


UNIVERSIDADE FEDERAL DE SANTA CATARINA
DEPARTAMENTO DE ENGENHARIA DE PRODUÇÃO E SISTEMAS

EVOLUÇÃO DA CONCENTRAÇÃO INDUSTRIAL NO
BRASIL (1973 - 77)



DISSERTAÇÃO SUBMETIDA À UNIVERSIDADE FEDERAL DE SANTA CATARINA
PARA OBTENÇÃO DO GRAU DE MESTRE EM ENGENHARIA

ENILSON MEDEIROS DOS SANTOS

FLORIANÓPOLIS
SANTA CATARINA - BRASIL
DEZEMBRO DE 1980

EVOLUÇÃO DA CONCENTRAÇÃO INDUSTRIAL NO
BRASIL (1973-77)

ENILSON MEDEIROS DOS SANTOS

ESTA DISSERTAÇÃO FOI JULGADA ADEQUADA PARA A OBTENÇÃO
DO TÍTULO DE

"MESTRE EM ENGENHARIA"

ESPECIALIDADE ENGENHARIA DE PRODUÇÃO E APROVADA EM SUA
FORMA FINAL PELO PROGRAMA DE PÓS-GRADUAÇÃO

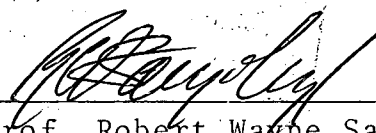


Prof. John Robert Mackness, Ph.D.

Coordenador do Programa

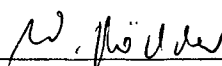
de Pós-Graduação em Engenharia de Produção

BANCA EXAMINADORA:

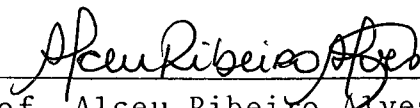


Prof. Robert Wayne Samohyl, Ph.D.

Presidente



Prof. Wilhelm Rödder, Ph.D.



Prof. Alceu Ribeiro Alves, M.Sc.



0.249.266-0

UFSC-BU

R E S U M O

A estrutura industrial brasileira é fortemente caracterizada pela presença de grandes empresas, em sua maioria transnacionais ou estatais. O presente trabalho visa analisar, a nível setorial, a evolução do grau de concentração na indústria de Transformação nacional, bem como o papel desempenhado por algumas variáveis na explicação dos níveis de concentração alcançados em um período recente da economia brasileira.

A fim de escolher os índices de concentração mais adequados ao estudo a ser desenvolvido, é feita uma análise das medidas de desigualdade disponíveis na literatura econômico-estatística, optando-se pela utilização do índice de Theil, do coeficiente de variação e do modelo log-logístico.

Posteriormente, aplica-se a análise de regressão para verificar os determinantes das diferenças inter-setoriais do grau de concentração, com destaque para o Departamento em que o setor está incluído e os investimentos estrangeiros e estatais no setor.

A B S T R A C T

The structure of the Brazilian industrial sector is strongly influenced by the presence of large firms which are either multinational or public. The present work analyses, on the sectorial level, the evolution of the degree of concentration in national manufacturing industry, and the role played by some variables in explaining this concentration in a recent period of the Brazilian economy (1973-1977).

In order to choose the most appropriate concentration index, an analysis is made of several inequality measures available in economic-statistical literature. Three measures - the Theil index, the coefficient of variation and the log-logistic model - were then chosen for application to the available data.

Afterwards, regression analysis is applied to examine the degree of sectorial concentration and its determinants, with special references to the Department where the sector is included and the foreign and public investments in that sector.

SUMÁRIO

	pag.
LISTA DE GRÁFICOS	viii
LISTA DE QUADROS	ix
 CAPÍTULO I	
1. INTRODUÇÃO	1
1.1. O Tema da Distribuição: Origem e Evolução	1
1.2. Uma Generalização Metodológica	2
1.3. Objetivos e Importância do Trabalho	3
1.4. Limitações do Trabalho	5
1.5. Estrutura do Trabalho	5
 CAPÍTULO II	
2. MEDIDAS DE CONCENTRAÇÃO E DESIGUALDADE	7
2.1. Generalidades	7
2.2. Características Gerais	8
2.3. Medidas de Dispersão: A Variância dos Logs	8
2.4. Coeficiente de Variação de Pearson	14
2.5. O Índice de Theil	18
2.6. Medidas Associadas à Curva de Lorenz	24
2.6.1. A razão de concentração de Gini	30
2.6.2. Outras medidas de desigualdade e curva de Lorenz	35
2.7. Considerações Gerais	40

CAPÍTULO III

3.	MODELOS DE DISTRIBUIÇÃO DE RENDA	42
3.1.	Generalidades	42
3.2.	A Função de Pareto	43
3.3.	A Distribuição log-Normal	46
3.4.	As Densidades Gama e Beta	51
3.5.	O Modelo de Singh-Maddala	53
3.6.	O Modelo Log-Logístico	57
3.7.	Considerações Gerais	63

CAPÍTULO IV

4.	EVOLUÇÃO RECENTE DA CONCENTRAÇÃO INDUSTRIAL NO BRASIL	65
4.1.	Introdução	65
4.2.	Observações Metodológicas	65
4.3.	Resultados Obtidos com o Índice de Theil	67
4.4.	Aplicação do Coeficiente de Variação e Resulta- dos Obtidos	79

CAPÍTULO V

5.	UMA ANÁLISE ESTRUTURAL DA CONCENTRAÇÃO INDUS- TRIAL NO BRASIL	84
5.1.	Introdução	84
5.2.	Aspectos Metodológicos Preliminares	86

	pag.
5.3. Obtenção dos Parâmetros da Função Log-Logística	88
5.4. Preliminares dos Modelos Testados	92
5.5. Análise da Evolução do Grau de Concentração Setorial	93
5.6. Determinantes do Grau de Concentração Setorial	101
5.7. Considerações Gerais	106
5.8. Recomendações para Futuras Pesquisas	107
REFERÊNCIAS BIBLIOGRÁFICAS	109

LISTA DE GRÁFICOS

		pag.
GRÁFICO 1	Uma Curva de Lorenz Genérica Mostrando a Linearização por Partes	27
GRÁFICO 2	Curvas de Lorenz que se Interceptam Fornecendo a Mesma Área de Concentração ...	33
GRÁFICO 3	Uma Curva de Lorenz Genérica Destacando o Ponto de Máxima Discrepância	39

LISTA DE QUADROS

	pag.
QUADRO 1	Subdivisão da Indústria de Transformação em Setores e Subsetores 68
QUADRO 2	Participação Relativa dos Setores e Subsetores no Total de Empresas e do Patrimônio Líquido da Indústria de Transformação 71
QUADRO 3	Valores do Índice de Theil por Subsetor para os Anos de 1973 e 1977 73
QUADRO 4	Valores do Índice de Theil por Setor e Global (1973 e 1977) 78
QUADRO 5	Valores do Coeficiente de Variação por Subsetor, Setor e Global (1973 e 1977).. 81
QUADRO 6	Subdivisão da Indústria de Transformação em Setores, Segundo a Secretaria da Receita Federal 87
QUADRO 7	Parâmetros da Curva Log-Logística por Setor e Ano 90
QUADRO 8	Principais Resultados Empíricos para a Evolução do Grau de Concentração Setorial 98

QUADRO 9	Principais Resultados Empíricos para o	
	Nível de Concentração Setorial	103

C A P I T U L O I

1. INTRODUÇÃO

1.1. O Tema da Distribuição: Origem e Evolução

O tema da repartição de um determinado atributo agregado entre os membros de uma população, hoje largamente considerado pelos economistas, parece ter sua origem na questão da distribuição funcional de renda. Ricardo considerava como "principal problema da Economia Política o de determinar as leis que regulam essa distribuição".¹

Tanto em Marx, originando o conceito de mais-valia, quanto nos neo-clássicos, assimilada implicitamente na estrutura das funções de produção, a remuneração diferenciada dos fatores produtivos tem estado no centro das contribuições teóricas ao conhecimento econômico.

Em fim do século passado e início deste, os trabalhos pioneiros de Lorenz, Corrado Gini e Vilfredo Pareto chamaram a atenção para o problema da distribuição de renda pessoal, acenando com a perspectiva de medidas de concentração ou desigualdade na repartição desse atributo. Em contrapartida à abordagem da distribuição (funcional) de renda por tipo, cuja estrutura teórica, em geral, não foi verificada empiricamente, os estudos das desigualdades interpessoais de renda se voltaram desde o início para a análise de processos distributivos reais.²

¹ Ricardo, David. "Principles of Political Economy" em Piero Straff (ed), "Words and Correspondence of David Ricardo" (Cambridge University Press, 1951)

² A "lei de Pareto" foi desenvolvida a partir da análise de dados de distribuição de renda.

Essa ênfase na aferição quantitativa dos processos de distribuição de renda, aliada ao debate político que facilmente se instaura quando a questão é mencionada, de certo modo justifica a negligência com que o tema foi tratado pelos economistas teóricos até recentemente. A partir de 1970, entretanto, pode-se notar que os trabalhos divulgados na literatura econômica sobre avaliação de processos distributivos têm se caracterizado pela preocupação com a fundamentação e interpretação econômica dos modelos, sem prejuízo à verificação empírica dos mesmos.

O interesse mais consequente pelo tema parece estar ligado à constatação de que "após duas décadas³ de empenho em relação ao desenvolvimento econômico em todo o mundo, mesmo o crescimento elevado não significa necessariamente uma melhoria da pobreza absoluta ou relativa".⁴

1.2. Uma Generalização Metodológica

Todo instrumental analítico, desenvolvido para a abordagem do problema da distribuição de renda pessoal, pode ser aplicado ao estudo dos graus de concentração e desigualdade de outros atributos que não a renda. Na verdade, os fundamentos teóricos desses modelos e medidas, por serem de natureza essencialmente estatística e probabilística, tornam possível sua aplicação a qualquer atributo agregado que, pelo menos a um nível abstrato, se possa imaginar como igualmente distribuído entre os elementos de uma população.

³ Anos 50 e 60

⁴ COSTA, Ramonaval A., "Distribuição da Renda Pessoal no Brasil - 1970". IBGE (Centro Editorial), Rio de Janeiro, 1977.

Os trabalhos já desenvolvidos nesse sentido incluem o uso das medidas e modelos de desigualdade em atributos tão diversos quanto o capital de firmas ou a área de terra de proprietários rurais ⁵. A experiência parece mostrar que, mesmo quando se trata de um modelo desenvolvido a partir de suposições sobre a distribuição da renda pessoal, ele permanece eficiente para explicar outros tipos de desigualdade econômica. Isto indica que, em geral, os pressupostos desses modelos também se aplicariam a outros processos distributivos, em especial àqueles associados à estrutura industrial como o do estoque de capital e da receita ou valor de produção entre firmas.

Numa época em que, mais do que nunca, ressalta-se o papel da pequena e média empresa no processo de crescimento econômico, principalmente, dos países de capitalismo tardio e verifica-se os efeitos nocivos do oligopólio sobre a distribuição de renda funcional e pessoal ⁶, o uso dos modelos e medidas de desigualdade para o estudo da concentração industrial fornece um instrumento coadjuvante de grande importância para a análise de estruturas econômicas.

1.3. Objetivos e Importância do Trabalho

A proposta deste trabalho pode ser colocada em dois níveis. No primeiro, pretende-se examinar as principais medidas de desigualdade e modelos de distribuição de renda propostos desde Pareto, sob o prisma de sua aplicabilidade potencial ao estu-

⁵ Por exemplo, Sabóia (1977) e Hoffman (1971)

⁶ Tolipan & Tinelli (1975), Considera (1980) e Fishlow (1972).

do da repartição de outros atributos; no segundo, à luz da análise anterior, estuda-se a evolução da estrutura industrial brasileira - indústria de transformação - quanto aos seus graus de concentração global e setorial durante o período 1973-1978.

Dessa forma, objetiva-se trazer uma contribuição efetiva ao entendimento mais exato da natureza das medidas e modelos de distribuição, expandindo seu campo de aplicabilidade, ao mesmo tempo em que, focalizando um período recente da economia nacional, procura-se fornecer subsídios para uma melhor compreensão dos mecanismos que regem as variações do grau de concentração industrial no Brasil.

Esse duplo objetivo, ao invés de fragmentar o trabalho, deve concorrer para sua unidade e consistência na medida em que os dois níveis aqui tratados se relacionam e se fundamentam mutuamente: pelo fornecimento da estrutura metodológica - da análise dos modelos para a aplicação - e, na direção inversa, pela verificação empírica da validade do instrumental analítico usado.

Assim, a importância deste trabalho pode ser medida não só pela atualidade e interesse do tema concentração industrial ou pela oportunidade de um estudo elucidativo das medidas de desigualdade e modelos de distribuição mais importantes, mas também pela sistematização de um procedimento cujas potencialidades ainda não foram completamente exploradas. A metodologia aqui aplicada, qual seja o uso generalizado - mas não indiscriminado - dos índices e modelos desenvolvidos com vistas às desigualdades de renda pessoal no estudo da estrutura industrial, pode ser utilizada em outras situações onde o interesse da pesquisa seja acompanhar a evolução, no sentido temporal ou "cross-section" ou

ambos, de um processo distributivo qualquer.

1.4. Limitações do Trabalho

Ao estudar um período recente da economia brasileira, o pesquisador invariavelmente se defronta com a dificuldade de obtenção dos dados. Ainda não integradas de forma rigorosa, as observações encontram-se dispersas por várias publicações com diferentes metodologias de classificação. Advém daí a principal limitação desse trabalho: o arquivo de dados foi obtido de estatísticas que não se voltavam para um trabalho desse tipo. Entretanto, tendo sido estes dados utilizados com os cuidados necessários, essa limitação não chega a ameaçar os resultados e conclusões aqui obtidos.

