram diferenças sensíveis entre a idade das empresas inovadoras e não inovadoras, a não ser para o caso das microempresas. As mesmas indicam que para a faixa de tamanho das microempresas, as empresas inovadoras têm data de constituição mais recente. É particularmente interessante notar que a maior diferença é aquela correspondente ao valor modal para a faixa das grandes empresas. Nessa faixa de tamanho, ainda que o valor da média indique que as empresas inovadoras são mais antigas que as não inovadoras, a informação modal revela o contrário.

Tabela 4
Estatísticas Descritivas – Ano de Constituição da Empresa

ESTATÍSTICA	FAIXA DE TAMANHO			
	MICRO	PEQUENA	MÉDIA	GRANDE
NÃO INOVA				
Média	1983	1979	1968	1963
Mediana	1986	1982	1973	1961
Moda	1986	1986	1974	1959
Desvio Padrão	10	13	64	18
INOVA				
Média	1985	1978	1969	1960
Mediana	1988	1982	1970	1961
Moda	1988	1984	1974	1973
Desvio Padrão	9	16	16	22

Na Tabela 5 são apresentadas medidas de tendência central para a média salarial do pessoal ocupado diretamente ligado à atividade produtiva, em moeda nacional, para o ano de 1996. De modo similar às tabulações anteriores, essas informações são classificadas segundo as faixas de tamanho das empresas e de acordo com a origem do capital. É notória a evidência de que tanto a média quanto a mediana da média salarial mensal das firmas

nacionais sejam inferiores ao das firmas estrangeiras e mistas, independentemente do fato de serem inovadoras ou não. A única exceção a essa evidência encontra-se no valor da mediana da faixa de microempresas não inovadoras.¹³ Na comparação entre empresas inovadoras e não inovadoras, em termos médios, existe uma diferença em favor da média salarial das empresas inovadoras. Particularmente na faixa de micro e pequenas empresas, predominam as evidências de médias salariais superiores para as empresas inovadoras independentemente da origem de capital. Já para as faixas de médias e grandes empresas, enquanto as empresas nacionais inovadoras apresentam médias salariais superiores ao das empresas nacionais não inovadoras, o inverso se verifica para o valor médio das empresas de capital estrangeiro e misto nas faixas de médias e grandes empresas. No entanto, para o valor da mediana essa diferença volta a ser favorável às empresas inovadoras.

Tabela 5
Estatísticas Descritivas – Média Salarial em Reais para 1996

		NACIONAL	ESTRANGEIRA	MISTO
NÃO INOVA				
MICRO	Média	347,0	377,3	562,3
	Mediana	294,8	244,8	389,5
PEQUENA	Média	426,0	1532,2	826,1
	Mediana	352,6	1180,6	703,9
MÉDIA	Média	573,0	1708,7	1296,6
	Mediana	501,4	911,5	992,4
GRANDE	Média	723,4	1207,5	1072,6
	Mediana	588,6	1025,5	n.d.
INOVA				
MICRO	Média	380,8	975,2	838,9
	Mediana	312,5	1271,1	839,6
PEQUENA	Média	487,9	1473,8	1175,2
	Mediana	417,5	1273,5	1013,6
MÉDIA	Média	666,8	1235,6	1077,3
	Mediana	562,8	1084,6	1009,7
GRANDE	Média	818,1	1169,5	965,1
	Mediana	681,5	1123,2	719,0

13 Ainda considerando como fator distintivo a origem do capital, relativamente as maiores disparidades salariais encontram-se no conjunto de empresas inovadoras nas faixas de micro e pequenas empresas (nacional com relação à estrangeira) e na faixa de pequena empresa não inovadora (nacional com relação à estrangeira).

A Tabela 6, apresentada a seguir, traz informações sobre a razão entre o valor adicionado e a receita líquida da empresa, discriminadas segundo as faixas de tamanho das empresas. Como se pode observar, os valores apresentados na tabela diferem sensivelmente apenas nas faixas de micro e pequenas empresas, sendo estas diferenças a favor das empresas não inovadoras.

Tabela 6
Estatísticas Descritivas – Valor Adicionado/Receita Líquida

	NÃO INOVA		INOVA	
	Média	Mediana	Média	Mediana
MICRO	0,47	0,56	0,40	0,53
PEQUENA	0,48	0,52	0,35	0,51
MÉDIA	0,48	0,51	0,49	0,52
GRANDE	0,46	0,47	0,43	0,46

Em suma, o que as evidências apresentadas na análise descritiva acima sugerem é que existe uma diferenciação mais acentuada entre empresas inovadoras e não inovadoras para o período de 1994 a 1996 quando se tomam como fatores distintivos o tamanho da empresa, o efeito setorial, sua orientação exportadora, a média salarial e a origem do capital. No que tange à origem do capital, essas evidências não são plenamente conclusivas em razão das ambigüidades apresentadas nas relações entre tamanho e origem do capital, e média salarial e origem do capital. Foi possível também observar que o ano de constituição da empresa, bem como a relação entre o valor adicionado e a receita líquida das empresas não são fatores que favorecem a distinção entre empresas inovadoras e não inovadoras. Essas considerações preliminares serão examinadas a seguir por meio de análise estatística.

3 Árvores de classificação para empresas inovadoras/não inovadoras

O objetivo dessa seção é produzir evidências empíricas capazes de auxiliar na classificação das empresas inovadoras, em contraposição às não inovadoras, baseadas no conjunto de variáveis constantes da seção 1. A seguir é apresentada, brevemente, a meto-

dologia de árvores de classificação utilizada nesse trabalho. Posteriormente, são apresentados e discutidos seus resultados.

3.1 Metodologia

Árvores de classificação são métodos relativamente novos empregados na exploração de relações entre dados em problemas de classificação. Uma árvore de classificação é uma regra para predição da classe de um objeto (variável dependente) a partir dos valores de suas variáveis explicativas. São procedimentos estatísticos não-paramétricos, baseados em algoritmos de busca exaustiva. Os resultados apresentados são estruturas hierarquizadas e flexíveis que permitem a observação de distintas relações entre a variável dependente e diversos subconjuntos de variáveis explicativas. As metodologias mais populares são a CHAID (Chi-Square Automatic Iterated Detection) proposta por Kass (1980), a C&RT (Classification and Regression Tree), de Breiman *et alii* (1984), e a QUEST (Quick, Unbiased, Efficient, Statistical Tree) proposta por Loh e Shih (1997).

