

Dinâmica Multissetorial da Indústria Brasileira: um Estudo Empírico

Luis Otávio Façanha
Marcelo Resende

Instituto de Economia, Universidade Federal do Rio de Janeiro
Instituto de Economia, Universidade Federal do Rio de Janeiro

RESUMO

O trabalho investigou aspectos da dinâmica da indústria de transformação no período 1995-1997 em termos de variáveis relativas a tamanho médio de planta, taxas de entrada e de saída e turbulência. Os resultados no nível de 4 dígitos evidenciaram um elevado grau de heterogeneidade de comportamento entre os diferentes setores. A tentativa de identificação de padrões baseou-se na Análise Discriminante com pré-classificações definidas a partir de Análise de Grupamento em setores a 2 dígitos e categorias de uso dos bens. Os resultados não foram suficientemente fortes no sentido de evidenciarem padrões setoriais definidos e, portanto, são consistentes com outros estudos desenvolvidos para os Estados Unidos.

PALAVRAS-CHAVE

dinâmica industrial, análise discriminante

ABSTRACT

The article investigated aspects of the manufacturing industrial dynamics in the period 1995-97 in terms of variables related to average plant size, entry and exit rates and turbulence. The results at the 4 digits level indicated a high degree of behavior heterogeneity across the different sectors. The attempt to identify patterns was based in Discriminant Analysis with pre-classifications defined upon Cluster Analysis in 2 digits sectors and users group. The results were not sufficiently strong in the sense of indicating salient sectoral patterns and therefore are consistent with other studies developed for the United States.

KEY WORDS

industrial dynamics, discriminant analysis

JEL Classification
L10 e L60

INTRODUÇÃO

A mobilidade perfeita no âmbito de um mercado, em termos da livre entrada e saída dos agentes econômicos, constitui uma condição importante para um equilíbrio competitivo de longo prazo sem a persistência de lucros anormais. A existência de imperfeições de mercado, expressas sobretudo em termos da presença de diferentes categorias de barreira à entrada, foi amplamente reconhecida pelo menos desde Bain (1956). A literatura empírica correspondente tem crescido de forma significativa, em que pesem as dificuldades de obtenção de dados mais desagregados (ver GEROSKI, 1991, 1995 e SIEGFRIED & EVANS, 1994, para resenhas bibliográficas). A referida literatura, contudo, centraliza sua atenção na superação dos problemas empíricos mencionados e no estudo de movimentos de entrada e saída de firmas em indústrias de países desenvolvidos, valendo mencionar, a respeito, o trabalho referencial de Dunne, Roberts e Samuelson (1988), baseado em dados censitários para os EUA, no qual os autores concluem pela existência de padrões de entrada e saída na indústria de transformação que são, em grande medida, setores-específicos. Assim sendo, parece existir uma importante lacuna no que tange ao estudo de estudos multissetoriais em economias heterogêneas e complexas como a do Brasil, o que se justifica, em grande medida, por dificuldades empíricas e de disponibilidade de dados a serem superadas.

Na verdade, estudos multissetoriais para o caso brasileiro são raros, e deixaram de configurar linhas sistemáticas de pesquisa, devendo-se mencionar os trabalhos de Tavares, Possas e Façanha (1978) e de Rodrigues e Façanha (1992). O primeiro trabalho mencionado examinou a estrutura industrial do Brasil com tabulações especiais relativas ao Censo de 1970 e à Pesquisa Industrial Anual de 1973, com dados de atividades para os setores industriais a 4 dígitos, e identificação das firmas proprietárias dos 4 maiores estabelecimentos de cada setor, o que permitiu identificar a origem da propriedade das “firmas líderes”, grau de internacionalização dos setores, indicadores de concentração, e de margens e taxas de lucros, dentre outros, para o conjunto dos setores e para as firmas líderes. O acesso aos dados de produção física viabilizou o cálculo de índices de preços e de quantidade para cada um dos 399 setores estudados, e estudo exploratório do período 1970-1973. O tra-

balho de Rodrigues e Façanha consistiu na aplicação de técnicas de estatística multivariada (principalmente análise fatorial e análise discriminante) aos dados de atividades relativos aos anos censitários de 1970, 1975 e 1980 e a faixas de tamanho de estabelecimentos, alcançando distinção de fases distintas de crescimento da indústria e categorização de grupos de setores representados por indicadores de lucratividade, de estratégias de preços e de aumento de produção, de concentração e relativos ao dinamismo de menores estabelecimentos.

O presente artigo pretende retomar o esforço de análise do comportamento da indústria brasileira, elaborando tratamento conjunto de indicadores ainda inexplorados em trabalhos anteriores, tais como taxas de entrada, taxas de saída de estabelecimentos, turbulência (conforme definida a seguir) e variáveis de estrutura que aproximam a escala eficiente mínima. Assim sendo, o estudo privilegiará aspectos relativos à mobilidade de firmas nos diferentes setores e menos aspectos mais convencionais de estrutura e desempenho. Para tanto, serão considerados microdados escassamente utilizados no Brasil, que abarcam o universo de estabelecimentos formais existentes no País para o período 1995-97. O trabalho está organizado da seguinte forma. Na próxima seção serão apresentadas as variáveis utilizadas e os valores assumidos pelas mesmas em 243 setores da indústria de transformação para os anos de 1996 e 1997, assim como estatísticas sintéticas relativas às variáveis utilizadas. Na seção 2 serão explorados padrões setoriais para as referidas variáveis e, para tanto, será utilizada a abordagem de Análise Discriminante. A última seção apresenta os comentários finais.