Uma outra limitação, esta quanto ao conteúdo, deve ser ressaltada: o critério de seleção das medidas e modelos de desigualdade não está ligado à menor ou maior veiculação destes na literatura. Prende-se, em vez disso, a uma avaliação feita pelo autor do que seria importante ou não considerar e, mesmo entre os modelos e medidas incluídos, houve-se por bem dar maior destaque àqueles para os quais um exame mais detalhado foi julgado necessário.

1.5. Estrutura do Trabalho

Pode-se distinguir duas partes neste trabalho, além deste capítulo inicial cuja finalidade é apresentar e discutir o tema proposto e a forma de execução dessa proposta.

Na primeira parte, constituída pelos capítu -

los 2 e 3, estudam-se, respectivamente, as medidas de desigualdade/concentração e os modelos de distribuição mais importantes divulgados neste século. Procura-se, para cada um deles, ressaltar a sua essência e natureza teóricas, o grau de dificuldade e os requisitos metodológicos para sua computação e os cuidados indispensáveis a sua aplicação.

Na segunda parte, aplica-se, no capítulo 4, aos dados disponíveis para caracterizar a estrutura da indústria de transformação nacional, o instrumental desenvolvido na parte anterior. Relacionando-se os resultados obtidos para a variação do grau de concentração com variáveis econômico-conjunturais no período , chega-se no quinto capítulo a conclusões que, agregadas, propiciam uma melhor percepção do funcionamento da estrutura industrial do país.

2. MEDIDAS DE CONCENTRAÇÃO E DESIGUALDADE

2.1. Generalidades

A preocupação da literatura econômica em pesquisar índices de concentração e/ou desigualdade na distribuição de um atributo qualquer (renda pessoal ou familiar, produção industrial, propriedade rural, etc.) surge da necessidade de comparação entre estruturas econômicas diversas, temporal e/ou espacialmente, com respeito a um processo distributivo específico. Nesse aspecto, a evolução de índices apropriados para a avaliação quantitativa de uma determinada situação econômica é de suma importância tanto para fundamentar estudos de caráter analítico-descritivo quanto para orientar decisões de política e planejamento econômicos.

Em geral, a obtenção desses índices tem sido possível a partir de duas abordagens do problema. Numa delas, que pode ser denominada de abordagem probabilística, o primeiro passo é o desenvolvimento ou adoção de uma função de densidade ou de distribuição para descrever o processo distributivo em pauta. Os parâmetros dessas funções são, então, estimados a partir das observações. (Esse procedimento é detalhado no capítulo 3 deste trabalho). Em uma outra abordagem, de natureza estatística, os índices são formulados como medidas de dispersão ou concentração dos dados, sem que se faça necessária a análise da forma da distribuição. Nas seções seguintes, os índices de concentração e desigualdade provenientes dessa segunda abordagem são discutidos em detalhe.

2.2. Características Gerais

Os índices de concentração e desigualdade de natureza estatística são, em última análise, avaliações analíticas de um determinado processo distributivo. Por esse motivo, esses índices, enquanto medidas do grau de concentração, são necessariamente referenciados a situações-limite, quais sejam a inexistência de desigualdades ou a desigualdade plena na distribuição do atributo¹ em estudo. Por uma questão de consistência, essas situações extremas são associadas aos limites do intervalo de variação do índice, limites que, portanto, não devem ser usados como base de comparação. Apenas a ordem definida sobre o intervalo de variação do índice, indicando a direção em que o grau de concentração medido aumenta ou diminui, deve ser tomada em consideração. Logo, um valor isolado para um determinado índice de concentração pouco ou nada significa. Afinal, é da própria essência dessas medidas serem apenas instrumentos de análise comparativa e, só com esse intento, elas devem ser aplicadas.

2.3. Medidas de Dispersão: A Variância dos Logs

A avaliação de um processo distributivo através de uma medida de dispersão absoluta como o desvio médio, a variância ou a diferença média é a idéia mais simples possível para retratar as desigualdades na distribuição. Essa idéia tem sido bastante refutada pelo argumento de que o conceito de desigualdade envolve outras características da distribuição como a assimetria e mesmo a média².

¹ Para simplificação do texto, a palavra "renda" será usada a partir de agora, designando um atributo genérico

² Ver Morgan (1962)

Por outro lado, as medidas de dispersão absolutas são normalmente expressas em unidades da variável considerada. Dessa forma, elas são dependentes da escala usada, o que não favorece comparações entre populações diferentes. No caso de o atributo em questão ser uma unidade monetária, por exemplo, a inflação (diferenças temporais) ou a taxa de câmbio (diferenças espaciais) afetam as medidas de dispersão absoluta.

A tentativa de expurgar a influência da escala em estudos de distribuição é que deu origem a um índice de desigualdade que usa o atributo logaritmado. Mesmo que essa medida - a variância dos logs, como é conhecida - permaneça apenas avaliando a dispersão dos dados, ela representa um avanço com relação à variância ou desvio médio da renda e tem sido bastante aplicado desde 1933 e até recentemente ³.

Podemos definir a variância dos logs, tal como vem sendo usada, a partir da definição de variância de uma variável aleatória contínua. Assim, sendo $L = \log R$, onde R representa a variável renda, uma v.a. com função de distribuição $F(1)$ contínua e diferenciável, define-se como variância de L o segundo momento em torno da média aritmética m :

$$\sigma^2(L) = \int_{-\infty}^{\infty} (l - m)^2 dF(1)$$

Já que não é de interesse conhecer a distri-

³ Ver Langoni (1973)

buição $F(1)$, a variância de L pode ser estimada ⁴ a partir de uma amostra de K elementos por

$$S^2(L) = \frac{1}{K-1} \sum_{i=1}^K (L_i - \bar{L})^2 \quad (01)$$

onde $\bar{L} = \frac{1}{K} \sum_{i=1}^K L_i$ é o estimador não-viesado para a média da população.

Pode-se verificar que esse índice independe da escala adotada para a variável renda, desenvolvendo-se (01) como segue:

$$\begin{aligned} S^2(L) &= \frac{1}{K-1} \sum_{i=1}^K (\log R_i - \frac{1}{K} \sum_{i=1}^K \log R_i)^2 \\ S^2(L) &= \frac{1}{K-1} \sum_{i=1}^K (\log R_i - \log \left(\prod_{i=1}^K R_i \right)^{1/K})^2 \\ S^2(L) &= \frac{1}{K-1} \sum_{i=1}^K \left(\log \frac{R_i}{R^*} \right)^2 \end{aligned} \quad (02)$$

onde $R^* = \left(\prod_{i=1}^K R_i \right)^{1/K}$ é a média geométrica das rendas.

Como se vê, cada termo do somatório em (02) é adimensional já que leva em conta apenas diferenças relativas

⁴ Nesta exposição, preferiu-se usar o estimador não-viesado da variância populacional (01), contrariamente a literatura que postula para a variância dos logs o estimador $S^2(L) = \frac{1}{K} \sum_{i=1}^K (L_i - \bar{L})^2$. O autor justifica o seu procedimento pelo fato de que os dados são sempre parciais na divulgação dos censos e, portanto, o pesquisador lida com amostras da população, desconhecendo a média populacional.

de renda. O fato de não penalizar desigualdades absolutas de renda, como o fazem outros índices baseados na dispersão do próprio atributo, é uma das vantagens mais consideradas da variância dos logs.

Uma outra vantagem deste índice é a possibilidade de ser decomposto. Às vezes, interessa ao pesquisador conhecer separadamente as desigualdades intra e inter-estratos. Para isso, estipula-se uma partição em classes de renda do intervalo $[0, \infty)$, distribuindo-se os dados individuais pelos estratos correspondentes.

Através de algumas transformações algébricas em (01), pode-se obter uma decomposição do índice em dois fatores: um relativo aos efeitos das desigualdades intra-classes e outro atribuído às desigualdades inter-classes. Sumariamente, tem-se:

$$S^2(L) = \frac{1}{K-1} \sum_{i=1}^K (\log R_i - \bar{L})^2$$

$$S^2(L) = \frac{1}{K-1} \sum_{j=1}^m \left[\sum_{i_j=1}^{K_j} (\log R_{i_j} - \bar{L}_j)^2 + K_j (\bar{L}_j - \bar{L})^2 \right]$$

onde j ($j = 1, 2, \dots, m$) designa cada classe de renda, K_j representa o total de indivíduos na classe j e

$$\bar{L}_j = \frac{1}{K_j} \sum_{i_j=1}^{K_j} \log R_{i_j} \quad j = 1(1)m$$

é a média das observações em cada classe. Dando sequência, vem

$$S^2(L) = \frac{1}{K-1} \sum_{j=1}^m \sum_{i_j=1}^{K_j} \left[\left(\log \frac{R_{i_j}}{R_j} \right)^2 + \left(\log \frac{R_j}{R^*} \right)^2 + 2 \log \frac{R_{i_j}}{R_j} \log \frac{R_j}{R^*} \right]$$

onde R_j é tal que $\log R_j = \bar{L}_j$, ou seja,

$$R_j = \prod_{i_j=1}^{K_j} R_{i_j}^{1/K_j} \quad j = 1(1)m$$

$$S^2(L) = \frac{1}{K-1} \sum_{j=1}^m \sum_{i_j=1}^{K_j} \left(\log \frac{R_{i_j}}{R_j} \right)^2 + \frac{1}{K-1} \sum_{j=1}^m \sum_{i_j=1}^{K_j} \left(\log \frac{R_j}{R^*} \right)^2 + \frac{2}{K-1} \sum_{j=1}^m \log \frac{R_j}{R^*} \sum_{i_j=1}^{K_j} \log \frac{R_{i_j}}{R_j}$$

O último termo do lado direito da igualdade se anula pois $\sum_{i_j=1}^{K_j} \log \frac{R_{i_j}}{R_j} = 0$ uma vez que se trata da soma

de desvios com relação à média. Introduzindo na expressão remanescente um pequeno artifício, tem-se

$$S^2(L) = \sum_{j=1}^m \frac{K_j - 1}{K-1} \frac{1}{K_j - 1} \sum_{i_j=1}^{K_j} \left(\log \frac{R_{i_j}}{R_j} \right)^2 + \sum_{j=1}^m \frac{K_j}{K-1} \frac{1}{K_j} \sum_{i_j=1}^{K_j} \left(\log \frac{R_j}{R^*} \right)^2$$

$$\begin{aligned}
S^2(L) = & \sum_{j=1}^m \frac{K_j - 1}{K - 1} \left[\frac{1}{K_j - 1} \sum_{i_j=1}^{K_j} \left(\log \frac{R_{i_j}}{R_j} \right)^2 \right] + \\
& + \frac{1}{K - 1} \sum_{j=1}^m K_j \left(\log \frac{R_j}{R^*} \right)^2 \quad (03)
\end{aligned}$$

Observa-se na equação (03) que a variância total deve-se à variância intra-estratos, correspondente à primeira parcela à direita do sinal de igualdade, e à variância inter-classes, expressa pela segunda parcela. Evidentemente, essa decomposição da desigualdade medida pela variância dos logs só pode ser feita se os dados individuais estão disponíveis.

Quando só estão disponíveis os dados agrupados em classes de renda, a utilização do índice só é possível fazendo-se uma simplificação na parcela de (03) referente às desigualdades entre classes. Essa simplificação consiste em substituir a média ponderada das rendas em cada classe e total - R_j e R^* , respectivamente -, pela média aritmética em cada classe, \bar{R}_j , e renda per capita \bar{R} . Em certas ocasiões, a média aritmética na classe não pode ser obtida e pode ser usado o ponto médio da classe. É nessa forma que o índice é mais correntemente usado:

$$S^2(L) = \frac{1}{K - 1} \sum_{j=1}^m K_j \left(\log \frac{\bar{R}_j}{\bar{R}} \right)^2 \quad (04)$$

A rigor, porém, a quantidade calculada por (04) não deve levar a denominação "variância dos logs". Na verda-

de, ela somente aproxima a segunda parcela em (03) convenientemente quando tanto o intervalo de classe como a faixa de variação do atributo são pequenos. Por esse motivo, a comparabilidade entre valores de $S^2(L)$ obtidos por (04) fica afetada: o índice é muito dependente da partição adotada. Sua utilização, portanto, requer muitos cuidados.

Quando aplicado nas formas (02) e (03), além das vantagens já discutidas, o índice permite ao pesquisador tratar não só classes de renda, no sentido de partição do domínio do atributo, mas também sub-populações quaisquer. Como exemplo, a desigualdade total pode ser decomposta em desigualdade entre os indivíduos de mesma ocupação e desigualdade entre ocupações. Isto fornece um instrumento de análise de muitas possibilidades no estudo de estruturas industriais como estratificação da população em empresas transnacionais, estatais e nacionais privadas ou em setores industriais, por exemplo.

Por último, vale salientar que, usado em qualquer de suas formas, notadamente (02) e (03), os indivíduos com renda zero devem ser excluídos do cálculo.

2.4. Coeficiente de Variação de Pearson

O coeficiente de variação, atribuído a Karl Pearson, é uma medida de dispersão relativa, como tantas outras, algumas das quais são discutidas neste trabalho. Como tal, além da vantagem da adimensionalidade, que evita o efeito inflacionário, o índice ainda incorpora uma outra característica da distribuição, no caso, a média. Em termos de uma distribuição contínua, pode-se formulá-lo como

$$V = \sigma/\mu = \left[\int_{-\infty}^{\infty} (x - \mu)^2 dF(x) \right]^{1/2} / \int_{-\infty}^{\infty} x dF(x)$$

A partir de uma amostra de K elementos, o coeficiente de variação pode ser estimado por ⁵

$$\hat{V} = \frac{s}{\bar{X}} = \frac{\left[\frac{1}{K-1} \sum_i (X_i - \bar{X})^2 \right]^{1/2}}{\frac{1}{K} \sum_i X_i} \quad (05)$$

Conhecido desde 1933 como medida de desigualdade de renda, o coeficiente de variação, entretanto, não tem sido muito utilizado com este fim. Sua utilização porém, está disseminada em estudos da concentração industrial ⁶ e é particularmente interessante quando focaliza-se apenas as firmas a partir de um determinado tamanho. Mesmo assim, algumas de suas características importantes não têm sido notadas como era de se esperar.