O diagrama 1 adiante ilustra uma árvore de partição binária com 3 níveis de classificação. Ela é composta por nós. Desde o nó raiz ao nó final, existem diversos nós que formam os ramos de classificação. O nó raiz contém todas as informações da amostra e representa a variável dependente Y , com o número de observações para cada uma das duas categorias (0 e 1 no caso) ou classes. À medida que a árvore se expande, os dados são ramificados em subconjuntos mutuamente exclusivos. Esses subconjuntos são ramos da árvore. No nível seguinte da árvore, a amostra é dividida segundo o melhor preditor de Y , a variável explicativa X_1 . O nó 1 apresenta o subconjunto de informações de Y , classificada segundo uma regra do tipo $X_1 < c$, enquanto que o nó 2 classifica as informações de Y segundo a regra $X_1 \geq c$, em que c é um número ou categoria qualquer correspondente à variável X_1 . Esses nós 1 e 2 contêm informações sobre as frequências da variável Y relacionadas a cada subgrupo, definido segundo as combinações $(Y, X_1, <c)$ e $(Y, X_1 \geq c)$. Os nós 1 e 2 são os pais dos nós 3, 4, 5 e 6. Para o nó 1 o melhor preditor é a variável X_2 , enquanto que para o nó 2 o melhor preditor é a variável X_3 . Esses nós filhos, obtidos a partir da divisão segundo os preditores X_2 e X_3 , contêm informações sobre as frequências da variável Y relacionadas a cada subgrupo definido segundo as combinações (Y, X_1, X_2) e (Y, X_1, X_3) . Estes últimos nós, por se encontrarem ao final da árvore, também são chamados de nós terminais.

Nessa análise será empregado o método de estimação QUEST. Esta metodologia permite a partição binária dos nós, evitando a extrema complexidade da análise, possibilita a inserção de custos diferenciados de classificação para as categorias e a realização da re-

dução da árvore (*prunning*) com base em uma relação benefício/custo. Resumidamente, a metodologia proposta por Loh e Shih (1997) trabalha com a partição dos nós e a seleção das variáveis explicativas separadamente. O algoritmo de partição é dependente do fato da variável preditora ser ordinal, contínua ou nominal.¹⁴ Nos dois primeiros casos, com uma variável X com um número J de classes superior a dois, a partição é realizada inicialmente com reclassificação da variável em duas superclasses, a partir da aplicação do algoritmo de *cluster* proposto por Hartigan e Wong (1979, *apud* Loh e Shih, 1997). Em seguida, é aplicada a análise de discriminante quadrática para se determinar o ponto de partição.¹⁵ Se a variável categórica for nominal, anteriormente a isso é necessária sua transformação em uma variável Z , por meio do mapeamento de X em vetores de variáveis *dummy* 0-1 e a projeção desses vetores sobre sua maior coordenada discriminante.¹⁶ A seleção de variáveis é também realizada dependendo do fato da variável preditora ser categórica nominal ou ordinal, ou contínua. Se a variável X for nominal, então são realizados testes Qui-Quadrado para a hipótese nula de independência entre X e a variável categórica dependente. Se X é contínua ou ordinal, são utilizados testes F de análise de variância.¹⁷

A estimação da árvore envolve o estabelecimento de vários parâmetros. Para a determinação da extensão da árvore são estabelecidas, pelo pesquisador, a extensão máxima de níveis de nós da árvore e as frequências mínimas nos nós pais e filhos. A classificação ainda é realizada em razão dos custos de má classificação e a probabilidade *a priori* para as categorias da variável dependente.¹⁸ Todos esses parâmetros influem nos resultados obtidos pela árvore. Particularmente aqui, é interessante analisar a influência desses dois

14 A título de esclarecimento, nesse trabalho define-se como variável categórica aquela que não apresenta medida contínua. As variáveis categóricas podem ser divididas em nominais ou ordinais. Para as primeiras, as categorias são estabelecidas segundo alguma qualificação da variável, enquanto que para as segundas, as categorias obedecem a algum ordenamento. Adicionalmente, é importante esclarecer que a metodologia QUEST somente é aplicável quando a variável dependente é categórica. No entanto, as variáveis predictoras podem ser contínuas ou categóricas.

15 A análise de discriminante quadrática tem como principal vantagem a possibilidade de lidar com variâncias desiguais para as classes.

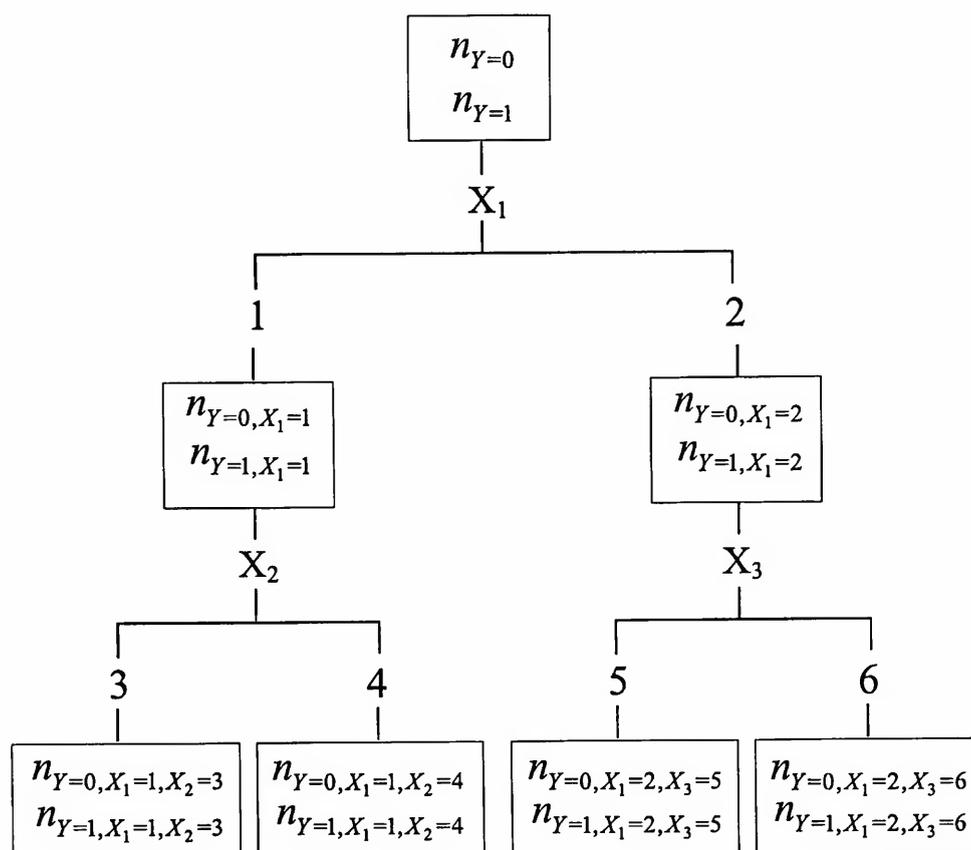
16 A maior coordenada discriminante é a projeção que maximiza a razão entre a soma dos quadrados entre-classes e intra-classes. Para mais informações sobre esse método de transformação linear, ver Gnanadesikan (1977).