1. CONSTRUÇÃO DE INDICADORES SETORIAIS

1.1 Base de Dados

O presente estudo faz uso de microdados da Relação Anual de Informações Sociais-RAIS de difícil acesso e pouca utilização. Essa fonte abarca informações sobre o número de empregados e abrange todo o conjunto de estabelecimentos formais no Brasil. O uso de registros administrativos para estudar a estrutura industrial já fora sugerido e implementado com base de dados

análoga para o caso português, conforme exemplificado nos trabalhos de Mata e Machado (1996) e de Görg *et al.* (2000) para o caso da Irlanda. No caso brasileiro, Najberg *et al.* (2000) utilizaram microdados da RAIS no período 1995-97 para investigar o padrão de sobrevivência das pequenas e médias firmas no Brasil.¹ No presente trabalho, considera-se a mesma base de dados naquele período (que inclui mais de 200.000 estabelecimentos a cada ano) e calculam-se indicadores de estrutura industrial, como o tamanho médio das plantas, e outros, referentes a questões de dinâmica industrial, tais como as taxas de entrada e de saída de firmas e o grau de turbulência, relativo a emprego e a firmas. As definições precisas das variáveis são as seguintes:²

- **tamanho médio das plantas-1996 (TMP96)**: total de pessoal ocupado do setor dividido pelo número de estabelecimentos no mesmo setor (para o ano de 1996)³;
- **tamanho médio das plantas-1997 (TMP97)**: idem para 1997;
- **taxa de entrada-1996 (TENT96)**: número de estabelecimentos novos em 1996 relativamente ao total de estabelecimentos em 1995. Considera-se como um novo estabelecimento aquele que não tenha respondido a RAIS em 1996, podendo verificar-se essa condição mediante a comparação de identificadores de empresas-estabelecimentos (cadastro nacional de pessoas jurídicas-CNPJ);
- **taxa de entrada-1997 (TENT97)**: idem para os anos de 1997 e 1996, respectivamente;

1 Os autores citados referem-se a estabelecimentos como firmas. De fato, os autores consideram o padrão de sobrevivência de novos estabelecimentos nos anos subsequentes para diferentes faixas de tamanho e, portanto, tratam de questões distintas dos padrões de saída evidenciados no presente estudo, já que não se foca aqui exclusivamente em unidades que sejam novas no período inicial de análise.

2 Não se pode garantir que o estudo esteja imune a omissões no preenchimento da pesquisa da RAIS que estão sendo interpretadas como saída de estabelecimentos. Se tal tipo de ocorrência fosse numeroso, esperar-se-ia uma abrupta mudança dos padrões das variáveis entre 1996 e 1997. Outrossim, a disponibilidade de microdados da RAIS com identificação tem difícil viabilidade para um intervalo de tempo maior. Vale ainda ressaltar que MATA (1995) resenhou diferentes trabalhos com base de dados análoga para Portugal, e em pelo menos um daqueles o critério de definição de firma nova é semelhante ao aqui adotado.

3 Os dados da RAIS consideram o pessoal ocupado em 31/12 de cada ano.

- **taxa de saída-1996 (TSAI96)**: número de estabelecimentos registrados pela RAIS em 1995 e que não constam da RAIS em 1996, relativamente ao total de estabelecimentos em 1995;
- **taxa de saída-1997 (TSAI97)**: idem levando-se em consideração a comparação dos anos de 1997 e 1996;
- **turbulência do emprego-1996 (TURBE96)**: construída pelo produto das taxas de criação e destruição de empregos em cada setor levando em conta os anos de 1996 e 1995.⁴ O cálculo das referidas taxas de criação e destruição de emprego são viabilizados em face da disponibilidade de microdados com identificação no nível de estabelecimento. Assim, pode-se identificar o volume de emprego criado ou destruído em cada estabelecimento, e essa informação pode ser agregada para cada setor a 4 dígitos, de modo que mediante uma comparação com o estoque de emprego do período anterior pode-se chegar às taxas que compõem a variável TURBE;
- **turbulência do emprego-1997 (TURBE97)**: idem para os anos de 1997 e 1996;
- **turbulência de firmas-1996 (TURBF96)**: é a soma das taxas de entrada e saída de firmas (estabelecimentos) em 1996 (TENT96+TSAI96);
- **turbulência de firmas-1997 (TURBF97)**: idem para 1997.⁵

1.2 Resultados Empíricos Preliminares

A Tabela 1A apresentada no anexo exhibe os valores das variáveis descritas anteriormente para setores a 4 dígitos da indústria de transformação brasileira (CNAE4). Alguns setores que apresentaram erros evidentes por conta de valores não plausíveis para as variáveis foram excluídos da análise; assim

4 Essa medida foi utilizada nos trabalhos de MATA & MACHADO (1996) e GÖRG *et al.* (2000). Como a variável capta a intensidade da criação e destruição simultânea de emprego em um dado setor, o último trabalho propõe que se interprete a mesma como uma *proxy* da prevalência de custos irrecuperáveis (*sunk costs*).

5 Esse conceito foi sugerido por BEESLEY & HAMILTON (1984).

sendo, o estudo se concentrará em 243 setores. Esses representam aproximadamente 91% relativamente aos 268 setores existentes a 4 dígitos, cabendo destacar a ampla cobertura em termos de estabelecimentos industriais ante o caráter censitário da RAIS.