Em primeiro lugar, essa medida tem um intervalo de variação finito. O limite inferior, corresponde à absoluta igualdade na distribuição (ou seja, cada elemento da população percebe a renda média), é o zero, já que o desvio-padrão σ de uma distribuição degenerada é nulo. O limite superior pode ser calcu-

⁵ Para simplificação do texto, a partir de agora, sempre que não houver possibilidade de dúvida, o símbolo $\sum_{i=1}^K$ será substituído por \sum_i ou mesmo Σ

⁶ Essa medida é aplicada no Brasil pela revista Visão na sua análise anual das "maiores firmas por setor" para avaliar o padrão de concentração de patrimônio líquido e faturamento.

lado facilmente para o caso de extrema desigualdade. Essa situação ocorre quando um único membro da população detém toda a renda R , enquanto os demais $N - 1$ indivíduos têm renda zero. Aplicando (05), temos ⁷

$$V = \frac{\sigma}{\mu} = \frac{1/N^{1/2} \left[(N-1) \left(\frac{R}{N}\right)^2 + \left(\frac{R}{N}\right)^2 (N-1)^2 \right]^{1/2}}{R/N}$$

$$V = \frac{\sigma}{\mu} = \frac{\left[\left(\frac{R^2}{N^2}\right) (N-1) \right]^{1/2}}{R/N} = \sqrt{N-1}$$

A dependência do número de elementos envolvidos não é interessante para um índice que se propõe a ser usado em comparações. Uma alternativa válida é a normalização do índice usando $V_s = V/\sqrt{N-1}$, obtendo-se assim uma medida com variação no intervalo $(0,1)$ e que viabiliza comparações entre populações de tamanhos diferentes.

Embora a decomposição de V não seja possível, a de V^2 o é. Seguindo a mesma linha dedutiva aplicada à variância dos logs na seção anterior, chega-se a ^{8,9}

⁷ Como estamos tratando da população inteira, não usamos o estimador \hat{V} de V . No caso de tratarmos com uma amostra de K elementos, o limite superior para \hat{V} é K . Os limites são compatíveis quando K tende para N .

⁸ A notação é similar à usada em (03), diferindo apenas na variável que, aqui, é o próprio atributo.

⁹ No caso de estarmos lidando com a população inteira, N e n_j devem substituir $(K-1)$ e (K_j-1) .

$$\hat{V}^2 = \frac{\sum_j \frac{K_j - 1}{K - 1} \left[\frac{1}{K_j - 1} \sum_{i_j=1}^{K_j} (X_{i_j} - X_j)^2 \right] + \frac{1}{K - 1} \sum_j K_j (X_j - \bar{X})^2}{\bar{X}^2}$$

onde a primeira parcela do numerador diz respeito às desigualdades intra-estratos, enquanto a segunda parcela refere-se às desigualdades inter-estratos. O uso conjunto de V e V^2 pode trazer informações substanciais ao analista de um processo distributivo.

No caso de dados grupados, o coeficiente de variação de Pearson, como a variância dos logs, desconhece as desigualdades dentro de cada classe de renda e é, portanto, calculado por

$$\hat{V} = \frac{\left[\frac{1}{K - 1} \sum_j K_j (X_j - \bar{X})^2 \right]^{1/2}}{\bar{X}}$$

que também pode ser normalizado por \sqrt{K} . Essa forma do índice permanece dependente da partição adotada e a generalização das correções de Sheppard (ver Kendall e Stuart, 1969) não é possível uma vez que, dificilmente, na divulgação em dados grupados de variáveis econômicas, o intervalo de classe é mantido constante ao longo da variação total do atributo.

O coeficiente de variação de Pearson, ou sua forma quadrática - proposta por Theil -, é essencialmente uma medida relativa de dispersão com uma interpretação bastante sim -

ples, inclusive do ponto de vista geométrico. Aplicada às desigualdades de renda, essa medida tem a desvantagem, comparativamente com a variância dos logs, de penalizar diferenças absolutas do atributo, o que às vezes não é exatamente o que importa, e a vantagem da não-exclusão dos elementos com renda zero.

Uma medida alternativa, o coeficiente de variação da variável $L = \log R$, baseada nas diferenças relativas de renda e incorporando a média, é dependente da escala adotada e, portanto, sua aplicação fica comprometida pela necessidade de deflação da variável renda, além de necessariamente excluir as rendas nulas.

Concluindo, o uso do coeficiente de variação como avaliador do grau de desigualdade de renda, incorpora algumas vantagens atraentes e não requer tantos cuidados na aplicação quanto a variância dos logs, embora ainda persista a questão das diferenças absolutas de renda, o que o torna pelo menos, discutível.

2.5. O Índice de Theil

O índice de Theil é uma medida de desigualdade relativamente recente, tendo sido divulgado com maior ênfase a partir de 1967, com a publicação de [40]. O desenvolvimento desse índice, devido a Henry Theil, fundamenta-se em alguns conceitos da Teoria da Informação, a saber:

a) Conteúdo de Informação: Seja E_i um evento cuja probabilidade de realização é $p(E_i)$. Define-se como conteúdo de informação da mensagem " E_i aconteceu" a função ¹⁰

¹⁰ Para simplificação do texto, escreve-se p_i para $P(E_i)$

$$h(p_i) = \log (1 / p_i) \quad 0 < p_i \leq 1$$

b) Entropia: Seja E_1, E_2, \dots, E_n uma partição do conjunto E , $p(E) = 1$ e $0 < p_i \leq 1$ para $i = 1(1)n$. Designam-se por p_i a probabilidade, "a priori", de ocorrência do evento E_i . Define-se como entropia ou valor esperado da informação, a função

$$H(p) = \sum_i p_i \cdot h(p_i) = -\sum_i p_i \log p_i$$

c) Informação esperada de mensagem indireta: Sejam E_i e p_i , $i = 1(1)n$, como antes e t_i , $i = 1(1)n$, as probabilidades a posteriori da realização de E_1, E_2, \dots, E_n . O conteúdo de informação de uma mensagem indireta é definido por

$$H(p_i, t_i) = h(p_i) - h(t_i) = \log (t_i/p_i)$$

e, dessa forma, a informação esperada de uma mensagem indireta pode ser calculada como

$$I(p, t) = \sum_i t_i \cdot \log (t_i/p_i)$$

Na formulação do índice de Theil, a fração de renda total que cada indivíduo receberia "a priori", igual a $1/n$, onde n é o número de elementos da população, é associada com as probabilidades a "priori" p_i . Por analogia similar, associa-se a fração v_i efetivamente percebida por cada elemento da população aos t_i acima definidos. O índice é, então, dado por

$$T(v, n) = \sum_i v_i \cdot \log \left(\frac{v_i}{1/n} \right) \quad (06)$$

A sua interpretação mais rica, em termos econômicos, é que ele significa a média ponderada (com base na repartição efetiva de renda) das distorções do processo distributivo. De outro modo, colocado na forma

$$T(v, n) = \log n - \sum_i v_i \log (1/v_i)$$

$$T(v, n) = \sum_i v_i \log \frac{1}{1/n} - \sum_i v_i \log (1/v_i)$$

ele pode ser interpretado, associando o conceito de entropia ao de repartição da renda. Neste caso, do valor da entropia a priori, $H(p) = \log n$, que representa o máximo de igualdade, é subtraído o valor da igualdade efetivamente encontrada, ou seja a entropia a posteriori $H(v)$.

O limite superior para o índice de Theil é alcançado quando considera-se o caso limite de desigualdade extrema. Nesta situação o indivíduo K apossa-se de toda a renda e

tem-se $v_k \rightarrow 1$ e $v_i \rightarrow 0$, para $i \neq k$. Aplicando a esta situação a equação (06), obtém-se¹¹

$$T(v, n) = v_k \log \left(\frac{v_k}{1/n} \right) = \log n$$

O limite inferior para o índice de Theil corresponde à situação de extrema igualdade, quando $v_i = 1/n$ para todo i . Neste caso, obtém-se

$$T(v, n) = \sum_i \frac{1}{n} \log \left(\frac{1/n}{1/n} \right) = 0$$

Como no caso do coeficiente de variação, a indesejável dependência de n para a fixação do intervalo de variação, dá lugar a uma normalização do índice para viabilizar comparações. Faz-se

$$T'(v) = T(v, n) / \log n \quad (07)$$

obtendo-se assim uma medida variando no intervalo (0;1) e não sendo afetada pelo tamanho da população.

Dada a sua formulação, que considera relações entre frações de renda, o índice de Theil independe da escala adotada, o que exclue o efeito da inflação. Além disso, leva a vantagem de considerar diferenças relativas do atributo o que

¹¹ Lembrando que $\lim_{x \rightarrow 0^+} (x \log x) = 0^-$

tem sido defendido, para o caso da renda, como mais razoável que considerar as diferenças absolutas.

Outro fator que concorre para sua importância é a possibilidade de decomposição do nível de desigualdade calculado por (06). Sendo n_j o número de elementos na classe j e

$v_j = \sum_{i_j=1}^{n_j} v_{i_j}$, a fração da renda total percebida por eles, segue:

$$T(v, n) = \sum_{i=1}^n v_i \log (nv_i)$$

$$T(v, n) = \sum_{j=1}^m \sum_{i_j=1}^{n_j} v_{i_j} \frac{v_j}{v_j} \log \left(nv_{i_j} \frac{v_j \cdot n_j}{n_j \cdot v_j} \right)$$

$$T(v, n) = \sum_j \sum_{i_j} v_{i_j} \frac{v_j}{v_j} \log \left(\frac{n \cdot v_j}{n_j} \right) + \sum_j \sum_{i_j} v_{i_j} \frac{v_j}{v_j} \log \left(\frac{v_{i_j} \cdot n_j}{v_j} \right)$$

$$T(v, n) = \sum_j v_j \sum_{i_j} \frac{v_{i_j}}{v_j} \log \left(\frac{nv_j}{n_j} \right) + \sum_j v_j \sum_{i_j} \frac{v_{i_j}}{v_j} \log \left(\frac{v_{i_j} \cdot n_j}{v_j} \right)$$

$$T(v, n) = \sum_j v_j \log \left(\frac{nv_j}{n_j} \right) + \sum_j v_j \sum_{i_j} \frac{v_{i_j}}{v_j} \log \left(\frac{v_{i_j} \cdot n_j}{v_j} \right)$$

Observando que, para $j = 1, 2, \dots, m$, tem -

se $\sum_{i_j} \frac{v_{i_j}}{v_j} = 1$, vem

$$T(v, n) = \sum_j v_j \log \left(\frac{v_j}{n_j/n} \right) + \sum_j v_j \sum_{i_j} (v_{i_j}/v_i) \log \left(\frac{v_{i_j} \cdot n_j}{1/n_j} \right) \quad (08)$$

onde a segunda parcela à direita da igualdade refere-se às desigualdades dentro dos estratos e a primeira parcela diz respeito às desigualdades entre os estratos.

No caso de os dados serem divulgados em classes de renda, o índice de Theil pode ser obtido sob a suposição que as desigualdades são nulas em cada estrato, o que equivale a considerar nula a segunda parcela em (08), obtendo-se a medida como

$$T(v, n) = \sum_j v_j \log \left(\frac{v_j}{n_j / n} \right)$$

O intervalo de variação do índice, calculado para dados grupados, permanece o mesmo $(0; \log n)$ e, portanto, pode ser normalizado como em (07).

Para dados grupados, o índice de Theil, como as outras medidas já analisadas, fica muito dependente da partição. Além disso, é necessária a exclusão dos indivíduos sem renda (renda nula) quando lida-se com observações individuais. Essas desvantagens, porém, não invalidam a aplicação da medida, particularmente atraente para estudos de concentração industrial, onde a inclusão da população com renda nula não faz necessariamente sentido pelo fato de que as firmas com capital nulo não existem e, no caso de faturamento zero, embora possível no caso de firmas em implantação ou em fase de paralização temporária, não é de interesse a inclusão destas unidades visto que não disputam efetivamente o mercado.

Em essência, o índice de Theil, apesar de ser uma medida de desigualdade relativamente sofisticada em sua formulação, avalia dispersões em torno da média da variável fração logaritmada de renda total e, introduzindo a ponderação, utiliza informações sobre o nível de assimetria dessas frações. Assim, representa um avanço sobre as demais medidas, principalmente, pela sua grande sensibilidade às mudanças na distribuição.

As formulações apresentadas para o índice de Theil não levam em conta a função de distribuição empírica da variável renda. Entretanto, o índice pode ser calculado conhecida a distribuição $F(x)$ por

$$T = \mu^{-1} \int_0^{\infty} x \log(x/\mu) dF(x)$$

onde $\mu = E[X]$ é a média da distribuição.

2.6. Medidas Associadas à Curva de Lorenz

A curva de Lorenz é um modelo gráfico destinado a analisar a concentração de um determinado atributo. Não obstante ter sido proposto visando as variáveis riqueza e renda, a sua utilização tem sido constante, também, como instrumento gerencial em inventário e produção (curva ABC). Sua derivação, em termos teóricos, é bastante simples e tem imediata interpretação econômica, como se vê a seguir.