17 Se o menor valor p de probabilidade é inferior ao valor p de significância (0,05 no caso) ajustado segundo Bonferroni para comparações múltiplas, a variável preditora é escolhida para a partição do nó. Se nenhum valor p é menor que o valor p limite, valores p de probabilidade são computados para testes estatísticos que são robustos a violações distribucionais, tal como os testes F de Levene para variâncias desiguais. Para obter mais detalhes sobre o algoritmo de seleção quando nenhum preditor tem valor p de probabilidade menor que algum valor limite especificado, ver Loh e Shih (1997).

18 As probabilidades *a priori*, ou "*priors*", especificam o quanto é provável, sem usar qualquer conhecimento *a priori* dos valores dos preditores, que um caso ou objeto pertença a uma das classes. No problema de classificação o pesquisador pode estabelecer que as probabilidades *a priori* sejam iguais às proporções de cada classe da variável dependente, equalizar esses valores para todas as classes, ou ainda estabelecer estes valores segundo uma idéia distinta sobre o fato. Já os custos de má classificação são penalidades numéricas para a classificação de um item em uma categoria ao qual este item não pertence de fato.

últimos parâmetros na possibilidade de se trabalhar com classes desequilibradas e custos distintos de má classificação.

Diagrama 1 – Árvore de Classificação



Este ponto pode ser observado por meio da análise de discriminante quadrática apresentada em Loh e Shih (1997) para a determinação do ponto de partição, com a introdução dos custos de má classificação. Adotando-se uma função de densidade de probabilidade normal univariada, a partição do nó pai para uma variável Y , com duas classes A e B , segundo uma variável explicativa X , é realizada a partir da determinação das raízes da seguinte equação quadrática:¹⁹

$$\left(s_A^2 - s_B^2\right)x^2 + 2\left(\bar{x}_A s_B^2 - \bar{x}_B s_A^2\right)x + \left(\bar{x}_B^2 s_A^2 - \bar{x}_A^2 s_B^2\right) - 2s_A^2 s_B^2 \ln\left(\frac{p_B s_A}{p_A s_B}\right) = 0$$

ou

$$ax^2 + bx + c = 0$$

(1)

19 Para ajudar no raciocínio imagine que A é não-inovador e B inovador, e X seja, por exemplo, o tamanho da empresa. Então o vetor de X seria particionado em X_A e X_B , gerando as médias e variâncias amostrais segundo cada uma das classes.

em que

p_j = probabilidade *a priori* de $j = A, B$

\bar{x}_j = média amostral de X para a classe $j = A, B$

s_j^2 = variância amostral de X para a classe de $j = A, B$

$$a = (s_A^2 - s_B^2)$$

$$b = 2(\bar{x}_A s_B^2 - \bar{x}_B s_A^2)$$

$$c = (\bar{x}_B^2 s_A^2 - \bar{x}_A^2 s_B^2) - 2s_A^2 s_B^2 \ln\left(\frac{p_B s_A}{p_A s_B}\right)$$

A incorporação dos custos de má classificação à análise é realizada por meio da alteração das probabilidades *a priori*, redefinidas da seguinte forma:

$$p'_A = \frac{C(B|A)p_A}{C(B|A)p_A + C(A|B)p_B} \quad \text{e} \quad p'_B = \frac{C(A|B)p_B}{C(B|A)p_A + C(A|B)p_B}$$

sendo que

p'_j é a probabilidade *a priori* ajustada, com $j = A, B$

$c(A | B)$ = custo de classificar erroneamente como A caso X pertença a B

$c(B | A)$ = custo de classificar erroneamente como B caso X pertença a A

Assim, à medida que são estabelecidos custos de má classificação não simétricos, isto implica ajustar as probabilidades *a priori* do modelo. Como é de conhecimento geral, as raízes dessa equação são dadas por:

$$d = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \tag{2}$$

Sendo assim, para qualquer valor de $\left(\frac{p'_B s_A}{p'_A s_B}\right) \neq 1$, isto implica alterar o valor da constante c e, conseqüentemente, o valor de d . Para simplicidade da análise admita o caso de variâncias iguais e médias desiguais. Com isso, a única raiz obtida é dada por:

$$d = \frac{(\bar{x}_A + \bar{x}_B)}{2} - \frac{s_A^2}{(\bar{x}_A - \bar{x}_B)} \ln\left(\frac{p'_A}{p'_B}\right) \quad (3)$$

Desse modo, o deslocamento do ponto médio dependerá do fato de $\bar{x}_A - \bar{x}_B < 0$ ou $\bar{x}_A - \bar{x}_B > 0$ em conjunção com o fato de $\ln\left(\frac{p'_A}{p'_B}\right) > 0$ ou $\ln\left(\frac{p'_A}{p'_B}\right) < 0$. Na presente análise, sendo A a classe de empresas não inovadoras, e B a classe de empresas inovadoras, em que $p_a > p_b$, e suponha que $\bar{x}_A < \bar{x}_B$ com $\bar{x}_A, \bar{x}_B > 0$. Segundo a expressão (3) acima, se os custos de má classificação forem equivalentes, haverá uma tendência de deslocamento do ponto de partição em direção à classe menos numerosa de empresas, minorando a má classificação de empresas não-inovadoras e aumentando a parcela de erro na classificação de empresas inovadoras. A reversão dessa tendência deve ocorrer à medida que $c(A|B) > c(B|A)$, de modo a tornar $\ln\left(\frac{p'_A}{p'_B}\right) < 0$. Esta discussão demonstra a importância do estabelecimento dos custos de má classificação e conseqüente ajuste das probabilidades *a priori* na determinação da classificação realizada pela árvore, na medida em que o ponto de partição é fundamental na verificação da relação entre a variável a ser predita e a variável preditora. Ao se alterar o ponto de partição, isto leva a uma alteração tanto na definição das variáveis preditivas selecionadas quanto nos resultados de classificação da árvore.