A inspeção inicial da tabela permite constatar um elevado grau de heterogeneidade entre os setores, com ênfase especial e mais visível para as variáveis tamanho médio de plantas. Como esses indicadores constituem boas *proxies* para a incidência de barreiras de escala à entrada de novas firmas, pode-se constatar, de imediato, que a indústria brasileira revela marcada diferenciação entre os diferentes setores.⁶ Menos nítidas, entretanto, são as variações intersectoriais apresentadas pelas demais variáveis, razão pela qual se recorrerá à Tabela 1, que apresenta as diferentes estatísticas descritivas das variáveis citadas, incluindo como medida de variabilidade o coeficiente de variação (CV).

TABELA 1 – INDICADORES DE DINÂMICA INDUSTRIAL: ESTATÍSTICAS DESCRITIVAS

| Variável | Mínimo | Máximo | Média | CV* |
|----------|--------|----------|--------|--------|
| TMP96 | 4,016 | 1011.375 | 55,167 | 184,92 |
| TMP97 | 4,627 | 897,056 | 52,569 | 180,17 |
| TENT96 | 0,040 | 0,423 | 0,192 | 35,16 |
| TENT97 | 0,023 | 0,436 | 0,217 | 32,95 |
| TSAI96 | 0,004 | 0,475 | 0,186 | 36,99 |
| TSAI97 | 0,081 | 0,507 | 0,189 | 33,12 |
| TURBE96 | 0,000 | 0,117 | 0,006 | 166,67 |
| TURBE97 | 0,000 | 0,083 | 0,009 | 155,56 |
| TURBF96 | 0,155 | 0,731 | 0,377 | 27,62 |
| TURBF97 | 0,183 | 0,754 | 0,406 | 25,18 |

(*): coeficiente de variação.

6 O reconhecimento da importância da escala de produção como fonte relevante de barreira à entrada vem sendo manifestado na literatura pelo menos desde BAIN (1956). Mais recentemente, GILBERT (1989), em sua abrangente resenha sobre barreiras à entrada, evidencia que o fator escala continua sendo de importância fundamental como condicionante da entrada de firmas. No presente estudo, as variáveis TMP96 e TMP97 pretendem aproximar as desvantagens incorridas por firmas menores que pretendessem entrar em um dado setor.

Como se pode observar, os resultados confirmam a percepção inicial relativa à variabilidade dos tamanhos de plantas, que variam de valor mínimo de cerca de 4 ao valor máximo de aproximadamente 1011, com coeficiente de variação de aproximadamente 180%. Taxas de entrada (TENT) e taxas de saída (TSAI) de firmas apresentam coeficientes de variação no intervalo de 30%-37%, valendo notar, a respeito, os expressivos coeficientes de variação das variáveis de turbulência do emprego (TURBE), de 166,67% e 155,56%, para 1996 e 1997, respectivamente. Esses valores contrastam com os coeficientes de variação das variáveis de turbulência de firmas (TURBF), de 27,62% e de 25,18%, para 1996 e 1997, respectivamente. Este último resultado é de certo modo esperado, já que a criação e destruição de firmas tendem a ser relativamente mais rígidas. Os dados da Tabela 1 merecem ser complementados pelas informações constantes da Tabela 2, que exhibe as correlações bivariadas entre as variáveis consideradas no estudo. No que diz respeito à Tabela 2, vale notar, em primeiro lugar, a ocorrência de coeficientes de correlação negativos (e significativos) entre variáveis tamanho de plantas e taxas de entrada de firmas, o que sugere, em parte e para 1997, a interpretação da variável tamanho de plantas como *proxy* de escala eficiente mínima e, portanto, de barreira à entrada. Mais importante, talvez, sejam os coeficientes de correlação obtidos para as variáveis taxas de entrada e taxas de saída de firmas.

TABELA 2 – MATRIZ DE CORRELAÇÕES DAS VARIÁVEIS

| | TMP96 | TMP97 | TENT96 | TENT97 | TSAI96 | TSAI97 | TURBE96 | TURBE97 | TURBF96 | TURBF97 |
|---------|-------|---------|--------|----------|---------|---------|---------|---------|---------|----------|
| TMP96 | 1.000 | 0.986** | 0.097 | -0.158* | 0.020 | -0.078 | -0.155* | -0.086 | -0.050 | -0.156* |
| TMP97 | | 1.000 | -0.084 | -0.187** | 0.014 | -0.065 | -0.155* | -0.081 | -0.045 | -0.169** |
| TENT96 | | | 1.000 | 0.435** | 0.166** | 0.401** | 0.337** | 0.244** | 0.758** | 0.545** |
| TENT97 | | | | 1.000 | 0.227** | 0.177** | 0.178** | 0.284** | 0.432** | 0.798** |
| TSAI96 | | | | | 1.000 | 0.458** | 0.127* | 0.218** | 0.769** | 0.437** |
| TSAI97 | | | | | | 1.000 | 0.322** | 0.382** | 0.563** | 0.735** |
| TURBE96 | | | | | | | 1.000 | 0.237** | 0.302** | 0.320** |
| TURBE97 | | | | | | | | 1.000 | 0.302** | 0.430** |
| TURBF96 | | | | | | | | | 1.000 | 0.643** |
| TURBF97 | | | | | | | | | | 1.000 |

Obs: (*) correlação significativa a 5 %; (**) correlação significativa a 1%.