Seja X uma variável aleatória contínua, com função densidade de probabilidade $f(x)$, necessariamente nula à

esquerda da origem ¹². Se a média μ da distribuição existe, ela é não-negativa, de formas que as variáveis F e L , definidas parametricamente em relação a x pelo sistema

$$F(x) = \int_{-\infty}^x f(t) dt$$

$$L(x) = \frac{1}{\mu} \int_{-\infty}^x t f(t) dt$$
(09)

variam no intervalo real $[0 ; 1]$. $F(x)$ e $L(x)$ representam, respectivamente, a função de distribuição e o momento incompleto de ordem um da variável X . Pode-se encarar o sistema (09) como estabelecendo uma relação entre as variáveis F e L . Cada ponto dessa relação, associa à percentagem da população com renda não superior a x , a fração da renda total correspondente. A curva $L(p)$ ¹³ tem características bem definidas, quais sejam:

- (i) A curva $L(p)$ é convexa com respeito ao eixo p
- (ii) $L(0) = 0$ e $L(1) = 1$
- (iii) $0 \leq L(p) \leq p$, $p \in [0 ; 1]$

A propriedade (i) pode ser demonstrada, verificando-se

¹² Alguns trabalhos recentes (Gastwirth, 1972, p.ex.) têm considerado rendas negativas no cálculo de medidas associadas à curva de Lorenz. Essa idéia não é considerada neste trabalho

¹³ Designa-se F por p para obter uma notação mais conveniente

$$\frac{dL}{dp} = \frac{dL}{dx} / \frac{dp}{dx} = \frac{x f(x)}{\mu f(x)} = \frac{x}{\mu} \geq 0$$

$$\frac{d^2L}{dp^2} = \frac{1}{\mu} / \frac{dp}{dx} = \frac{1}{\mu f(x)} \geq 0$$

e é equivalente, em termos econômicos, à constatação de que a um incremento dp em qualquer parcela p da população, a parcela $\Delta L(p)$ da renda total acrescentada é não-negativa e não-decrescente com p ¹⁴.

Quanto a (ii), basta fazer $x = 0$ em (09), observada a restrição imposta a $f(x)$, para obter $F(0) = 0$ e $L(0) = 0$. Da mesma forma, fazendo x tender a infinito (ou ao valor máximo finito da variável se a distribuição é truncada), tem-se $\lim_{x \rightarrow \infty} L(x) = 1$ e $\lim_{x \rightarrow \infty} F(x) = 1$.

Com relação à propriedade (iii), tem-se que o limite inferior para $L(p)$ é obtido considerando que $f(x)$ é nula à esquerda de $x = 0$, ou seja, não são consideradas rendas negativas. O limite superior é dedutível combinando (i) e (ii). Considerando a variável renda, pode-se interpretar economicamente (iii), observando que a percentagem da renda total $L(p)$ atribuída aos 100 p% mais pobres da população não pode superar p , caso contrário verifica-se inconsistência na ordenação das rendas individuais.

¹⁴ $\frac{dL(p)}{dp} = 0$ para $0 \leq p < t$, onde t é a percentagem da população com renda zero. No caso de $t > 0$, a curva de Lorenz coincide com o eixo p no intervalo $[0; t]$.

No gráfico 1, está plotada uma curva de Lorenz genérica $L(p)$. A linha AB , correspondente à função $L_0(p) = p$, representa a curva de Lorenz para uma situação de perfeita equidade distributiva, onde os 100 p% mais pobres da população contribuem com 100 p% da renda total. A área hachurada A_c , entre a curva $L(p)$ e a reta igualitária $L_0(p)$ é denominada área de concentração. Como $L(p)$ se afasta de $L_0(p)$ quando aumenta o grau de concentração, a área de concentração A_c também aumenta a partir de zero, tendendo à área do triângulo ABC ($=1/2$) quando $L(p)$ tende à linha quebrada ACB . Essa linha quebrada é o equivalente geométrico da máxima desigualdade, quando $N - 1$ membros da população não têm renda alguma, $L(p) = 0$ para $0 \leq p \leq (N - 1)/N$ e, portanto, p tende à unidade quando N tende ao infinito.

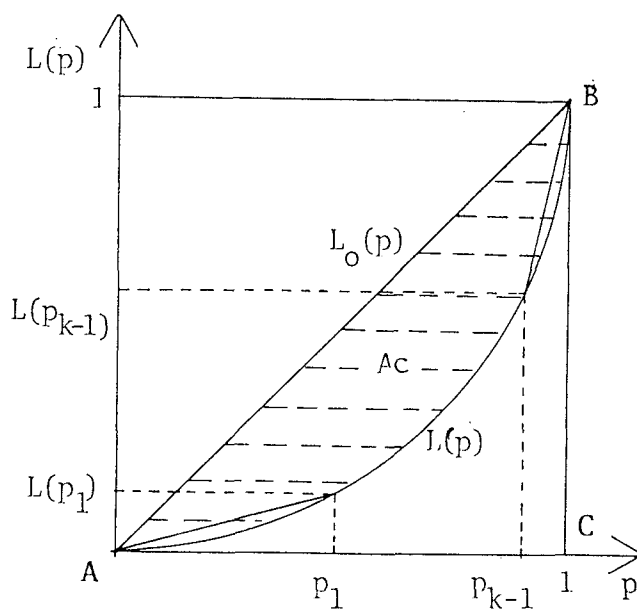


Gráfico 01: Uma curva de Lorenz genérica mostrando a linearização por partes.

A construção da curva de Lorenz via (09) exige a estimação da função de densidade $f(x)$.

A função $L(p)$ é obtida diretamente por

$$L(p) = \frac{1}{\mu} \int_0^p F^{-1}(t) dt \quad (10)$$

onde $F^{-1}(x)$ é a inversa da função de distribuição da variável X . A estimação de $f(x)$ ou de $F(x)$ envolve frequentemente procedimentos computacionais sofisticados que podem ser não condizentes com as informações a serem extraídas da curva de Lorenz. Para a maior parte dos estudos de processos distributivos, formas mais simples de obtenção da curva de Lorenz têm sido apresentadas.

O processo de linearização por partes da curva de Lorenz é bastante útil para o caso em que os dados agrupados são disponíveis. Conhecendo-se, para cada intervalo de classe $[a_{j-1}, a_j)$, a frequência relativa f_j e a renda total $R_j, j = 1, 2, \dots, K$, obtêm-se $K - 1$ pontos $(p_j, L(p_j))$ tais que

$$p_j = \sum_{i=1}^j f_i \quad \text{e} \quad L(p_j) = \sum_{i=1}^j R_i / R \quad (11)$$

onde $R = \sum_{j=1}^K R_j$ é a renda total da população.

Unindo-se os pontos $(p_{j-1}, L(p_{j-1}))$ e $(p_j, L(p_j))$ por segmentos de reta obtêm-se uma poligonal $(0;0), (p_1; L(p_1)), \dots$,

$(p_{K-1} ; L(p_{K-1}))$, $(1 ; 1)$ que é tangenciada nos vértices pela curva de Lorenz real (ver gráfico 1). Para efeito de análise gráfica, a curva pode ser interpolada sobre os vértices da poligonal e, para o cálculo da área de concentração, uma estimacão é dada pela área do polígono formado pelos pontos $(p_j, L(p_j))$.

Outro processo utilizado é a estimacão da curva de Lorenz sem que se faça uso da função de densidade $f(x)$. Essa idéia foi pouco abordada na literatura pelo simples fato de que ao postular-se uma forma funcional para a função $L(p)$, implicitamente determina-se uma forma funcional de $f(x)$ e vice-versa. Um artigo recente de Kakwani e Podder [23], através de uma abordagem nova do problema, propõe que a curva de Lorenz seja estimada num sistema de coordenadas rs obtido por rotacão dos eixos p e $L(p)$, de formas que

$$r = (p + L(p)) / \sqrt{2} \quad \text{e} \quad s = (p - L(p)) / \sqrt{2}$$

são, respectivamente, a distância do ponto $(p, L(p))$ à origem sobre a linha igualitária e a distância perpendicular desse ponto à linha igualitária. As primeiras experiências empíricas com a função $s = ar^b (\sqrt{2} - r)^c$ foram satisfatórias. No entanto, outras funções mais complexas têm sido apontadas como mais convenientes, não havendo ainda consenso quanto às formas funcionais a serem usadas.

A obtenção das medidas de concentração associadas à curva de Lorenz é geralmente simples e a sua interpretação geométrica é imediata. A compreensão total do significado des

sas medidas, entretanto, fica condicionada ao conhecimento das suas raízes estatísticas.

Nas seções seguintes, procura-se detalhar essas medidas em sua essência, bem como discutir sua aplicabilidade e limitações.

2.6.1. A razão de concentração de Gini

A área de concentração A_c , compreendida entre a curva de Lorenz $L(p)$ e a reta igualitária $L_0(p) = p$, tem seu valor definido por

$$A_c = \frac{1}{2} - \int_0^1 L(p) dp$$

A medida de concentração mais importante entre as associadas à curva de Lorenz é a razão de concentração de Gini, definida como a razão entre a área de concentração real A_c e a área de concentração máxima $1/2$

$$G = A_c / (1/2) = 2A_c = 1 - 2 \int_0^1 L(p) dp$$

O índice de Gini pode ser obtido diretamente da função de distribuição $F(x)$, observando a equação (10), por

$$G = 1 - 2 \int_0^1 \left(\frac{1}{\mu} \int_0^p F^{-1}(t) dt \right) dp$$

Visto que a área de concentração varia no intervalo $[0 ; 1/2]$, a variação do índice de Gini se dá no intervalo $[0 ; 1]$, com os limites inferior e superior avaliando as situações extremas de absoluta equidade distributiva e completa de sigualdade, respectivamente.

Através dessa formulação geométrica, a razão de concentração de Gini pode ser entendida facilmente. Entretanto, uma interpretação estatística da medida pode revelar bem melhor sua essência.

Pode ser mostrado¹⁵ que a razão de concentração de Gini, em termos estatísticos, é dado por

$$G = \Delta / 2\mu$$

onde μ é a média da variável renda e Δ

$$\Delta = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |x - y| dF(y) dF(x)$$

ou

$$\Delta = \frac{1}{N(N-1)} \sum_{i=1}^N \sum_{j=1}^N f_i f_j |x_i - x_j| \quad i \neq j$$

é a diferença média da variável x com funções de distribuição $F(x)$ para o caso contínuo, ou função frequência absoluta $f(x_i)=f_i$, para o caso discreto.

¹⁵ Kendall e Stuart (1989), pag. 48.

A diferença média Δ é uma medida de dispersão absoluta com valor mínimo zero e valor máximo 2μ . Logo, nessa formulação, o Gini pode ser entendido como uma medida de dispersão relativa das rendas individuais e, como tal, incorpora as vantagens de independência da escala adotada e tornar viável com comparações entre populações diversas, com a vantagem adicional de considerar desvios entre todos os valores da variável e não apenas desvios com relação à média. Para o caso da renda, o índice de Gini tem a desvantagem de penalizar diferenças absolutas de renda.

Um outro problema da razão de concentração, que pode ser visualizado no gráfico 2, é o fato de, por não incorporar informações acerca da assimetria ("skewness") da distribuição, fornecer, eventualmente, o mesmo valor para distribuições diferentes. É o caso de duas curvas de Lorenz que se interceptam, fornecendo áreas de concentração iguais. Esse problema pode ser resolvido, calculando-se, além do Gini para a população total, índices parciais para os elementos que percebem abaixo e acima da média.

Mesmo sem o uso desse artifício, o índice de Gini tem sido considerado como a medida simples de desigualdade mais eficiente¹⁶, embora reconheça-se nele um certo grau de insensibilidade a variações pequenas na distribuição. A aparente dificuldade de computação que ressalta das formas já apresentadas no texto para o seu cálculo, é contornável desde que se lance mão de um procedimento veiculado como o método padrão para

¹⁶ Ver Morgan (1962)

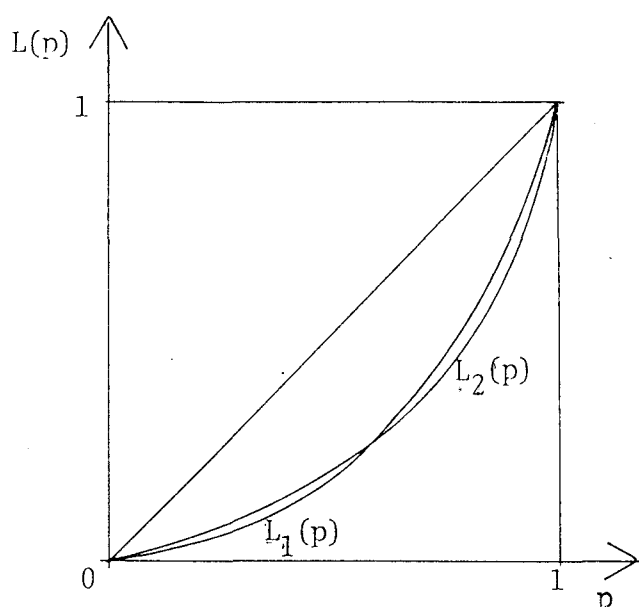


Gráfico 02: Curvas de Lorenz que se intercentam fornecendo a mesma área de concentração.

sua estimação¹⁷, já citado na seção anterior. Trata-se de obter u ma estimativa da área de concentração A_c através da linearização por partes da curva $L(p)$. Esse procedimento sempre conduz a uma subestimação da área de concentração e, por extensão, do índice G , face à convexidade da curva de Lorenz. Na verdade, o valor de G obtido via área da poligonal no gráfico 1, refere-se apenas às desigualdades inter-estratos, pois a suposição implícita à linearização da curva entre os pontos $(p_{j-1}, L(p_{j-1}))$ e $(p_j, L(p_j))$ é a de que a renda total R_j de cada classe é equitativamente distribuída pelos elementos incluídos na j -ésima classe. Portanto, tra-

¹⁷ Ver Morgan (1962)

ta-se nitidamente da obtenção de um limite inferior para a razão de concentração.