A medida de capacidade, ou acurácia, preditiva de uma árvore é construída com base em uma matriz de classificação dada por:

	Valor Observado	
Valor Predito	Classe A	Classe B
Classe A	N_{AC}	N_{AM}
Classe B	N_{BM}	N_{BC}

em que

N_{jC} é número de observações corretamente classificadas com $j = A, B$

N_{jM} é número de observações incorretamente classificadas com $j = A, B$

$N_j = N_{jC} + N_{jM}$ com $j = A, B$

Essa medida é igual a $1 - \text{Risco}$, em que o Risco (ou erro aparente) é dado por:

$$\text{RISCO} = R(T) = \left(\frac{N_{AM} + N_{BM}}{N_A + N_B} \right) * 100 \quad (4)$$

Neste trabalho o risco será calculado utilizando o procedimento de validação cruzada. Esse procedimento determina um número de subamostras aleatórias de igual tamanho, se possível. No presente caso, igual a 10 amostras. Com isso, a árvore de classificação é computada 10 vezes, sendo que em cada uma dessas rodadas, uma dessas subamostras é deixada de fora da amostra e utilizada como amostra de testes para validação da árvore. O risco estimado apresentado em (4) será igual à média dos riscos obtidos em cada uma dessas estimações.

A complexidade de análise de árvore de classificação pode ser um fator de custo para o pesquisador. Assim, à medida que a árvore se expande, por um lado produz o benefício de redução do risco, mas, por outro lado, traz um custo com o aumento de sua complexidade. Um critério de redução da árvore é aquele que considera essa relação benefício/custo. Sendo assim, para a redução da árvore é utilizado um índice custo-complexidade que mede o risco e complexidade da árvore, dado por:

$$R_\alpha(T) = R(T) + \alpha |\tilde{T}| \quad (5)$$

em que α é o custo de complexidade e $|\tilde{T}|$ é o número de nós terminais para uma árvore T .²⁰ O processo de redução é realizado por meio da estimação de seqüência de subárvores em que são removidos sub-ramos que promovem a menor redução do risco relativamente ao cálculo recursivo de uma medida α de complexidade. A escolha da árvore reduzida é tal que o risco da menor das subárvores seja menor ou igual ao de risco da árvore de mínimo risco mais um erro padrão do risco.

3.2.2 Resultados

A amostra para a estimação das árvores é constituída de 38.574 empresas.²¹ As estimativas de árvores foram realizadas estabelecendo como 5 a extensão máxima do nível de nós da árvore, 50 a freqüência mínima nos nós pais e 25 a freqüência mínima nos nós filhos. Inicialmente são utilizados como probabilidades *a priori* para as classes inovadoras e não inovadoras as respectivas parcelas dessas classes na amostra utilizada e estabelecidos custos de má classificação iguais a 1 para as classes não inovadoras e inovadoras. O nível de significância para os testes sobre as variáveis preditivas é de 5%. As variáveis selecionadas para análise e as transformações utilizadas estão apresentadas no Quadro 1. A árvore 1, apresentada a seguir, corresponde àquela estimada com processo de redução baseado no critério de risco mínimo. Ou seja, considerando os parâmetros acima descritos e o processo de validação cruzada, é selecionada a árvore que tem o risco médio mínimo. A seguir é apresentada na árvore 2 a estimação da árvore reduzida baseada no critério de custo-complexidade.

20 Uma árvore poderia crescer até o momento em que cada nó terminal contivesse somente uma informação. Essa árvore teria um risco mínimo. Isto ocorreria se α fosse igual a zero e T fosse suficientemente grande, combinado à ausência de restrições para as freqüências nos nós pais e filhos. Assim, à medida que α aumenta, a partir de zero, produz-se uma seqüência de subárvores T_1, T_2, T_3 com, progressivamente, um número menor de nós.

21 Essa amostra foi reduzida por causa de variáveis sem informação e devido à exclusão das empresas com valores para a variável de valor adicionado negativos.

Quadro 1
Variáveis Empregadas na Estimação das Árvores de Classificação

Variável	Variável Empírica	Transformações
Não Inova/Inova (dependente)	Declaração das empresas para o período 1994/96	Categórica 0 = Não Inova 1 = Inova
Tamanho da Firma	Pessoal Ocupado Total na Empresa	Logarítmica
Estrutura de Mercado	Medida CR4 para indústria paulista	Nenhuma
Orientação Exportadora	Empresas que exportam 1% ou mais do seu total de vendas	Categórica 0 = Não Exporta 1 = Exporta
Aprendizado Acumulado	Década de Constituição da Empresa	Categórica 1 = até 1949 2 = 1950 a 1959 3 = 1960 a 1969 4 = 1970 a 1979 5 = 1980 a 1989 6 = 1990 e depois
Qualificação do Trabalho	Média Salarial Mensal	Logarítmica
Origem do Capital Controlador	Nacional, Estrangeiro ou Misto em 1996	Categórica 1 = Nacional 2 = Estrangeiro 3 = Misto
Apropriabilidade	Relação entre Valor Adicionado e Receita Líquida da Empresa	Logarítmica
Efeitos Setoriais	Divisão CNAE ^a Classificações Segundo as Oportunidades Tecnológicas (INTES3) e Segundo a Intensidade do Fator de Produção (2)	Categórica (INTENS3) 1 = Intens. em Recursos Naturais 2 = Intens. em Trabalho 3 = Intens. em Capital e Tecnologia Categórica (2) 1 = Alta Oportunidade 2 = Média Oportunidade 3 = Média-Baixa Oportunidade 4 = Baixa Oportunidade

a. As classificações aqui utilizadas foram propostas por Moreira e Najberg (1998).

As Tabelas 7 e 8 apresentam os sumários estatísticos das árvores 1 e 2. Essas árvores possuem custos simétricos de classificação, sendo a primeira de risco mínimo e a segunda reduzida de acordo com o critério de custo-complexidade. Como é possível perceber, o risco estimado para essas árvores é igual a 0,23^{22,23}. Os nós terminais, em ordem decrescente, com maior percentual de empresas inovadoras são indicados pela coluna lucro nessas tabelas.²⁴ Para a árvore 1, os nós considerados representativos de empresas inovadoras, em ordem decrescente, são os de números 24, 15, 10, 13, 20, 17, 25 e 22. Já para árvore 2, os nós representativos das empresas inovadoras são os de números 8, 4, 12 e 9. Em termos gerais, as árvores 1 e 2 destacam, também em ordem decrescente, a capacidade distintiva do fato da empresa ser exportadora, a relação positiva com o tamanho, os efeitos setoriais, a concentração setorial e a origem estrangeira ou mista do capital na classificação das empresas inovadoras.