O fenômeno observado, de correlações positivas (elevadas) entre taxas brutas de entrada e taxas brutas de saída de firmas, foi detalhadamente estudado por Dunne *et al.* (1988), para o caso do setor industrial americano. Sutton (1997) discute o problema da turbulência como uma das quatro regularidades importantes relacionadas ao estudo de idade e tamanho de firmas como determinantes do crescimento.⁷ Segundo os autores, a maior parte dos padrões observados de entrada e saída de firmas seriam correlacionados de maneira significativa⁸ e teriam pouco efeito sobre as maiores firmas das indústrias consideradas. Além disso, as comparações entre países indicam que existiria alguma correspondência entre o ordenamento de indústrias em diferentes países segundo medidas de turbulência, o que, por sua vez, reforça a percepção de que existiriam fatores sistemáticos e específicos de indústrias que explicariam os níveis de turbulência.

O tratamento completo dessa questão específica foge ao escopo do presente estudo, valendo, entretanto, chamar a atenção, mais uma vez, para os coeficientes de variação das variáveis turbulência de emprego e turbulência de firmas, assim como para os coeficientes de correlação obtidos entre as mesmas. A turbulência do emprego (TURBE), cujo coeficiente de variação é substancialmente elevado, exhibe correlações positivas significativas com as taxas de entrada e taxas de saída (TENT e TSAI, respectivamente), sem se revelarem, entretanto, muito elevadas, como se as variações de emprego carregassem dinâmica específica ou estivessem associadas a movimento de plantas e firmas de menor tamanho (conforme se comentará a seguir). No que diz respeito às variáveis de turbulência de firmas (TURBF), vale assinalar, inclusive para fins de contraste com os resultados anteriormente mencionados, as elevadas correlações com as variáveis taxas de entrada e taxas de saída de fir-

7 As demais regularidades diriam respeito a (1) tamanho e crescimento, (2) ciclo de vida de firmas e (3) saída repentina de firmas (*shakout*).

8 DUNNE *et al.* (1988) encontram variações substanciais das taxas de entrada e de saída entre indústrias a 4 dígitos, inclusive no âmbito de setores a 2 dígitos, com correlação positiva entre padrões de entrada e saída, nos anos estudados e ao longo do tempo. Como assinalam os autores, a persistência ao longo do tempo da co-variabilidade desses padrões sugere que fatores específicos de indústria afetam, por sua vez, os graus de entrada e saída de firmas. A título de comparação, os autores obtêm correlações entre taxas de entrada e taxas de saída que variam de 0.180 a 0.321, levando em conta os censos de 1963, 1967, 1972, 1977 e 1982. Para referências futuras, deve-se notar que os autores também encontram similaridades nos padrões de entrada e saída (medidos por taxas médias) de firmas nos setores a 2 dígitos. SUTTON (1997) relata a ocorrência de coeficiente de correlação de 0.796 para amostra de 95 indústrias no Reino Unido (cf. GEROSKI, 1991).

mas. Esses são, entretanto, resultados que receberão tratamento tentativamente mais completo em partes seguintes do presente texto.

A abordagem que se pretende adotar na próxima seção procura ir além das correlações bivariadas das variáveis, explorando cenários multidimensionais propiciados pelas variáveis disponíveis, e padrões de comportamento das indústrias que emergiriam da consideração conjunta das mesmas.. Para tanto, o estudo recorrerá à Análise Discriminante, que é técnica da Estatística Multivariada destinada à identificação de grupamentos de observações (no caso, os 243 setores selecionados da indústria de transformação brasileira) e de suas características. Isto é o que se pretende desenvolver na seção que se segue. Cabe ressaltar que a utilização de Análise Estatística Multivariada no âmbito de estudos em Organização Industrial é bastante restrita. Além de Rodrigues e Façanha (1992), merecem menção Baldwin e Gorecki (1994), que investigaram a associação entre medidas de concentração e mobilidade industrial no Canadá, mediante o emprego de Correlação Canônica, e ainda Resende (1995), que identificou redundâncias entre diferentes medidas de concentração industrial para os Estados Unidos por meio da técnica de Componentes Principais.

2. ANÁLISE MULTIVARIADA: TESTANDO REGULARIDADES SETORIAIS

2.1 Metodologia

Como já assinalado, os dados da tabela do anexo serão analisados por meio da Análise Discriminante – AD, e a presente subseção fornecerá elementos básicos para o entendimento da mesma, extraídos principalmente de Kleka (1980) e Manly (1994). Referências técnicas adicionais podem ser obtidas em Rodrigues e Façanha (1992). Para os propósitos do estudo, é importante mencionar que a AD tem por objetivo precípuo a avaliação das possibilidades de se obter separação satisfatória de indivíduos (no caso, os setores da indústria utilizados), e grupamentos dos mesmos, a partir de medidas de características e de variáveis relativas a cada um dos indivíduos considerados. No exercício que será desenvolvido a seguir, os setores foram pré-cate-

gorizados em 4 grupos. Assim sendo, tudo se passa como se o estudo dispusesse de 4 (quatro) amostras aleatórias relativas aos grupos j , com $j = 1, \dots, 4$, cujos números de observações serão denotados pelo índice i , com i variando de 1 a n_j no grupo j . As 10 variáveis disponíveis serão genericamente denotadas por $X_1 \dots X_p$, com $p = 10$ na presente aplicação, e suas realizações específicas serão referidas por X_{ijr} (para $r = 1, \dots, p$). Com esses esclarecimentos básicos, deve-se mencionar que o objetivo da AD é o de obter funções discriminantes (canônicas), relativas aos casos e aos grupos, que serão representadas da forma que se segue:

$$Y_{ij} = c_0 + c_1 X_{ij1} + \dots + c_p X_{ijp} \quad (1)$$

na qual, como já assinalado, Y_{ij} é o valor da função discriminante para o caso i no grupo j ; os X 's representam os valores das variáveis discriminatórias (identificadas pelo primeiro subscrito) e os c 's são coeficientes que produzirão características desejáveis na função discriminante. Os coeficientes da primeira função discriminante são calculados de forma a tornar a média dos grupos tão diferentes quanto possível. A segunda função discriminante obedece ao mesmo princípio, com o requisito adicional de ortogonalidade das combinações lineares, e assim sucessivamente para as demais funções discriminantes.