O efeito das desigualdades intra-estratos pode ser analisado decompondo-se a diferença média Δ por

$$\Delta = \sum_{i \neq j} \sum f_i f_j |\mu_i - \mu_j| + \sum_{j=1}^K f_j^2 \Delta_j$$

onde f_j é a frequência relativa da classe j ($f_j = p_j - p_{j-1}$), μ_j é a renda média dos indivíduos incluídos na classe j e Δ_j é a diferença média da classe j . Gastwirth mostrou que

$$0 \leq \Delta_j \leq 2(\mu_j - a_{j-1})(a_j - \mu_j) / (a_j - a_{j-1})$$

ou

$$0 \leq \Delta_j \leq 2(\mu_j - a_{j-1})$$

quando a_j não é finito.

Com base nesse resultado, pode-se derivar um limite superior e, portanto, um intervalo de variação para o Gini, chamando de G_1 o limite inferior obtido pelo método Standard, tem-se

$$G_1 \leq G \leq G_1 + \bar{D} = G_s$$

onde

$$\bar{D} = \frac{1}{\mu} \left(\sum_{j=1}^K f_j^2 (\mu_j - a_{j-1})(a_j - \mu_j)(a_j - a_{j-1}) \right)$$

e a K-ésima parcela do somatório deve ser substituída por $2(\mu_K - a_{K-1})$ quando a última classe é aberta.

Os resultados obtidos por Gastwirth trouxeram uma contribuição muito valiosa à utilização do índice de Gini em estudos de desigualdade de renda, principalmente pelo fato de mostrar que, mesmo sem ajustar funções de distribuição aos dados grupados, pode-se obter informações bem precisas sobre o valor real do índice de Gini.

Não obstante ser o Gini uma medida de desigualdade muito usada, é a curva de Lorenz que fornece ao pesquisador a informação completa sobre a estrutura da distribuição. Assim sendo, outras medidas podem ser definidas sobre a curva de Lorenz e o conjunto dessas medidas, de naturezas diversas, podem definir o perfil da curva, trazendo mais informações que o Gini isoladamente. Por isso, na seção seguinte, outras medidas de desigualdade que podem ser formuladas usando a curva de Lorenz são apresentadas e discutidas.

2.6.2. Outras medidas de desigualdade e a curva de Lorenz

A característica mais marcante da curva de Lorenz é a sua convexidade. Sendo $L(p)$ convexo, a função $d(p) = p - L(p)$ é côncava sendo nula em $p = 0$ ou $p = 1$. Dessa forma, existe um ponto p_0 tal que $p - L(p)$ é máximo. O valor de p_0 pode ser obtido fazendo nula a primeira derivada de $d(p)$ com respeito

a p. Assim

$$\frac{d}{dp} (p - L(p)) = 0$$

$$1 - \frac{d}{dp} \int_0^p \frac{1}{\mu} F^{-1}(t) dt = 0$$

$$1 - \frac{1}{\mu} F^{-1}(p_0) = 0$$

$$p_0 = F(\mu)$$

O valor da máxima discrepância $d(p_0)$ pode ser obtido fazendo

$$d(p_0) = p_0 - L(p_0) = F(\mu) - L(\mu)$$

$$d(p_0) = \int_0^\mu dF(x) - \int_0^\mu \frac{x}{\mu} dF(x)$$

$$d(p_0) = \int_0^\mu (1 - x/\mu) dF(x)$$

$$d(p_0) = \int_0^\mu \frac{(\mu - x)}{\mu} dF(x)$$

$$d(p_0) = \frac{1}{\mu} \frac{\delta}{2}$$

$$\text{onde } \delta = \int_0^{\infty} |x - \mu| dF(x) = 2 \int_0^{\mu} (\mu - x) dF(x)$$

é uma medida de dispersão conhecida como desvio médio, cujo limite superior é 2μ . Assim, o valor da máxima discrepância entre a reta igualitária e a curva de Lorenz é uma medida de dispersão relativa e, portanto, avalia de certa forma o grau de desigualdade existente.

Também reduzem-se a $\delta/2\mu$ outras medidas de desigualdade propostas sobre a curva de Lorenz como a de Schutz (1951) que mede a desigualdade pela integral

$$S = \int_0^{p_0} [1 - L'(p)] dp$$

$$S = p_0 - L(p_0) = d(p_0) = \delta/2\mu$$

e a de Pietra nos anos 30 que sugeriu como medida de concentração a razão entre a área do maior triângulo inscrito na curva de Lorenz e a área sob a reta igualitária. Como as bases dos dois triângulos são as mesmas, a medida P de Pietra se reduz à razão entre a maior distância perpendicular h_0 da curva de Lorenz à reta $L_0(p) = p$ e a metade da diagonal $(1, 0)(0, 1)$. Como $h(p) = d(p)/\sqrt{2}$, tem-se $h_0 = d(p_0)/2 = \delta/2\sqrt{2}\mu$ e

$$P = \frac{2h_0}{\sqrt{2}} = \frac{\delta}{2\mu}$$

O ponto $(p_0, L(p_0))$ está também intimamente ligado ao sistema de três medidas de desigualdade propostos por Éltetö e Frigyes [12]. As medidas são

$$u = m/m_1, \quad v = m_2/m_1, \quad w = m_2/m, \quad (12)$$

onde $m = E(x)$, $m_1 = E(x/x < m)$ e $m_2 = E(x/x \geq m)$. As medidas u , v e w têm uma interpretação econômica simples pois relacionam a média de renda da população e as médias dos elementos da população com renda acima ou abaixo da média. O intervalo de variação das três medidas é $[1, \infty)$, sendo ambas crescentes com o grau de concentração.

A ligação do sistema de Éltetö e Frigyes com a curva de Lorenz vem das relações abaixo:

$$u = \frac{m}{m_1} = \frac{F(m)}{m^{-1} \int_0^m x dF(x)} = \frac{p_0}{L(p_0)}$$

$$w = \frac{m_2}{m} = \frac{m^{-1} \int_m^\infty x dF(x)}{1 - F(m)} = \frac{1 - L(p_0)}{1 - p_0}$$

$$v = u, w = \frac{p_0}{L(p_0)} \cdot \frac{1 - L(p_0)}{1 - p_0}$$

A interpretação geométrica para essas medidas, que pode ser visualizada no gráfico 3, é obtida unindo-se os pontos $(0, 0)$, $(p_0, L(p_0))$ e $(1, 1)$ da curva de Lorenz por segmentos de reta. Os ângulos formados por esses dois segmentos com a horizontal, respectivamente α e β , são dados por

$$\cotg \alpha = u \quad \text{e} \quad \tg \beta = w$$

enquanto $v = uw = \tg \beta / \tg \alpha$.

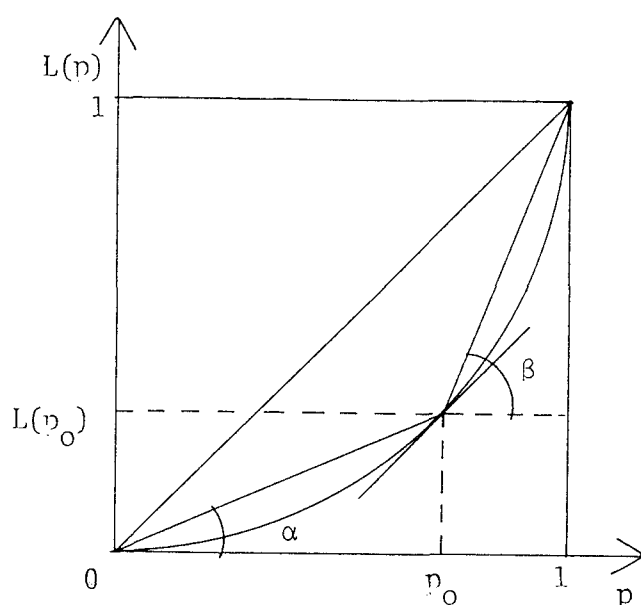


Gráfico 03: Uma curva de Lorenz genérica destacando o ponto de máxima discrepância p_0 .

O sistema (12) é independente em apenas duas medidas. Kondor [25] mostrou que essas duas medidas não trazem muitas informações além das obtidas com o desvio médio relativo δ/μ . Não obstante, a pouca informação adicional é importante pois reflete a assimetria ("skewness") da curva de Lorenz.

É interessante também notar que os segmentos que dão origem a α e β são os lados desconhecidos do triângulo de Pietra, o que indica um ciclo em torno da medida $\delta/2\mu$. Com efeito, em combinação com o índice de Gini, o desvio médio relativo completa a idéia de dispersão relativa dos dados, pertinente à curva de Lorenz.

2.7. Considerações Gerais

As medidas de concentração e desigualdade avaliam os processos distributivos quantificando determinada (s) característica(s) da população. A menos que o propósito do trabalho não exija um perfil mais detalhado da situação distributiva, a caracterização desta, a rigor, não é tarefa para uma medida isolada. Em sendo o grau de concentração o que se pretende medir, cada um dos índices existentes o avalia segundo uma definição do que seja desigualdade. A essência de cada índice e, explicitamente, a sua formulação são moldadas por um conceito de desigualdade a ele subjacente e, por esse motivo, a configuração completa do grau de concentração, em uma determinada situação econômica só é possível com um conjunto de medidas.

A curva de Lorenz, por compactar medidas importantes de desigualdade, pode exercer o papel de sistema de medidas. Sua obtenção diretamente a partir dos dados é uma propos-

ta recente e não se encontra suficientemente desenvolvida. A alternativa é a construção da curva de Lorenz indiretamente, a partir de funções de distribuição que adiram às observações e sejam, por si só, instrumentos auxiliares de análise.

No capítulo seguinte, as funções mais utilizadas para explicar dados de distribuição de renda são estudadas e apresentados os reflexos da aceitação dessas funções sobre a curva de Lorenz e as medidas de desigualdade aqui consideradas,

3. MODELOS DE DISTRIBUIÇÃO DE RENDA

3.1. Generalidades

O desenvolvimento e aplicação de modelos de natureza probabilística em distribuição de renda é uma idéia antiga. Já em fins do século passado, Pareto propunha a sua "lei" distributiva para explicar as diferenças de renda em diversos países e períodos. A "lei" de Pareto, formulada em termos de frequência absoluta dos membros da população, conduz, implicitamente, como se verá adiante, a uma função de distribuição.

Contribuições a essa abordagem da distribuição de renda, entretanto, são esparsas, mesmo que significativas, na primeira metade deste século. A rigor, apenas Amoroso (1925) e Gibrat (1931), respectivamente com a função gama generalizada e a distribuição log-normal, destacam-se entre as referências. Nas décadas de 50 e 60, surgiram os primeiros esforços para derivação de funções a partir de suposições consistentes sobre o processo de geração ou distribuição da renda. Os trabalhos de Champernowne (1953) e Fisk (1961) são os exemplos mais representativos.

É na década de 70, entretanto, que surgem as maiores contribuições à análise da renda via funções de distribuição. A distribuição log-normal deslocada (Metcalf, 1969), a densidade beta (Thurow, 1970), a densidade gama a dois parâmetros (Salem e Mount, 1974), a distribuição de Weibull (Bartels e Van Metelen, 1975), o modelo log-logístico (Dagum, 1977 e Sabóia, 1978) e

uma função de Burr (Singh e Maddala, 1976) foram aplicadas e, embora não se tenha chegado a uma definição de qual a função de distribuição que mais bem se ajusta a dados de renda, tem-se, pelo menos, a certeza de que algumas funções destacam-se dentre as outras, seja pelo bom ajustamento aos dados que propiciam, seja pela facilidade de estimação e interpretação dos seus parâmetros.

Nas seções seguintes, as propriedades dessas funções e suas características principais, enquanto modelos de distribuição de renda, são analisadas e verificadas as suas implicações sobre as medidas de desigualdade apresentadas no capítulo anterior.

3.2. A Função de Pareto

A "lei" de Pareto pode ser representada pela equação

$$N(r) = \beta r^{-\alpha} \quad \beta > 0, \quad \alpha > 1, \quad r \geq A > 0$$

onde $N(r)$ representa o número de elementos da população com renda igual ou superior a r e A é a renda mínima, considerada estritamente positiva. O número total de elementos da população equivale a $N(A)$ e, logo, a função de distribuição $F(r)$ implícita à lei de Pareto é

$$F(r) = \frac{N(A) - N(r)}{N(A)} = 1 - \left(\frac{A}{r}\right)^\alpha \quad (13)$$

e a função densidade de probabilidade é dada por:

$$f(r) = \frac{\alpha}{A} r^{-(\alpha+1)} \quad r \geq A > 0; \quad \alpha > 1$$

A distribuição de Pareto tem média $E(R) = \frac{\alpha A}{\alpha - 1}$ e variância $V(R) = \frac{\alpha A^2}{(\alpha - 2)}$, para $\alpha > 2$. Logo, o coeficiente de variação é $V = \sigma / \mu = \frac{\alpha - 1}{\sqrt{\alpha(\alpha - 2)}}$.

A curva de Lorenz obtida sob a suposição que a distribuição de renda segue uma função de Pareto é

$$L(p) = 1 - (1 - p)^k$$

onde $k = 1 - (1/\alpha)$ e a razão de concentração de Gini é dada por $G = 1/(2\alpha - 1)$. O valor da máxima discrepância entre a curva de Lorenz e a reta igualitária, equivalente ao desvio médio relativo, é $d_0 = \delta / 2\mu = 1 - k^{\alpha-1}$. A variância da variável $L = \log R$ é $V(L) = 1/\alpha^2$ e o índice de Theil para distribuições de renda seguindo a função de Pareto é dada por

$$T(R, \alpha) = \frac{1}{\alpha - 1} - \log \frac{\alpha}{\alpha - 1}$$

Como se vê, os índices de desigualdade calculados sob a suposição de que a distribuição de renda segue a "lei" de Pareto, dependem unicamente de α e, à medida que esse parâmetro cresce, diminui o grau de concentração na forma medida por todos os índices considerados. Portanto, α pode ser entendido como um parâmetro de igualdade, sendo esta tanto maior quanto mais α se distancia da unidade.