Como se pode perceber, os dois primeiros nós em poder de classificação na árvore são filhos do ramo das empresas exportadoras com até 55 empregados e pertencentes a setores com grau de concentração superior a 0,50. A diferença entre o nó 24 e o 15 é o fato das empresas classificadas no primeiro nó pertencerem a setores intensivos em capital e tecnologia e serem constituídas a partir da década de 1970, enquanto que no nó 15 as empresas selecionadas pertencem a setores intensivos em trabalho ou recursos naturais. Ainda no mesmo ramo de empresas exportadoras (com 55 ou menos empregados), é importante destacar a classificação promovida pelo nó 13 de empresas atuantes em setores com grau de concentração inferior a 0,50 e de alta oportunidade tecnológica.

Também no ramo das empresas exportadoras, o nó 10 classifica como empresas tipicamente inovadoras aquelas com mais de 224 empregados. Partindo ainda do mesmo ramo da árvore, o nó 17 apresenta uma outra classificação de empresas inovadoras como sendo aquelas pertencentes a setores com média-baixa, média e alta oportunidade tecnológica na faixa de 55 a 224 empregados.

22 Essa estimativa de erro é inferior à estimativa ingênua de 0,5, e também inferior à estimativa de que todas as empresas são não inovadoras.

23 A robustez desses resultados pode ser verificada por meio da comparação com o resultado do modelo logit apresentado no apêndice.

24 É importante esclarecer que a árvore teria como objetivo a classificação estrita em cada nó. Ou seja, todas as informações, em cada um dos nós, deveriam ser de apenas uma das classes (Não Inovadora ou Inovadora). Sendo assim, se, por exemplo, um nó da árvore contém um percentual da classe Inovadora superior a 50%, esse nó está dizendo que mais provavelmente sob aquela partição esse nó representaria a classe de empresas inovadoras segundo classificação proposta pelo ramo ao qual este nó pertence. Dessa forma, a medida de risco do nó é a quantidade complementar ao percentual de empresas Inovadoras no nó. É nesse sentido que a tabela de lucros indica os nós em que o grau de homogeneidade é maior, significando que nesses nós a classificação é mais clara e, portanto, menos sujeita a riscos.

Tabela 7
Sumário Estatístico da Árvore de Classificação 1

Nós	Frequência	Porcentual	Lucro	Matriz de Classificação		
					Valor Observado	
24	90	0.2	0.79			
15	90	0.2	0.72	Valor Predito	Não Inova	Inova
10	998	2.6	0.68	Não Inova	27771	7902
13	37	0.1	0.68	Inova	950	1783
20	283	0.7	0.66			
17	926	2.4	0.62	Risco	0.230485	
25	46	0.1	0.59			
22	263	0.7	0.59			
19	370	1.0	0.48			
18	287	0.7	0.48			
26	1306	3.4	0.41			
14	804	2.1	0.34			
23	48	0.1	0.33			
5	32858	85.6	0.21			

Tabela 8
Sumário Estatístico da Árvore de Classificação 2

Nós	Frequência	Porcentual	Lucro	Matriz de Classificação		
					Valor Observado	
8	228	0.6	0.67			
4	2211	5.8	0.63	Valor Predito	Não Inova	Inova
12	263	0.7	0.59	Não Inova	27429	7622
9	653	1.7	0.56	Inova	1292	2063
11	1352	3.5	0.42			
7	841	2.2	0.35	Risco	0.232099	
5	32858	85.6	0.21			

Já no ramo de empresas não exportadoras, a árvore 1 demonstra que o tamanho nessa categoria de empresas é um bom preditor para a classe de empresas inovadoras, já que é a primeira característica selecionada para a distinção de empresas inovadoras e não inovadoras. No entanto, diferentemente do caso das empresas exportadoras, o nível de classificação sobe para 90 empregados. Isto é, no caso das não exportadoras o tamanho é uma categoria mais relevante na explicação da probabilidade da firma inovar ou não. Nesse ramo, três nós terminais são representativos das empresas inovadoras: o nó 20, o 25 e o 22. O nó 20 seleciona as empresas com mais de 90 empregados, atuantes em setores de média e alta oportunidade tecnológica, com graus de concentração superior a 0,37. O nó 22 seleciona empresas com mais de 365 empregados, enquanto que o nó 25 empresas com mais de 90 e menos de 365 empregados de origem estrangeira ou mista do capital.

A árvore 2 demonstra claramente que, no ramo das empresas exportadoras, as características com maior poder de discriminação são o tamanho e o grau de concentração, enquanto que no ramo das empresas não exportadoras o tamanho é a principal variável distintiva.

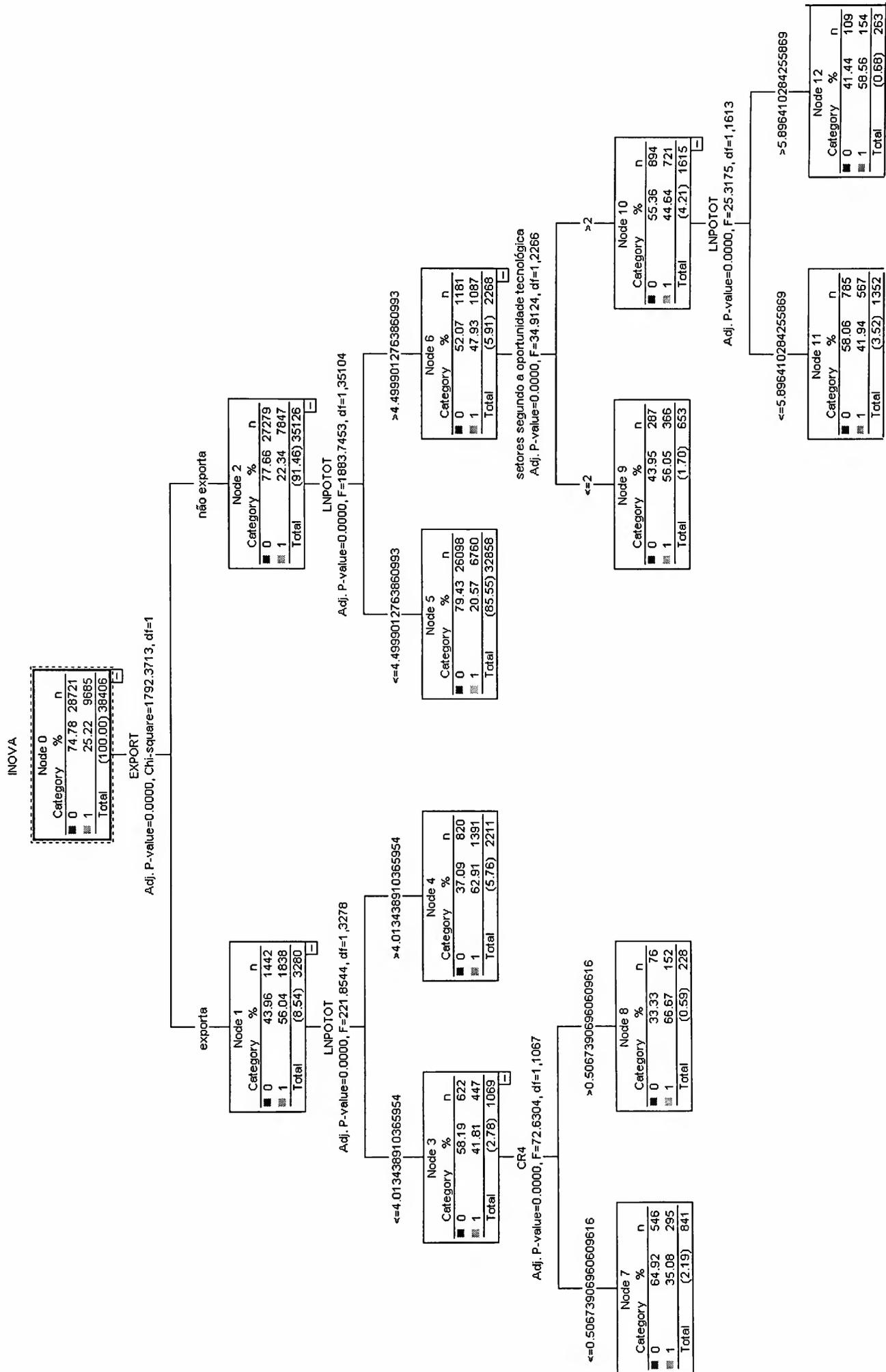
Na árvore 3 foi realizado um esforço de classificação para as empresas inovadoras pertencentes ao nó 5 das árvores 1 e 2, classificado tipicamente como um nó não inovador nessas árvores. Ainda assim, este nó contém mais de 70% das empresas inovadoras da amostra. Para isso, a árvore 3 foi estimada estabelecendo-se um custo igual a 3 para a classificação errônea de uma empresa não inovadora como inovadora e custo igual a 1 para classificação errônea de uma empresa inovadora como não inovadora. Ou seja, a árvore 3 deve classificar mais precisamente a classe de empresas não inovadoras. A estimativa apresentada refere-se àquela reduzida segundo o critério de custo-complexidade. O sumário estatístico dessa árvore está apresentado na Tabela 9.