Mais especificamente, a AD utiliza-se da matriz de variação/dispersão total T , da matriz de variação W interna aos grupos e da matriz de variação B entre grupos. Naturalmente, $B = T - W$, e o cálculo dos c 's procura tornar a relação entre os elementos de B e os elementos de W , ou $W^{-1} B$ tão elevada quanto possível, assegurando que os valores das funções sejam não-correlacionados.⁹ Ou seja, a AD busca um conjunto de combinações lineares das variáveis, cujos valores são tão próximos quanto possível no interior dos grupos, e tão afastados quanto possível quando referidos aos grupos considerados (ou seja, variância mínima nos grupos e variância máxima entre grupos). Conseqüentemente, para dada observação, os valores de Y_{ij} , (o

9 Onde $t_{su} = \sum_j \sum_i (X_{ijs} - \bar{X}_s)(X_{iju} - \bar{X}_u)$ denota um elemento típico da matriz T de soma de quadrados e produtos cruzados e $w_{su} = \sum_j \sum_i (X_{ijs} - \bar{X}_{js})(X_{iju} - \bar{X}_{ju})$ representa um elemento típico da matriz W de somas de quadrados e produtos cruzados no interior de cada grupo.

valor da função discriminante para o caso i no grupo j) representam as coordenadas da observação no espaço das funções discriminantes.

Os valores dos c 's são alcançados por meio de operações matemáticas relativamente complexas, cujo detalhamento foge ao escopo do presente trabalho, valendo notar, entretanto, que nos resultados que se apresentará na subseção seguinte o leitor encontrará valores para **autovalores** e a porcentagem da variabilidade total dos dados obtida por meio dos **autovetores** correspondentes, que são os valores dos c 's para cada função discriminante. Vale ainda ressaltar que os autovalores maiores estarão indicando que maior será a separação de grupos (maior a relação entre B e W). A função discriminante pode ser interpretada como eixo coordenado, já que os valores observados das variáveis para cada caso podem ser transformados num valor dado pela função, que será a posição do caso no eixo. Assim sendo, funções discriminantes (e três delas estarão sendo obtidas a seguir) representam eixos coordenados – ortogonais entre si, na medida em que cada uma das funções carrega informação diferente das demais – e cada observação, que inicialmente estaria localizada no espaço p – dimensional das variáveis, passará a estar representado no espaço que tem a dimensão do total de grupos menos um. Vale dizer, no exercício que se segue, no qual se utilizará 4 grupos de observações, cada observação pode ser localizada no espaço tridimensional definido pelas funções. Naturalmente, quando se obtém mais de uma função discriminante, é importante examinar as magnitudes relativas do poder discriminatório das mesmas. Os resultados apresentam as porcentagens da variância total dos dados explicada por cada função, assim como a porcentagem acumulada. Esta é, sem dúvida, uma das mais importantes informações a serem avaliadas, uma vez que porcentagem pequena de poder discriminatório associada à função indicam que a mesma estaria adicionando pouco entendimento a respeito da diferença entre grupos. A par disso, constarão também dos resultados os valores do **coeficiente de correlação canônica**, que é medida sintética do grau de associação dos grupos com cada uma das funções discriminantes. Um valor elevado (próximo de 1) indica que a função discriminante a que se refere o autovalor tem forte poder de discriminação dos grupos. Caso os grupos não sejam muito diferentes à luz das variáveis consideradas, o coeficiente de correlação canônica será baixo. Adicionalmente, os resultados também fornecerão valores para o *Wilk's lambda*, que é, também, medida de discriminação entre os grupos alcançada

pelas funções e variáveis utilizadas. Como se trata de medida da variância nos grupos em relação à variância total, valores baixos do *Wilk's lambda* denotam que as variáveis utilizadas diferenciam satisfatoriamente os grupos, valendo notar que o primeiro valor da medida refere-se à capacidade de discriminação das variáveis antes da estimação das funções correspondentes. Por sua vez, a “discriminação residual”, que se refere à capacidade de discriminação depois de estimada uma ou mais funções, pode ser avaliada por intermédio dos valores do **qui-quadrado** e dos valores de prova associados aos mesmos.

A literatura empírica costuma relativizar a importância dos testes estatísticos mencionados, especialmente quando os dados não constituem dados amostrais. Em se tratando de dados para populações de objetos, maior importância é atribuída à caracterização dos grupos e ao exame das relações entre as variáveis individuais e os valores das funções. Com esse intuito, no lugar das funções discriminantes, que são, em geral, de difícil interpretação (por encerrarem correlações entre as variáveis e das variáveis com os valores das funções), a próxima subseção fornecerá, em lugar das funções discriminantes, a **matriz de estrutura** (conforme esclarecimentos a seguir), que exhibe as correlações bivariadas entre os valores das funções discriminantes e os valores das variáveis (correlações estas que são depuradas das correlações entre variáveis).