A estimação da distribuição de Pareto segue um processo bastante simples, pois é possível uma linearização de (13) para

$$\ln(1 - F(r)) = \alpha \ln A - \alpha \ln r \quad (14)$$

de formas que pode-se obter as estimativas de α e A por mínimos quadrados ordinários. Em (14), disponíveis as observações censitárias em classes de renda, associa-se, ao limite superior r_j ($j=1,2,\dots,m$) de cada classe, a percentagem $F(r_j)$ da população com renda inferior a r_j . Como a última classe é geralmente aberta, ou seja, r_m não é finito, dispõe-se de $m-1$ observações para que a estimação seja realizada.

A maior desvantagem da função de Pareto na descrição de dados de renda é que sua densidade é zero-modal, enquanto, normalmente, os histogramas de processos distributivos são uni-modais. Assim, o ajuste da função não é satisfatório à esquerda do valor modal da distribuição real, embora tenha um desempenho razoável nas faixas mais altas. Por esse motivo, a função de Pareto tem sido usada mais recentemente em distribuições truncadas à

esquerda (com uma origem A arbitrária), apenas para fornecer uma estimativa da média de renda na última classe (aberta) quando não se dispõe da renda total em cada classe¹.

A aplicação da função de Pareto a estudos de concentração industrial é possível desde que considere-se apenas as firmas maiores (segundo a variável usada no estudo), para aproveitar a característica de boa aderência da função nas faixas mais altas do atributo considerado.

3.3. A Distribuição Log-Normal

A aplicação da distribuição log-normal a desigualdades econômicas remonta à década de 30. Um dos motivos principais para o seu uso é que essa distribuição surge naturalmente, quando a variável estudada pode ser interpretada como um produto de variáveis. A variável renda "pode ser interpretada como resultante de uma renda inicial que sofre o efeito de taxas de crescimento a cada unidade de tempo" (Sabóia, 1978). O mesmo acontece com outras variáveis econômicas como o faturamento de firmas ou o seu capital. Com efeito, a distribuição log-normal tem sido também usada em estudos de concentração industrial².

Uma variável aleatória X é dita log-normal se a variável $Y = \ln(X - \tau)$ é normalmente distribuída com parâmetros μ e σ . A função densidade da variável X é

¹ Neste caso, para os outros intervalos, usa-se o ponto médio como ponto representativo da classe.

² Ver Sabóia (1977).

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}(x-\tau)} \exp \left\{ -\frac{1}{2\sigma^2} [\ln(x-\tau) - \mu]^2 \right\} \quad x > \tau$$

e a função de distribuição correspondente $F(x)$, não sendo explicitável, é designada por $\Lambda(x-\tau|\mu,\sigma)$.

A distribuição log-normal a dois parâmetros é mais usada, correspondendo à situação em que o parâmetro de inicialização³ τ é nulo. Esta é a situação mais comum quando a variável analisada é a renda pessoal ou familiar, embora alguns pesquisadores trabalhem com rendas negativas. Em estudos de concentração industrial, é interessante, às vezes, considerar apenas as unidades produtoras com valor-atributo superior a um mínimo arbitrariamente escolhido. Ainda assim, não é necessária a utilização da log-normal a três parâmetros, vez que basta uma simples mudança de variável $X_1 = X - \tau$ para que se tenha a distribuição a dois parâmetros.

A média e a variância de uma variável log-normal a dois parâmetros são dadas por

$$E(x) = \exp\left(\mu + \frac{\sigma^2}{2}\right)$$

$$V(x) = e^{2\mu + \sigma^2} (e^{\sigma^2} - 1)$$

³ Threshold, em inglês. Refere-se ao fato de, à esquerda de $X = \tau$, o valor da função densidade é nulo.

O coeficiente de variação $V = \sigma/\mu$ é, portanto,

$$V = (e^{\sigma^2} - 1)^{1/2}$$

A curva de Lorenz obtida sob a suposição de log-normalidade da variável renda é simétrica em relação à diagonal secundária do quadrado ACBD no gráfico 1 e a razão de concentração de Gini é

$$G = 2N(\sigma/\sqrt{2} | 0,1) - 1$$

onde $N(z|\mu, \sigma^2)$ denota a probabilidade de que uma variável normalmente distribuída, com média μ e variância σ^2 , assumira um valor menor que z .

O desvio médio relativo, equivalente ao valor da máxima diferença entre a curva de Lorenz e a reta igualitária é $d_0 = \delta/2\mu = N(\sigma/2|0,1) - N(-\sigma/2|0,1)$.

Como era de se esperar, as medidas de desigualdade extraídas da distribuição log-normal são funções apenas do parâmetro σ , identificado como medida de dispersão. O parâmetro μ pode ser interpretado como uma medida de localização pois sua influência é exercida apenas na definição da posição da curva em relação ao eixo horizontal.

A estimação dos parâmetros μ e σ da distri-

buição log-normal que mais bem se ajusta aos dados observados pode ser realizada através de várias técnicas de estimação, como o método de máxima verossimilhança, o método de Pearson (estimadores de mínimos qui-quadrados) e o método de mínimos quadrados ordinários. Os dois últimos no caso da função log-normal e de outras funções de distribuição não-linearizáveis, conduzem a estimações não-lineares, o que exige programas de computador nem sempre disponíveis. Por outro lado, a maximização da função de verossimilhança nem sempre é possível em termos analíticos, provocando soluções através de procedimentos iterativos específicos.

Um procedimento mais simples para a estimação dos parâmetros de uma distribuição log-normal, que pode ser denominado de método dos "fractiles", envolve apenas a aplicação do método da regressão linear simples. Consiste do seguinte: conhecido o histograma da variável renda pode-se obter, tomando-se os logaritmos dos limites das classes de renda, o histograma da variável $\log R$. Sendo $\log R$ normalmente distribuída com média μ e variância σ^2 , a variável Z definida como

$$Z = \frac{\log R - \mu}{\sigma}$$

segue uma distribuição normal "standard". Então, se p_j é a percentagem da população com renda menor que a_j ($j = 1, 2, \dots, m$), sendo a_j o limite superior da j -ésima classe, pode-se extrair de uma tabela dos quantis da distribuição normal padrão o valor z_j para o qual

$$P(Z < z_j) = p_j$$

De acordo com a padronização da variável $\log R$, pode-se escrever, para $j = 1, 2, \dots, m-1$ (em geral, a_m não é finito), a relação

$$\log a_j = \hat{\mu} + \hat{\sigma} z_j + \varepsilon_j$$

onde ε_j é o erro aleatório sob as suposições do modelo básico de regressão linear simples⁴.

Seja qual for o método de estimação utilizado, a distribuição log-normal, em geral, provê um bom ajustamento aos dados. Estudos recentes têm mostrado que outras funções de distribuição dão uma melhor aderência que a log-normal pelo menos para os dados de renda usados nestes trabalhos⁵. Apesar disso, não é possível desmerecer a capacidade da distribuição log-normal em explicar as diferenças de renda. Além disso, a facilidade de interpretação de μ e σ por um método simples e rápido como o método dos "fractiles" favorecem a aplicação dessa distribuição não só em estudos de concentração de renda pessoal ou familiar mas também em distribuições de outros atributos.

⁴ As suposições do modelo básico são resíduos não-auto-correlacionados e homocedásticos.

⁵ McDonald e Ransom (1979), p.ex.

3.4. As Densidades Gama e Beta

A densidade Gama foi proposta como modelo de distribuição de renda por Salem e Mount (1974), particularizando um estudo pioneiro de Amoroso (1925). A função

$$f(x) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)} \quad x > 0 \text{ e } \alpha, \beta > 0 \quad (15)$$

é a tipo III do sistema de Pearson.

A média da distribuição gama e a sua variância, são dadas por $E(X) = \alpha/\beta$ e $V(X) = \alpha/\beta^2$, de formas que o coeficiente de variação de Pearson é dado por $V = \sigma/\mu = \alpha^{-1/2}$.

Por outro lado, os valores do coeficiente de Gini e do índice de Theil podem ser evoluídos com relativa facilidade, obtendo-se, respectivamente,

$$G = \frac{\Gamma(\alpha + 1/2)}{\Gamma(\alpha + 1)\sqrt{\pi}}$$

$$T = 1/\alpha - \ln \alpha + \psi(\alpha)$$

onde a função $\psi(\alpha) = \frac{d \ln \Gamma(\alpha)}{d\alpha}$ é designada como a função "digamma".

Como se vê nas expressões acima, o parâmetro β é apenas um fator de escala, não interferindo no valor dos índices de desigualdade. Mesmo assim, a distribuição gama tem relativa flexibilidade podendo assumir as formas "U" invertido ou "J" (unimodal ou zero-modal). A função (15) é zero-modal para $(0 < \alpha < 1)$

A interpretação dos parâmetros da densidade gama é imediata vez que, sendo β um fator de escala, α é o parâmetro único de concentração, similarmente aos parâmetros μ e σ da distribuição log-normal.

A distribuição beta, considerada por Thurow (1970) como modelo de distribuição de renda, é dada por⁶

$$f(x) = \frac{x^{r-1} (t-x)^{s-1}}{t^{r+s-1} B(r,s)} \quad 0 < x < t \quad (16)$$

com $r, s > 0$, onde $B(r, s) = \frac{\Gamma(r) \Gamma(s)}{\Gamma(r+s)}$ é a função beta de Euler.

Na forma (16), a distribuição beta é tripamétrica o que, além de maior flexibilidade para a descrição de dados, introduz uma inovação, qual seja, a de considerar o truncamento da variável renda. O modelo, portanto, incorpora o fato de que existe uma renda máxima finita na população.

A função (16) é zero-modal para $r > 1$ e $s < 1$ ou vice-versa e unimodal para $r, s > 1$ (ou < 1). Quando $r > s > 1$,

⁶ A função beta mais divulgada na literatura é a bipamétrica, que equivale a (16) com $t = 1$.

o gráfico da função é inclinado para a frente, na direção do ponto $x = t$, sendo inclinado para a origem quando $s > r > 1$.

A média e o coeficiente de Gini para a distribuição beta são dados por

$$E(x) = tr/(r+s)$$

$$e \quad G = \frac{\Gamma(r+s) \Gamma(r+1/2) \Gamma(s+1/2)}{\Gamma(r+s+1/2) \Gamma(r+1) \Gamma(s) \sqrt{\pi}}$$

A função beta tem a peculiaridade de, quando $s \rightarrow \infty$ mantendo-se constante a média, tender em forma à distribuição gama. Esse fato leva à conclusão de que a aderência da função beta é, pelo menos, idêntica à propiciada pela função gama.

Não se tem conhecimento de aplicação de qualquer dessas funções a dados relativos a concentração industrial. Não há, entretanto, motivos que levem a descrer das possibilidades dessas funções. Dado a sua flexibilidade, como também aos bons resultados obtidos com as aplicações a dados de renda pessoal ou familiar é bastante provável que as distribuições beta e gama tenham desempenho satisfatório na descrição de observações da estrutura industrial.

3.5. O Modelo de Singh-Maddala

Em um trabalho recente, Singh e Maddala desenvolveram uma função para a distribuição de renda a partir do conceito de "taxa de falha" ("hazard rate" ou "failure rate"), aplicado em teoria da confiabilidade e modelos de manutenção e substituição (ver Barlow e Proschan, 1965 e Rødder, 1979).

Se o tempo que um equipamento, em operação contínua, permanecerá em funcionamento é uma variável aleatória, com função densidade de probabilidade $f(x)$, a probabilidade que ele se encontre em funcionamento no tempo x é $R(x) = 1 - \int_0^x f(x) dx = 1 - F(x)$. Dado que o equipamento "sobreviveu" até o instante x , a probabilidade de falha entre x e $x+dx$ é $s(x) = f(x)dx/R(x)$ e a "taxa de falha" instantânea, no instante x é dada por

$$r(x) = \frac{d}{dx} \left[\int_0^x f(x) dx / R(x) \right] = f(x) / R(x)$$

Com relação à variável renda, $r(x)$ pode ser interpretado como avaliando a tendência de uma renda x gerar rendas maiores. Enquanto que, no caso da vida do equipamento, a função $r(x)$ deve ser crescente com x , vez que, quanto maior o tempo de funcionamento, maior a tendência à falha imediata, o caso da renda envolve outras considerações.

A suposição usual é que, para a faixa inferior da renda, $r(x)$ deve ser crescente, seguindo-se uma seção com "taxa de falha" decrescente. A justificativa envolve a consideração de que, abaixo de um determinado valor, rendas crescentes não habilitam o agente econômico (indivíduo) a faturar mais. Acima

desse valor, entretanto, existe essa possibilidade. Uma interpretação equivalente a esta, envolvendo a idéia de ascensão (mobilidade) social: considerando o valor onde $r'(x) = 0$ como delimitador de classes, vê-se que $r(x)$ crescente na classe baixa indica que a possibilidade de ascensão dos indivíduos da classe baixa decresce com a renda, enquanto que $r(x)$ decrescente na classe alta indica que a possibilidade de ascensão nesta classe cresce com a renda. O menor nível de mobilidade encontra-se no delimitador de classes e, portanto, a ascensão é mais difícil onde ela significa mudança de classe.

Singh e Maddala utilizaram, na evolução de seu modelo, um conceito alternativo qual seja o de "taxa de falha proporcional". Esse conceito se traduz, matematicamente, por uma função $r^*(z)$, onde $z = \log x$ e

$$r^*(z) = \frac{dF(z)}{dz} \frac{1}{1-F(z)}$$

Sob a suposição de que $r^*(z)$ cresce com z a princípio a uma taxa crescente, depois a uma taxa decrescente até tornar-se assintoticamente constante, Singh e Maddala propuseram a equação diferencial

$$y'' = ay'(\alpha - y''). \quad (17)$$

onde $y = \log(1 - F)^{-1}$, $y' = r^*(z)$ e $y'' = r^{*'}(z)$

A função de distribuição obtida como solução de (17) é

$$F(x) = 1 - \frac{1}{(1 + a_1 x^{a_2})^{a_3}} \quad x \geq 0, \quad a_1, a_2, a_3 > 0 \quad (18)$$

que é uma das famílias de distribuição propostas por Burr em 1942, embora com outros propósitos (ver Kendall e Stuart, 1969).