Tabela 9
Sumário Estatístico da Árvore de Classificação 3

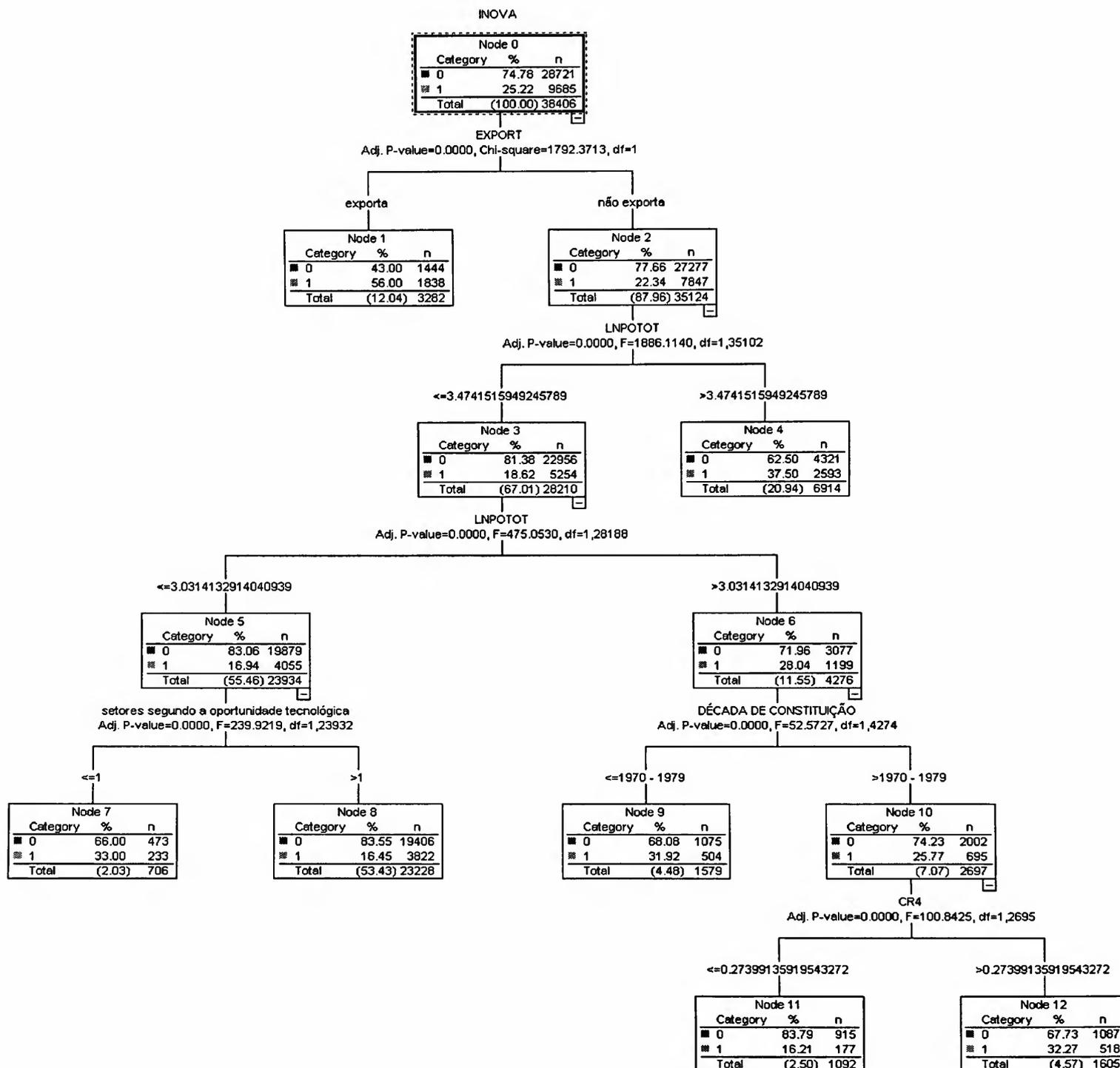
Nós	Frequência	Porcentual	Lucro	Matriz de Classificação		
1	3282	12.0	0.79	Valor Observado		
4	6914	20.9	0.64	Valor Predito	Não Inova	Inova
7	706	2.0	0.60	Não Inova	20321	3999
12	1605	4.6	0.59	Inova	8400	5686
9	1579	4.5	0.58			
8	23228	53.4	0.37	Risco	0.353036	
11	1092	2.5	0.37			