2.2. Resultados Empíricos

A exploração multivariada dos dados da tabela do anexo desenvolveu-se a partir de tentativas iniciais de categorização dos setores. Métodos de grupamento (*clustering*) foram empregados, com resultados bastante insatisfatórios (ver, entretanto, RODRIGUES & FAÇANHA, *op.cit.*, para cotejo), assim como grupamentos baseados nos setores a 2 dígitos (com 18 grupos de indústrias), por se tratar de alternativa mais convencional. No caso, os resultados também foram convincentemente insatisfatórios, excetuando o fato dos grupamentos obtidos com a AD sugerirem a redução do número de grupamentos e a tentativa de se explorar a categorização dos setores a 4 dígitos segundo os “grupos de uso” final dos bens produzidos.¹⁰ Assim sendo,

10 Procedimento semelhante foi adotado em TAVARES *et al.* (1978).

foram definidos 4 grupamentos, denotados por (1) bens de consumo, (2) partes e componentes, (3) intermediários e (4) bens de capital, conforme indicado na tabela do anexo.

A AD pretende identificar padrões de comportamento que sejam particulares a essas classes de setores. Vale dizer, pretende-se examinar em que medida a ampla heterogeneidade observada na tabela do anexo estaria associada a fatores comuns a setores de uma mesma classe em vez de serem específicos a setores individuais, como parece ser o caso nos Estados Unidos.

Os resultados iniciais estão apresentados nas Tabelas 3 e 4, a seguir:

TABELA 3 – AUTOVALORES

| Função | Autovalor | % Variância | % Cumulativa | Correlação Canônica |
|--------|-----------|-------------|--------------|---------------------|
| 1 | 0,303 | 83,2 | 83,2 | 0,483 |
| 2 | 0,042 | 11,4 | 94,7 | 0,200 |
| 3 | 0,019 | 5,3 | 100,0 | 0,138 |

TABELA 4 – WILK'S LAMBDA E QUI-QUADRADO

| Teste de funções | Wilk's lambda | Qui-quadrado | Graus de liberdade | Significância |
|------------------|---------------|--------------|--------------------|---------------|
| 1 a 3 | 0,722 | 76,727 | 24 | 0,000 |
| 2 a 3 | 0,942 | 14,176 | 14 | 0,437 |
| 3 | 0,981 | 4,534 | 6 | 0,605 |

Como se pode notar, no caso de grupamentos por grupos de uso dos bens, os resultados estatísticos não são totalmente convincentes (ver qualificações a seguir), observando-se, de início, valores baixos para os autovalores. Como já foi assinalado, valores baixos para os autovalores indicam que os grupos são pouco discriminados via funções correspondentes. Em particular, a primeira função discriminante explica 83,2% da variação dos dados observados, sendo a única função para a qual se observa valor mais expressivo para o coeficiente de correlação canônica. Complementarmente, os resul-

tados fornecem valores elevados para o *Wilk's lambda*, com exceção, digna de ênfase, para o valor e significância do qui-quadrado associado ao grau de discriminação das variáveis antes de estimadas as funções, o que justifica em parte, pelo menos, recorrer-se à estimação da primeira função. De maneira geral, a regra operacional da AD recomenda que a derivação de funções prossiga até que a discriminação residual (medida pelo *Wilk's lambda*) se torne não significativa.

No caso, a apresentação dos resultados completos obedeceu a propósitos heurísticos, ao fato das demais funções adicionarem algum grau de explicação à variabilidade dos dados e, adicionalmente, ao fato do estudo tratar de dados populacionais, para os quais a **matriz de estrutura** constitui elemento de avaliação talvez tão importante quanto os testes estatísticos. Os resultados associados à matriz estão apresentados a seguir, na Tabela 5, da qual foram omitidos valores menores e de menor significação, para melhor apreensão visual das informações, valendo lembrar, a propósito, que valores absolutos baixos dos coeficientes indicam que a função tem muito pouco em comum com a variável em questão.

TABELA 5 – MATRIZ DE ESTRUTURA

| Variáveis | Função 1 | Função 2 | Função 3 |
|-----------|----------|----------|----------|
| TENT97 | 0,889 | | |
| TURBF97 | 0,737 | | |
| TURBF96 | 0,643 | | |
| TENT96 | 0,476 | | |
| TSAI96 | | 0,520 | |
| TURBE96 | | -0,276 | |
| TSAI97 | | 0,366 | |
| TEMP96 | | | 0,447 |
| TEMP97 | | | 0,358 |
| TURBE97 | | | -0,210 |

Antes de passar à caracterização tentativa dos grupos e, antes disso, à melhor caracterização das funções discriminantes, convém lembrar ao leitor que os resultados apresentados na Tabela 6 referem-se a correlações bivariadas entre as variáveis e o valor das funções, o que depura as informações das correlações, já comentadas, entre as variáveis. Vale dizer, as funções discriminantes calculadas (usualmente de difícil interpretação) forneceriam a localização das observações no espaço tridimensional das funções, como já comentado. Entretanto, e dado que a discriminação entre os grupos revelou-se precária, preferiu-se recorrer à correlação das variáveis com os valores das funções, o que constitui procedimento convencionalmente recomendado pela AD. Assim sendo, pode-se observar que a função 1 estaria altamente correlacionada com taxas de entrada de firmas, e com a turbulência de firmas (TENT96, TENT97, TURBF96 e TURBF97, respectivamente) nos dois anos considerados. À luz de comentários anteriores, pode-se dizer que o movimento de entrada de firmas nos setores é o principal fator de separação das observações, segundo a primeira função discriminante.