As expressões para a média e a razão de concentração de Gini (Cronin, 1979) são, respectivamente,

$$E(x) = \frac{\Gamma(1/a_2 + 1) \Gamma(a_3 - 1/a_2)}{a_1^{1/a_2} \Gamma(a_3)}$$

e

$$G = 1 - \frac{\Gamma(2a_3 - 1/a_2) \Gamma(a_3)}{\Gamma(2a_3) \Gamma(a_3 - 1/a_2)}$$

Uma das maiores virtudes da função (18) é que ela inclui, como casos particulares, as distribuições de Weibull, Pareto e a função Sech^2 considerada por Fisk (1961) (ver referência [14]). Assim, dependendo dos valores dos parâmetros, a função densidade associada a (18) pode ser unimodal ou zero-modal.

A estimação dos parâmetros pode ser feita utilizando-se métodos de estimação convenientes, mas o mais indicado é o método dos mínimos quadrados não-linear.

Apesar de não haver sido ainda aplicado em estudos de concentração industrial, o modelo de Singh-Maddala é obtido de suposições sobre o processo de distribuição de renda que são válidas para o caso de outros atributos. A excelente aderência da função a dados de renda pessoal e familiar, bem como a flexibilidade do modelo dado que envolve três parâmetros, são indicadores positivos para uma boa performance da função em outras aplicações.

Uma desvantagem do modelo é a dificuldade de interpretação dos parâmetros. Excetuando o parâmetro a_1 , nitidamente um fator de escala, os demais não têm significado explícito como é comum em outras funções aplicadas à distribuição de renda.

3.6. O Modelo Log-Logístico

As observações empíricas da distribuição da variável renda revelam algumas características regulares em regiões e períodos de tempo diferentes. A unimodalidade da função densidade e a existência de uma percentagem finita de agentes econômicos com renda nula são algumas destas características. A elas, Dagum (1977) acrescentou a suposição de que a elasticidade-ingresso da distribuição $F(x)$ é uma função monotonicamente decrescente e linear de $F(x)$, a partir da origem $\alpha > 0$.

Para esse conjunto de características, a especificação matemática conveniente é a equação diferencial

$$\frac{x}{F(x) - \alpha} \frac{dF(x)}{dx} = \frac{\delta}{1 - \alpha} (1 - F(x))$$

$$\begin{array}{l} x > 0 \\ 0 \leq \alpha < 1 \\ \delta > 1 \end{array}$$

cujas soluções são as funções

$$F(x) = \alpha + \frac{1 - \alpha}{1 + \lambda x^{-\delta}} \quad \lambda > 0 \quad (19)$$

A função densidade de probabilidade é dada por

$$f(x) = \begin{cases} \alpha & x = 0 \\ (1 - \alpha) \frac{\lambda \delta x^{-(\delta+1)}}{(1 + \lambda x^{-\delta})^2} & x > 0 \end{cases}$$

A esta altura, é interessante notar que para $0 < \delta \leq 1$ a função $f(x)$ é zero-modal e, portanto, é útil para descrever densidades com forma "Ω" ou "J", indistintamente.

A especificação da função (19) para explicar dados de renda foi conseguida, independentemente de Dagum e de forma bem diversa, por Sabóia (1978). A linha de raciocínio de Sabóia parte do reconhecimento de que a distribuição log-normal é um bom modelo para distribuição de renda.

A distribuição logística

$$F(y) = \frac{1}{1 + e^{-z}}$$

onde z é uma função linear de y com $\frac{dz}{dy} > 0$ foi postulada por Berkson (ver Van der Waarden, 1967) como excelente aproximação para a função de distribuição de uma variável aleatória normal Y . Uma simples substituição de variáveis $Y = \log X$ conduz ao modelo log-logístico.

$$F(x) = \frac{1}{1 + e^{-a_1 - b \log x}} \quad x > 0, \quad b > 0$$

do qual, pode-se obter, fazendo $e^{-a_1} = a$

$$F(x) = \frac{1}{1 + a e^{-b \log x}} \quad a, b > 0, \quad x > 0 \quad (20)$$

A curva de Lorenz para o modelo log-logístico (19) foi obtida por Dagum, para $\delta > 1$, como

$$L(p) = \begin{cases} 0 & p \leq \alpha \\ \frac{\delta}{\pi} \operatorname{sen} \frac{\pi}{\delta} B\left(\frac{p-\alpha}{1-\alpha}; 1 + \frac{1}{\delta}, 1 - \frac{1}{\delta}\right) & p > \alpha \end{cases}$$

onde $B(t; r, s)$ é a função beta incompleta de Euler⁷. A média e a variância da distribuição proposta por Dagum são dadas por

$$E(x) = \frac{\pi(1-\alpha)\lambda^{1/\delta}}{\delta \operatorname{sen} \frac{\pi}{\delta}} \quad \delta > 1$$

$$V(x) = \lambda^{1/\delta} E(x) \left[\frac{1}{\cos(\pi/\delta)} - \frac{\pi(1-\alpha)}{\delta \operatorname{sen}(\pi/\delta)} \right] \quad \delta > 2$$

A razão de concentração de Gini, obtida a partir da curva de Lorenz, sob a suposição de que a distribuição de renda é log-logística, é dado por

$$G = \alpha + \frac{1-\alpha}{\delta} \quad \delta > 1$$

Como se vê, o parâmetro λ é um parâmetro de escala, não influenciando na avaliação do grau de concentração. O parâmetro α pode ser entendido como um parâmetro de desigualdade na medida em que contribue positivamente para o incremento do Gini. O parâmetro δ , inversamente, pode ser interpretado como um avaliador do grau de igualdade entre os membros da população com

⁷ A função beta incompleta é definida pela integral $\int_0^t t^{r-1}(1-t)^{s-1} dx$ enquanto a função beta (completa) é definida pela mesma expressão com $t=1$.

com renda não-nula.

O modelo proposto por Sabôia desconsidera, como é convencional na abordagem probabilística da distribuição de renda, a existência de indivíduos com renda nula. Portanto, equivale ao modelo de Dagum com α suposto nulo⁸. A média, variância, curva de Lorenz e a razão de concentração correspondentes à função (19) são obtidas das expressões já apresentadas, simplesmente fazendo $\alpha = 0$.

A estimação das funções apresentadas nesta seção é bastante simples e pode ser feita utilizando o método de mínimos quadrados para regressão linear. A partir da função de distribuição (20) obtém-se a forma linearizada

$$\log \frac{1 - F(x)}{F(x) - \alpha} = \log \lambda - \delta \log x$$

No caso do modelo tri-paramétrico, pode-se aplicar um método iterativo de mínimos quadrados (ver Dagum, 1977). No caso da função (20), fazendo $\alpha = 0$ em (19), pode-se usar o procedimento padrão para regressão linear simples sobre a equação.

$$\log \frac{1 - F(x)}{F(x)} = \log a - b \log x$$

⁸ Das equações (19) e (20), a coerência dos parâmetros é verificada por $a = \lambda$ e $b = \delta$.

Algumas observações interessantes sobre o modelo log-logístico podem ser feitas com relação à variável $y = \log x$. O parâmetro b não é afetado pela escala adotada que, por extensão, não afeta as medidas de desigualdade. Por outro lado, a média da variável y , $E(y)$, é dada por

$$E(y) = \frac{1}{b} \log a$$

e a variância $V(y) = \pi^2/3b^2$. Pode-se notar a relação entre o parâmetro b e $V(y)$ que é a variância dos logs se a renda tem distribuição log-logística.

Uma interpretação geométrica para o parâmetro b vem do fato que

$$\left. \frac{dF(y)}{dy} \right|_{y = \frac{1}{b} \log a} = \frac{b}{4}$$

sendo b , portanto, um avaliador da inclinação da curva no seu ponto médio. Quanto maior b , então, maior a percentagem da população concentrada em torno da média. O parâmetro a pode ser usado em estudos de desigualdade de renda, verificando-se a característica $F(x=1) = 1/(1+a)$. Como a escala usada não interfere em b , pode-se usar uma unidade conveniente para aproveitar a informação transmitida por a . Assim, por exemplo, em estudos de renda pessoal, se a unidade monetária considerada for o salário mínimo, $1/(1+a)$ é a percentagem de sub-empregados na população.

A aplicação da distribuição log-logística em estudos de concentração industrial vem imediatamente da aderência entre esta curva e a log-normal que, já utilizada, propiciou bons resultados.

O modelo log-logístico biparamétrico, por ser de estimação simples e de fácil interpretação tanto no plano $(x, F(x))$ como no plano $(\log x, F(\log x))$, bem como apresentar a versatilidade de descrever densidades zero-modais e unimodais, apresenta-se como um modelo adequado de distribuição de renda. Seja por aproximar-se da curva log-normal, ou porque as suposições de Dagum são também aceitáveis para o caso da distribuição do capital ou receita setorial entre firmas, optou-se por utilizar este modelo no âmbito deste estudo.

3.7. Considerações Gerais

O desenvolvimento e aplicação de modelos mais sofisticados à distribuição de renda tem provocado intensa investigação sobre o assunto. A interpretação e a análise comparativa dos modelos, o desenvolvimento de testes de aderência e de uma teoria de amostragem particular para a área⁹, a generalização dos modelos para truncamento à esquerda e à direita e a necessidade de evolução de algoritmos de estimação eficientes, tudo isso constitui uma preocupação dos estudiosos de uma área que, por muito tempo relegada a segundo plano, passa a ocupar lugar de destaque na investigação econômica.

⁹ Ver Gastwirth (1972 e 1974).

Se, de um prisma quantitativo, o interesse pela distribuição de renda é notável, do ponto de vista qualitativo, a relação entre essa questão e o desenvolvimento econômico é matéria de particular interesse e, talvez seja, atualmente, "o principal problema da Economia Política" como o queria Ricardo.

Neste contexto, a importância da estrutura industrial como elemento preponderante na fixação das parcelas lucros e salários do produto gerado cresce de significado. A percepção dos movimentos dessa estrutura, no sentido de maior ou menor grau de concentração, bem como o entendimento dos mecanismos que guiam esses movimentos é tarefa primordial dos estudiosos da Economia.

A modelagem e quantificação da situação distributiva dos atributos econômicos ligados à estrutura industrial, como estoque de capital, valor da produção, lucros, receita e outros, vêm fornecer subsídios mais consistentes às análises e estudos na área. Desse ponto de vista, a adequabilidade dos modelos de distribuição de renda para descrever outros processos distributivos, caso verificada, fornece instrumental muito importante para a caracterização instantânea da situação econômica sob análise.

4. EVOLUÇÃO RECENTE DA CONCENTRAÇÃO INDUSTRIAL NO BRASIL

4.1. Introdução

Nos capítulos anteriores, foi examinada toda uma gama de índices de concentração e modelos de distribuição de atributos econômicos, procurando avaliar as suas possibilidades de aplicação ao estudo da concentração industrial. Neste capítulo, duas das medidas de concentração previamente analisadas são empregadas para a verificação da evolução da estrutura industrial brasileira entre 1973 e 1977.

Os índices de concentração utilizados são o índice de Theil e o coeficiente de variação de Pearson¹. Estas medidas, que se completam por avaliarem o grau de concentração de um determinado atributo econômico por formas bastante diversas, foram aplicadas à distribuição do atributo patrimônio líquido das firmas conforme a listagem da revista Visão, em sua edição especial anual "Quem é quem na Economia Brasileira". O nível de agregação utilizado na análise corresponde aos 46 (quarenta e seis) subsetores em que a referida publicação divide a indústria da transformação nacional, agrupados em 13 (treze) setores.

4.2. Observações Metodológicas

A aplicação dos índices selecionados aos da

¹ Ver seções 2.4 e 2.5.

dos foi feita de formas a obter o índice global, decomposto pelas contribuições de cada subsetor, como em 2.4 e 2.5. Obtém-se, assim, tanto a avaliação da desigualdade em cada subsetor, como a participação que cada um deles tem na desigualdade da indústria da transformação como um todo.

O conjunto de empresas do "Quem é quem" corresponde àquelas cujo patrimônio líquido à época do balanço seja não inferior a 6,5 milhões de cruzeiros (edição de 1974) ou 17,5 milhões de cruzeiros (edição de 1978). O piso adotado corresponde aproximadamente ao milhão de dólares americanos^{2,3}.

Assim sendo, estamos conceitualmente lidando com uma população: aquela apresentada na publicação já referida. Os índices a serem calculados não são estimativas do valor do índice real da indústria de transformação; representam o valor real do índice para a população considerada.

A população, entretanto, varia entre os dois períodos considerados. Dado ao critério adotado para a entrada de empresas no conjunto do "Quem é quem", haverá indivíduos que, incluídos na edição de 1974 estarão excluídos da de 1978 e vice-versa. Por outro lado, para cada subsetor considerado, a população varia entre as duas edições não somente com a diferença interna de crescimento (entre as empresas do subsetor) mas também conforme o crescimento do nível de atividade econômica subsetorial seja maior ou menor que a variação da taxa cambial entre os períodos considerados.

² Ver Visão, "Quem é quem" (agosto / 1978), pg 45 a 48.

³ A edição de 1978 diz respeito a balanços encerrados no 2º semestre de 1977 ou no 1º semestre de 1978. A edição de 1974 refere-se a período equivalente.

A compatibilidade temporal dos índices obtidos, todavia, não é afetada pelas questões expostas acima, embora faça-se necessário realçá-las para determinar exatamente o que (e como) está sendo medido.