Na tabela supracitada, o porcentual de empresas inovadoras corretamente classificadas chega a aproximadamente 59%, demonstrando um grande avanço em termos da classificação dessa categoria de empresas com relação às classificações realizadas pelas árvores 1 e 2. A árvore 3 reforça a existência de uma relação positiva entre tamanho e a probabilidade de inovar das empresas não exportadoras, mesmo para as empresas médias e pequenas. Essa relação positiva é verificada em todas as partições da árvore, mas principalmente no nó 4, em que a única característica distintiva é o fato de a empresa ter mais de 32 empregados. No entanto, para as demais classificações, outras características também são importantes. Um primeiro subconjunto apresentado no nó 7 são as empresas com 20 ou menos empregados, atuantes em setores de alta oportunidade tecnológica, provavelmente representando as empresas de base tecnológica desses setores. Já o nó 12 destaca a caracterização de empresas com mais de 20 empregados e menos de 32, com constituição após a década de 1970 e pertencentes a setores com grau de concentração superior a 0,27, enquanto que o nó 9 apresenta as empresas originadas antes da década de 1970.

Árvore 2 Não Inovadoras/Inovadoras



Árvore 3 Não Inovadoras/Inovadoras



ficações aqui estabelecidas são úteis para se definir as características relevantes para a distinção das empresas inovadoras em contraposição àquelas não inovadoras.

Trata-se de um primeiro trabalho nessa linha sobre o tema, abrindo perspectivas para trabalhos posteriores em que poderão ser analisadas categorias mais restritas, ou particulares, de atividade inovadora. Para sua realização foram utilizadas informações constantes na PAEP para o período de 1994 a 1996 e realizadas estimações de árvores de classificação. A hierarquização e o formato apresentado pelas árvores referendaram a ampla maioria das proposições teóricas acima apresentadas na definição das variáveis da pesquisa.

Os resultados, em termos gerais, demonstram que a probabilidade de uma empresa ser inovadora está diretamente associada à orientação exportadora da empresa. A distinção causal entre inovação e inserção internacional não foi possível de ser realizada neste estudo, ficando essa questão em aberto e a ser examinada em estudos posteriores. Adicionalmente, em ordem decrescente, são fatores que alteram positivamente essa probabilidade o seu tamanho, a participação dessas empresas em setores intensivos em tecnologia, ou com maiores oportunidades tecnológicas, o grau de concentração setorial, a idade da empresa e a origem não nacional do capital.

Um resultado interessante apresentado pelas árvores de classificação é o relativo à variável CR4, representativa da estrutura de mercado. Diferentemente da explicação usualmente oferecida pela teoria, em que possivelmente esta variável ajudaria na definição da probabilidade de grandes empresas atuantes como líderes em setores industriais concentrados, esta variável esteve associada às micro e pequenas empresas industriais. Em alguns casos também esteve associada a empresas atuantes em setores dinâmicos tecnologicamente, denotando as pressões competitivas oferecidas por estes setores mais concentrados sobre suas empresas.

Em razão dos resultados obtidos com a utilização da variável “Década de Constituição da Empresa”, é possível estabelecer-se uma relação positiva entre a atividade inovadora e o conhecimento tácito. Nas duas circunstâncias em que essa variável foi útil para a classificação de empresas inovadoras, ela esteve associada às pequenas empresas. Considerando a alta taxa de mortalidade das micro e pequenas empresas, pode-se concluir, pelas datas de constituição selecionadas, que indicam que são empresas com, no mínimo, mais de quinze anos, que o conhecimento tácito dessas empresas é um fator que contribui para definir a probabilidade dessas empresas serem inovadoras.

Na estimativa das árvores de classificação, a variável representativa para o grau de apropriabilidade apresentou baixo poder distintivo entre as empresas inovadoras e não inovadoras. O mesmo ocorreu para a variável representativa do nível de qualificação do trabalho. No entanto, pode ser observado, na estimativa do modelo logit apresentada em apêndice, que o nível de qualificação tem um efeito positivo e significativo estatisticamente sobre a probabilidade de inovar, enquanto que o grau de apropriabilidade tem um efeito negativo.

Bibliografia

- Braga, H.; Willmore, L. Technological imports and technological effort. *Journal of Industrial Economics*, Blackwell Publishers, v. 39, n. 4, p. 421-432, Jun./1991.
- Breiman, L.; Friedman, J.; Olshen, R.; Stone, C. *Classification and regression trees*. Belmont, California: Wadsworth, 1984.
- Chandler, Alfred D. Jr. What is a firm? A historical perspective. *European Economic Review*, v. 36, n. 2, p. 483-492, 1992.
- Cohen, W. M. Empirical studies of innovative activity. In: Stoneman, P. (ed.), *Handbook of the economics of innovation and technological change*. Oxford: Blackwell, 1995.
- Cohen, W. M.; Levin, R. C. Empirical studies of innovative activity and market structure. In: Schmalensee, R.; Willig, R. (eds.), *Handbook of industrial organization*. Amsterdam: North Holland, 1989.
- Dosi, G. Sources, procedures, and microeconomic effects of innovation. *Journal of Economic Literature*. American Economic Association, v. 26, n. 3, p. 1120-1171, Sep./1988.
- Gnanadesikan, R. *Methods for statistical data analysis of multivariate observations*. New York: John Wiley & Sons, Inc, 1977.
- Henderson, R. Underinvestment and incompetence as responses to radical innovation: evidence from the photolithographic alignment equipment industry. *The RAND Journal of Economics*. RAND Corporation, v. 24, n. 2, p. 248-270, Summer/1993.
- Kass, G. V. An exploratory technique for investigating large quantities of categorical data. *Applied Statistics*. Royal Statistical Society, v. 29, n. 2, p. 119-127, 1980.
- Kumar, N.; Siddharthan, N. S. *Technology, market structure and internalization: issues and policies for developing countries*. London: Routledge, 1997.

Loh, W. Y.; Shih, Y. S. Split selection methods for classification trees. *Statistica Sinica*, 7, p. 815-840, 1997.

Moreira, M. M.; Najberg, S. Abertura comercial: criando ou destruindo empregos? *Pesquisa e Planejamento Econômico*. IPEA, v. 28, n. 2, Ago./1998.

Quadros, R.; Furtado, A.; Bernardes, R; Franco, E. Technological innovation in Brazilian industry: an assessment based on the São Paulo innovation survey. *Technological forecasting and social change*, v. 67, n. 2-3, p. 203-219, Jun./2001.

Pavitt, K. Sectoral patterns of technical change: towards a taxonomy and a theory. *Research Policy*, 13, p. 343-373, 1984.