Com menor poder discriminatório, a função discriminante 2 separa os setores predominantemente pelas taxas de saída de firmas (TSAI96 e TSAI97) e pela turbulência de emprego (TURBE96), devendo-se notar, a propósito, a correlação negativa significativa desta última variável com a função discriminante. Na função 3 prevalecem os tamanhos médios de planta (TEMP96 e TEMP97) e a correlação negativa do valor da função discriminante com a turbulência do emprego (TURBE97). Note-se que os setores discriminados por meio da segunda função discriminante, que se caracterizam por taxas de saída expressivas, constituem também setores em que é menor a turbulência no emprego, o que denotaria a ocorrência de custos irrecuperáveis menores, e saída de firmas de tamanho menor, cujo comportamento não chegaria a afetar significativamente os setores de forma a também caracterizá-los como de elevada turbulência no emprego. Por sua vez, nos setores discriminados pela função discriminante 3, e mais fortemente por tamanhos médios de plantas maiores, observa-se também maior estabilidade nos níveis de emprego, o que pode estar denotando, mais uma vez, a ocorrência de mobilidade de plantas e firmas de menor tamanho (ver comentários a seguir). A Tabela 5 omitiu dados de menor significação, mas é útil mencionar que os valores da função discriminante 3 revela correlação positiva de 0,347

com a taxa de saída de firmas para 1997, e correlação negativa com a taxa de entrada de firmas, o que reforça a interpretação apresentada anteriormente.

Finalmente, e como complemento aos comentários anteriores, a Tabela 6 exibe as pré e pós-classificações dos setores, a partir da categorização adotada *a priori*, e segundo a classificação obtida por meio da AD, respectivamente.

TABELA 6 – PRÉ E PÓS-CLASSIFICAÇÃO DOS SETORES

| Grupos | 1* | 2* | 3* | 4* | Totais |
|--------|----|----|----|----|--------|
| 1 | 62 | 11 | 20 | 9 | 102 |
| 2 | 14 | 16 | 16 | 8 | 54 |
| 3 | 5 | 9 | 31 | 15 | 60 |
| 4 | 0 | 5 | 10 | 12 | 27 |
| Totais | 81 | 41 | 77 | 44 | 243 |

(*) Referem-se aos grupos nos quais os setores foram pós-classificados pela Análise Discriminante.

Esses resultados suprem, em parte, tentativas de se representar, por meio de recursos gráficos, a nuvem de pontos produzida pelas funções discriminantes, que se revelariam pouco esclarecedoras, por razões já mencionadas, associadas à precariedade do poder discriminatório entre grupos das funções. No que diz respeito à Tabela 6, deve-se notar, inicialmente, que apenas 49,8% dos 243 setores foram “corretamente” classificados. Observa-se que o grupo 1, de bens de consumo, destaca-se em termos de classificações corretas, com aproximadamente 60% de casos bem classificados e, em menor grau, o grupo 3, de bens intermediários, com cerca de 50% dos casos corretamente classificados. No grupo 4, a porcentagem de casos corretamente classificados está próxima dos 45%, mas no grupo 2 as taxas de acertos revelaram-se bem mais baixas. Isto quer dizer que (v. KLEKA, 1980, p. 51) a análise estaria incorrendo em apenas 33% menos erros do que no caso de a classificação dos setores ser definida aleatoriamente (com probabilidade *a priori* de $\frac{1}{4}$ para cada grupo). No entanto, outros comentários, de caráter mais exploratório, merecem ser destacados. Examinando-se o grupo 3 e o grupo 4 (de bens de capital) conjuntamente, pode-se notar que há uma taxa de acertos na classificação consolidada bastante significativa. Somando-se os

casos corretamente classificados nos grupos 3 e 4 com os casos pós-classificados nos grupos 3 e/ou 4, chega-se a um total de 68 casos pré ou pós-classificados nos grupos 3 ou 4. O total de setores pré-classificados nos grupos 3 e 4 é de 87 casos, o que aponta para uma taxa de sucesso na classificação destes casos de cerca de 78%. Tais qualificações não devem, entretanto, ser tomadas de forma conclusiva (aproximadamente 40% menos erros estariam sendo cometidos diante da classificação aleatória dos grupos), mas tão-somente exploratória e subsidiária à conclusão principal inicialmente apresentada. Na Tabela 1A o leitor encontra os grupos nos quais os setores foram inicialmente pós-classificados, o que permite avaliar diretamente a expressiva heterogeneidade dos grupamentos utilizados, assim como possibilidades recomendáveis de a classificação adotada vir a ser revisada, valendo notar que exercícios anteriores, com grupamentos de uso distintos, foram executados, com resultados menos satisfatórios do que os encontrados no presente trabalho.

Vale dizer, a despeito das ressalvas anteriormente apresentadas, os comentários a respeito da caracterização dos setores não devem ser aplicados automaticamente aos grupamentos de setores segundo os grupos de uso dos bens, devendo-se diagnosticar que inexistem um padrão inequívoco associado ao comportamento do setor industrial à luz das variáveis utilizadas no presente estudo. Os resultados obtidos não desfavorecem a tentativa de categorização adotada (e exercícios de aprimoramento das mesmas, desejavelmente com variáveis mais numerosas), mas recomendam, talvez mais convincentemente, retornar a percepções já assinaladas de que, à luz das variáveis aqui utilizadas, existiriam padrões específicos de comportamento setorial a serem melhor explorados. Algumas outras linhas futuras de trabalho podem ser destacadas de imediato, e dizem respeito, pelo menos, à obtenção de dados longitudinais, e acerca de outras características da concorrência não contempladas pelo presente estudo.¹¹

11 Para verificar a robustez dos resultados e comentários posteriores, consideramos uma definição alternativa de turbulência de firmas, conforme sugerido por um dos pareceristas. A medida calculada para 1996 e 1997 (respectivamente TURBP96 e TURBP97) foi definida em termos do produto da taxa de entrada e de saída de firmas, ao invés da medida baseada na soma, utilizada anteriormente. Os resultados obtidos foram bastante semelhantes. De fato, mesmo uma análise preliminar de correlação revelou expressivos coeficientes entre TURBF e TURBP, que nos dois anos ultrapassam 0.97. Adicionalmente, a AD não revelou diferenças dignas de nota, seja em termos das estatísticas, seja em termos da comparação entre pré e pós-classificações de setores. A esse respeito, por sinal, houve mais perdas do que ganhos em termos de resultados corretamente classificados. Os resultados completos podem ser solicitados dos autores.

3. COMENTÁRIOS FINAIS

O trabalho teve como objetivo primordial a investigação de diferentes aspectos do comportamento da indústria brasileira ao longo do período 1995-97. Para tanto, teve-se acesso a microdados da RAIS anteriormente não explorados nesse contexto. Tal base fornece dados no nível de 4 dígitos, e permite a consideração do universo de estabelecimentos industriais formais no Brasil. O mapeamento das variáveis relativas a tamanho médio de planta, taxas de entrada e saída e turbulência evidenciou um comportamento bastante heterogêneo das referidas variáveis entre os diferentes setores da indústria de transformação. Como desdobramento lógico, o trabalho explorou a vertente empírica de detecção de eventuais padrões de comportamento, que poderiam se manifestar de forma similar em diferentes grupos de setores. Nesse sentido, a utilização de Análise Discriminante mostrou-se útil, mas não satisfatoriamente convincente na identificação grupamentos de setores e de padrões de comportamento diferenciados para o conjunto de setores tratado no estudo. Tal diagnóstico é consistente com resultados de estudo de Dunne *et al.* (1988) para a economia norte-americana, que apontam a importância de fatores explicativos específicos a setores na determinação do comportamento e da dinâmica de indústrias.

REFERÊNCIAS

- BAIN, J. *Barriers to new competition*. Cambridge-MA: Harvard University Press, 1956.
- BALDWIN, J. R.; GORECKI, T. K. Concentration and mobility statistics in Canada's manufacturing sector. *Journal of Industrial Economics*, 42, p. 93-103, 1994.
- BEESELEY, M. E.; HAMILTON, R. T. Small firms' seedbed role and the concept of turbulence. *Journal of Industrial Economics*; 33, p. 217-231, 1984.
- DUNNE, T.; ROBERTS, M. J.; SAMUELSON, L. Patterns of firm entry and exit in U.S. manufacturing industries. *Rand Journal of Economics*, 19, p. 495-515, 1988.
- GEROSKI, P. A. *Market dynamics and entry*. Oxford: Basil Blackwell, 1991.

- _____. What do we know about entry? *International Journal of Industrial Organization*, 13, p. 421-40, 1995.
- GILBERT, R. J. Mobility barriers and the value of incumbency. In: SCHMALENSEE, R.; WILLIG, R. D. (eds.), *Handbook of industrial organization*. Amsterdam: North-Holland, 1989, p. 475-535.
- GÖRG, H.; STROBL, E.; RUANE, F. Determinants of firm start-up size: an application of quantile regression for Ireland. *Small Business Economics*, 14, p. 211-22, 2000.
- KLEKA, W. R. *Discriminant analysis*. Beverly Hills: Sage Publications, 1980.
- MANLY, B. F. J. *Multivariate statistical methods*. London: Chapman & Hall, 1994.
- MATA, J. Entrada e saída na indústria transformadora. *Economia*, 19, p. 57-86, 1995.
- MATA, J.; MACHADO, J. A. F. Firm start-up size: a conditional quantile approach. *European Economic Review*, 40, p. 1305-23, 1996.
- NAJBERG, S.; PUGA, F. P.; OLIVEIRA, P. A. S. Sobrevivência das firmas no Brasil: dez. 1995/dez. 1997. *Revista do BNDES*, 7, p. 33-48, 2000.
- RESENDE, M. On the characterization of the market structure in U.S. industry: a principal components analysis. *Economia*, 19, p. 87-94, 1995.
- RODRIGUES, D. A.; FAÇANHA, L. O. Indústria brasileira na década de 70: interpretação de resultados de estatística multivariada e de aspectos da dinâmica concorrencial. *Revista Brasileira de Economia*, 46, p. 447-476, 1992.
- SIEGFRIED, J. J.; EVANS, L. B. Empirical studies of entry and exit: a survey of the evidence. *Review of Industrial Organization*, 9, p. 121-155, 1994.
- SUTTON, J. Gibrat's legacy. *Journal of Economic Literature*, XXXV, p. 40-59, 1997.
- TAVARES, M. C.; POSSAS, M. L.; FAÇANHA, L. O. *Estrutura industrial e empresas líderes*. Financiadora de Estudos e Projetos – FINEP, Relatório de Pesquisa, 1978. Mimeografado.

Os autores agradecem à assistência de pesquisa de Renato Dirk e Bruno Caselli. L.O. Façanha e M. Resende agradecem respectivamente o apoio do CNPq e da FAPERJ. E ainda a colaboração de Vera Marina do Ministério do Trabalho no sentido de permitir acesso aos microdados utilizados neste estudo.

(Recebido em março de 2003. Aceito para publicação em julho de 2004).