Apêndice

Modelo de Regressão Logística: Probabilidade da Empresa ser Inovadora – Indústria Paulista

Variável dependente = probabilidade de a empresa ser inovadora					
Variáveis Independentes...	Coefficientes	Wald	Significância	Rz de chance	
LNPOTOT	0.51127	1266.76119	0.00000	1.66741	
EXPORT	0.98840	50.64082	0.00000	2.68693	
EXPORT*LNPOTOT	-0.12245	15.73555	0.00007	0.88475	
ORGCAP	0.15810	7.10808	0.00767	1.17128	
LGMESAL	0.26407	34.44356	0.00000	1.30221	
LNVAREL	-0.10310	28.52549	0.00000	0.90204	
CR4	0.19639	5.23443	0.02214	1.21700	
Década S/ Inform.	0.31504	6.07538	0.01371	1.37031	
ATE 1949	-0.17451	4.53031	0.03330	0.83987	
1950 – 1959	-0.09975	2.51156	0.11301	0.90506	
1960 – 1969	-0.02722	0.32821	0.56672	0.97314	
1970 – 1979	-0.19416	24.14804	0.00000	0.82353	
1980 – 1989	0.03826	1.31155	0.25211	1.03900	
Dummies	D10	-4.83958	0.41304	0.52043	0.00791
Setoriais	D11	2.64021	0.03825	0.84495	14.01613
	D13	0.37990	0.23146	0.63044	1.46214
	D14	-1.21833	41.61466	0.00000	0.29572
	D15	-0.23692	10.28542	0.00134	0.78906
	D16	-6.52141	0.68415	0.40816	0.00147
	D18	-0.36919	22.40124	0.00000	0.69130
	D19	0.02648	0.08882	0.76568	1.02683
(Referência =	D20	-1.08847	58.03043	0.00000	0.33673
Têxtil= 17)	D21	-0.09381	0.86660	0.35190	0.91045
	D22	0.26524	12.08762	0.00051	1.30374
	D23	-0.61653	6.10230	0.01350	0.53981
	D24	0.59841	57.23194	0.00000	1.81923

Variável dependente = probabilidade de a empresa ser inovadora				
Variáveis Independentes...	Coeficientes	Wald	Significância	Rz de chance
D25	0.52086	49.14005	0.00000	1.68347
D26	0.12271	2.51564	0.11272	1.13056
D27	0.23506	6.90024	0.00862	1.26498
D28	0.34648	24.20216	0.00000	1.41407
D29	0.42668	33.79831	0.00000	1.53215
D30	1.73186	59.91801	0.00000	5.65114
D31	0.21155	5.27726	0.02161	1.23559
D32	0.80170	42.68619	0.00000	2.22934
D33	0.99808	84.73289	0.00000	2.71306
D34	0.28820	9.81648	0.00173	1.33402
D35	0.33122	3.64129	0.05636	1.39266
D36	0.47157	41.64751	0.00000	1.60250
Constante	-3.67986	791.68145	0.00000	0.02523
Ajuste do Modelo				
	Observado	Predito (o valor de corte é 0,5)		
-2 Log likelihood = 38380.69		Não inova	Inova	% de acerto
Cox & Snell R Square = 0.11279	Não Inova	27258.28	1279.794	95.52
Nagelkerke R Square = 0.16685	Inova	7549.928	2015.496	21.07
				76.83

Classificação CNAE – Indústria de Transformação

- 10 EXTRAÇÃO DE CARVÃO MINERAL
 - 11 EXTRAÇÃO DE PETRÓLEO E SERVIÇOS RELACIONADOS
 - 13 EXTRAÇÃO DE MINERAIS METÁLICOS
 - 14 EXTRAÇÃO DE MINERAIS NÃO-METÁLICOS
 - 15 FABRICAÇÃO DE PRODUTOS ALIMENTÍCIOS E BEBIDAS
 - 16 FABRICAÇÃO DE PRODUTOS DO FUMO
 - 17 FABRICAÇÃO DE PRODUTOS TÊXTEIS
 - 18 CONFECÇÃO DE ARTIGOS DO VESTUÁRIO E ACESSÓRIOS
 - 19 PREPARAÇÃO DE COUROS E FABRICAÇÃO DE ARTEFATOS DE COURO, ARTIGOS DE VIAGEM E CALÇADOS
 - 20 FABRICAÇÃO DE PRODUTOS DE MADEIRA
 - 21 FABRICAÇÃO DE CELULOSE, PAPEL E PRODUTOS DE PAPEL
 - 22 EDIÇÃO, IMPRESSÃO E REPRODUÇÃO DE GRAVAÇÕES
 - 23 FABRICAÇÃO DE COQUE, REFINO DE PETRÓLEO, ELABORAÇÃO DE COMBUSTÍVEIS NUCLEARES E PRODUÇÃO DE ÁLCOOL
 - 24 FABRICAÇÃO DE PRODUTOS QUÍMICOS
 - 25 FABRICAÇÃO DE ARTIGOS DE BORRACHA E PLÁSTICO
 - 26 FABRICAÇÃO DE PRODUTOS DE MINERAIS NÃO-METÁLICOS
 - 27 METALURGIA BÁSICA
 - 28 FABRICAÇÃO DE PRODUTOS DE METAL-EXCLUSIVE MÁQUINAS E EQUIPAMENTOS
 - 29 FABRICAÇÃO DE MÁQUINAS E EQUIPAMENTOS
 - 30 FABRICAÇÃO DE MÁQUINAS PARA ESCRITÓRIO E EQUIPAMENTOS DE INFORMÁTICA
 - 31 FABRICAÇÃO DE MÁQUINAS, APARELHOS E MATERIAIS ELÉTRICOS
 - 32 FABRICAÇÃO DE MATERIAL ELETRÔNICO E DE APARELHOS E EQUIPAMENTOS DE COMUNICAÇÕES
 - 33 FABRICAÇÃO DE EQUIPAMENTOS DE INSTRUMENTAÇÃO MÉDICO-HOSPITALARES, INSTRUMENTOS DE PRECISÃO E ÓPTICOS, EQUIPAMENTOS PARA AUTOMAÇÃO INDUSTRIAL, CRONÔMETROS E RELÓGIOS
 - 34 FABRICAÇÃO E MONTAGEM DE VEÍCULOS AUTOMOTORES, REBOQUES E CARROCERIAS
 - 35 FABRICAÇÃO DE OUTROS EQUIPAMENTOS DE TRANSPORTE
 - 36 FABRICAÇÃO DE MÓVEIS E INDÚSTRIAS DIVERSAS
 - 37 RECICLAGEM
-