No quadro 01, estão listados os setores e sub-setores nos quais está dividida a indústria de transformação nacional, segundo a classificação da revista Visão. Ao lado do código alfabético utilizado pela publicação e da descrição da atividade, apresenta-se um código numérico que será utilizado no decorrer deste capítulo.

4.3. Resultados Obtidos com o Índice de Theil

O índice de Theil foi calculado para cada um dos subsetores apresentados no Quadro 01, para as tabelas das edições de 1974 e 1978 do "Quem é quem" da Revista Visão. O atributo utilizado foi o patrimônio líquido, escolhido pelo fato de que as tabelas citadas são organizadas a partir desta grandeza contábil. O cálculo foi feito a partir dos dados individuais não-deflacionados e, obtido o valor do índice para cada subsetor, a formulação (08) da seção 2.5 foi usada para evoluir os valores do índice a nível setorial e global.

No quadro 02, é apresentado o peso relativo de cada um dos setores e subsetores com relação ao número de empresas e ao total do patrimônio líquido das empresas incluídas na análise. Estas grandezas desempenham papel importante no cálculo do índice de Theil, ao mesmo tempo em que permitem uma avaliação do crescimento da importância do (sub)setor no contexto da indústria da transformação. A comparação entre as participações relati

Quadro 01

SUBDIVISÃO DA INDÚSTRIA DE TRANSFORMAÇÃO EM SETORES E SUBSETORES
(Classificação da Editora Visão LTDA *)

Descrição da Atividade	Código Alfabe tico	Código Numérico	
		Setor	Subsetor
PRODUTOS DE MINERAIS NÃO-METÁLICOS		0100	
Cal e cimento	CIM		01
Cerâmica e artefatos de cimento, gesso e amianto	CER		02
Artefatos de vidro e cristal	VID		03
Outros produtos de minerais não-metálicos	ONM		04
METALURGIA		0200	
Siderurgia	SID		05
Metalurgia dos não-ferrosos	MNF		06
Produtos metalúrgicos diversos	MET		07
MECÂNICA		0300	
Máquinas, motores e equipamentos industriais	MAQ		08
Máquinas, aparelhos e instrumentos para escritório	INS		09
MATERIAL ELETRO-ELETRÔNICO		0400	
Material elétrico	MEL		10
Aparelhos domésticos, material de comunicação, lâmpadas, acessórios e utensílios	DOM		11
MATERIAL DE TRANSPORTE		0500	
Construção Naval	NAV		12
Material Ferroviário	FER		13
Veículos automotores	AUT		14
Autopeças e carrocerias	APÇ		15
Tratores, máquinas de terraplanagem e implementos	TRA		16
Aviões e outros veículos	AVV		17

cont.

Quadro 01 - (Cont.)

Descrição da Atividade	Código Alfabé- tico	Código Numérico	
		Setor	Subsetor
MÓVEIS E MADEIRA		0600	
Móveis e artefatos de decorações	MOV		18
Madeira, compensados, laminados, artefatos	MAD		19
Papel e papelão	PAP		20
QUÍMICA E FARMACÊUTICA		0700	
Química e petroquímica	QUI		21
Plásticos e derivados	PLA		22
Petróleo (refinação e distribuição)	PET		23
Gás (de petróleo)	GAS		24
Produtos farmacêuticos, medicinais e veterinários	FAR		25
Produtos de perfumaria, higiene doméstica, velas, etc.	PER		26
COURO E BORRACHA		0800	
Couros e peles	COU		27
Borracha e artefatos	BOR		28
TÊXTIL E VESTUÁRIO		0900	
Fiação e tecelagem	TEX		29
Artefatos de tecidos	ART		30
Artigos de vestuário e acessórios	VES		31
PRODUTOS ALIMENTÍCIOS		1000	
Moinhos	MOI		32
Carnes frigoríficas e industrializadas	FRI		33
Pescado (pesca e industrialização)	PES		34
Laticínios	LAT		35
Açúcar e álcool	AÇU		36
Café	CAF		37
Óleos vegetais	OLV		38
Produtos alimentares diversos	ALM		39

Cont.

Quadro 01 - (Cont.)

Descrição da Atividade	Código Alfabé- tico	Código Numérico	
		Setor	Subsetor
BEBIDAS E FUMO		1100	
Bebidas e refrigerantes em geral	BEB		40
Fumo e produtos de fumo	FUM		41
EDITORIAL E GRÁFICA		1200	
Jornais	JOR		42
Editoras e Gráficas	GRF		43
INDÚSTRIAS DIVERSAS		1300	
Conglomerados	CGL		44
Instrumentos e aparelhos profissionais e de pre- cisão	APP		45
Artigos diversos	DIV		46

* A subdivisão base é a da edição de 1974 do "Quem é quem". Assim ,
o subsetor CALÇADOS está incluído no subsetor VES.

Quadro 02

Participação Relativa dos Setores e Subsetores no Total de Empresas e do Patrimônio Líquido da Indústria de Transformação (Valores monetários em Cr\$ 1.000.000.000,00 do respectivo ano)

i	j	1 9 7 3				1 9 7 7				Variação(x10 ²)
		n _j ^(a)	P _j ^(b)	n _j /n	P _j /P	n _j	P _j	n _j /n	P _j /P	
01		<u>137</u>	<u>6,13</u>	<u>0,0547</u>	<u>0,0505</u>	<u>152</u>	<u>24,52</u>	<u>0,0479</u>	<u>0,0452</u>	
	01	45	2,83	0,0179	0,0233	41	11,60	0,0129	0,0214	-0,19
	02	70	2,01	0,0279	0,0165	78	7,50	0,0246	0,0158	-0,27
	03	12	0,83	0,0048	0,0068	20	4,08	0,0063	0,0075	+0,07
	04	10	0,46	0,0040	0,0038	13	1,34	0,0041	0,0025	-0,13
02		<u>308</u>	<u>20,07</u>	<u>0,1226</u>	<u>0,1652</u>	<u>412</u>	<u>89,61</u>	<u>0,1298</u>	<u>0,1651</u>	
	05	38	11,26	0,0151	0,0927	58	50,87	0,0183	0,0937	+0,10
	06	31	2,39	0,0123	0,0197	43	12,75	0,0135	0,0235	+0,38
	07	239	6,42	0,0951	0,0528	311	25,99	0,0980	0,0479	-0,49
03		<u>175</u>	<u>5,51</u>	<u>0,0696</u>	<u>0,0454</u>	<u>273</u>	<u>32,34</u>	<u>0,0860</u>	<u>0,0596</u>	
	08	162	5,02	0,0645	0,0413	255	29,12	0,0803	0,0537	+1,24
	09	13	0,49	0,0052	0,0040	18	3,22	0,0057	0,0059	+0,19
04		<u>104</u>	<u>4,59</u>	<u>0,0414</u>	<u>0,0378</u>	<u>135</u>	<u>22,36</u>	<u>0,0425</u>	<u>0,0412</u>	
	10	49	1,78	0,0195	0,0147	76	9,97	0,0239	0,0184	+0,37
	11	55	2,81	0,0219	0,0231	59	12,39	0,0186	0,0228	-0,03
05		<u>147</u>	<u>12,39</u>	<u>0,0585</u>	<u>0,1020</u>	<u>177</u>	<u>51,02</u>	<u>0,0557</u>	<u>0,0940</u>	
	12	10	1,11	0,0040	0,0091	17	8,96	0,0054	0,0165	+0,74
	13	8	0,70	0,0032	0,0058	8	4,14	0,0025	0,0076	+0,18
	14	8	5,93	0,0032	0,0488	13	19,07	0,0041	0,0351	-1,37
	15	84	3,13	0,0334	0,0258	97	12,56	0,0306	0,0231	-0,27
	16	30	1,14	0,0119	0,0094	35	3,46	0,0110	0,0064	-0,30
	17	7	0,38	0,0028	0,0031	7	2,83	0,0022	0,0052	+0,21
06		<u>216</u>	<u>6,39</u>	<u>0,0860</u>	<u>0,0526</u>	<u>287</u>	<u>29,25</u>	<u>0,0904</u>	<u>0,0539</u>	
	18	20	0,39	0,0080	0,0032	49	2,22	0,0154	0,0041	+0,09
	19	106	2,31	0,0422	0,0190	132	10,30	0,0416	0,0190	+0,00
	20	90	3,69	0,0358	0,0304	106	16,73	0,0334	0,0308	+0,04
07		<u>317</u>	<u>29,37</u>	<u>0,1261</u>	<u>0,2417</u>	<u>392</u>	<u>148,63</u>	<u>0,1235</u>	<u>0,2739</u>	
	21	164	8,33	0,0653	0,0686	205	46,60	0,0646	0,0859	+1,73
	22	52	1,49	0,0207	0,0123	84	7,39	0,0265	0,0136	+0,13
	23	25	16,79	0,0099	0,1382	21	86,90	0,0066	0,1601	+2,19
	24	9	0,72	0,0034	0,0059	10	1,75	0,0031	0,0032	-0,27
	25	46	1,41	0,0183	0,0116	46	4,19	0,0145	0,0077	-0,39
	26	21	0,63	0,0084	0,0052	26	1,80	0,0082	0,0033	-0,19
08		<u>66</u>	<u>1,99</u>	<u>0,0263</u>	<u>0,0164</u>	<u>82</u>	<u>7,17</u>	<u>0,0258</u>	<u>0,0132</u>	
	27	31	0,52	0,0123	0,0043	37	2,07	0,0117	0,0038	-0,05
	28	35	1,47	0,0139	0,0121	45	5,10	0,0142	0,0094	-0,27
09		<u>385</u>	<u>10,04</u>	<u>0,1552</u>	<u>0,0826</u>	<u>456</u>	<u>38,16</u>	<u>0,1436</u>	<u>0,0703</u>	
	29	264	7,09	0,1051	0,0584	273	24,13	0,0860	0,0445	-1,39
	30	41	1,08	0,0163	0,0089	58	4,26	0,0183	0,0079	-0,10
	31	80	1,87	0,0318	0,0154	125	9,77	0,0394	0,0180	+0,26
10		<u>453</u>	<u>13,05</u>	<u>0,1803</u>	<u>0,1074</u>	<u>563</u>	<u>58,07</u>	<u>0,1773</u>	<u>0,1070</u>	
	32	44	1,35	0,0175	0,0111	50	5,12	0,0157	0,0094	-0,17
	33	56	1,52	0,0223	0,0125	80	6,87	0,0252	0,0127	+0,02
	34	34	0,81	0,0135	0,0067	41	2,12	0,0129	0,0039	-0,28
	35	16	0,76	0,0064	0,0063	30	1,98	0,0094	0,0036	-0,27
	36	154	4,25	0,0613	0,0350	162	20,20	0,0510	0,0372	+0,22
	37	8	0,45	0,0032	0,0037	21	1,93	0,0066	0,0036	-0,01
	38	43	1,05	0,0171	0,0086	48	5,40	0,0151	0,0100	+0,14
	39	98	2,86	0,0390	0,0235	131	14,45	0,0413	0,0266	+0,31
	11		<u>67</u>	<u>4,25</u>	<u>0,0267</u>	<u>0,0350</u>	<u>73</u>	<u>15,28</u>	<u>0,0230</u>	<u>0,0282</u>
40		61	2,67	0,0243	0,0220	67	9,69	0,0211	0,0179	-0,41
41		6	1,58	0,0024	0,0130	6	5,59	0,0019	0,0103	-0,27
12		<u>67</u>	<u>1,79</u>	<u>0,0267</u>	<u>0,0147</u>	<u>78</u>	<u>5,09</u>	<u>0,0246</u>	<u>0,0094</u>	
	42	13	0,45	0,0052	0,0037	17	1,58	0,0054	0,0029	-0,08
	43	54	1,34	0,0215	0,0110	61	3,51	0,0192	0,0063	-0,45
13		<u>71</u>	<u>5,94</u>	<u>0,0283</u>	<u>0,0489</u>	<u>95</u>	<u>21,14</u>	<u>0,0299</u>	<u>0,0390</u>	
	44	7	4,82	0,0028	0,0397	9	16,08	0,0028	0,0296	-1,01
	45	22	0,36	0,0086	0,0030	29	2,12	0,0091	0,0039	+0,09
	46	42	0,76	0,0167	0,0063	57	2,94	0,0180	0,0054	-0,09
TOTAIS		<u>2513</u>	<u>121,49</u>	<u>1,0000</u>	<u>1,0000</u>	<u>3175</u>	<u>542,64</u>	<u>1,0000</u>	<u>1,0000</u>	<u>0,00</u>

Obs.: a) n_j é o número de empresas no (sub)setor j e $n = \sum_j n_j$

b) P_j é o patrimônio líquido agregado do (sub)setor j e $P = \sum_j P_j$

vas de um determinado (sub)setor no total do patrimônio líquido pode mostrar em quanto o sub(setor) cresceu relativamente ao crescimento do conjunto de empresas consideradas. A última coluna do quadro 02 registra a variação do peso do (sub)setor na formação do patrimônio líquido agregado. As variações positivas indicam que essa conta do (sub)setor cresceu mais que a conta agregada, enquanto que as variações negativas indicam crescimento menor que o do agregado.

No quadro 03, são apresentados os valores do índice de Theil obtidos para cada subsetor, em cada um dos anos considerados (1973 e 1977), os fatores de normalização do índice, os valores normalizados do índice, a variação do índice normalizado entre os dois períodos e, para cada ano, a posição de cada setor no "ranking" de concentração.

Diferentemente do que foi proposto na equação (07) da seção 2.5, o fator de normalização aqui utilizado é deduzido a partir da constatação de que o índice foi calculado para um conjunto de empresas com um valor mínimo a_0 para o atributo em questão. Neste caso, o valor da máxima desigualdade possível, ou seja, o fator de normalização, é dado por

$$f(a_0, P, n) = \frac{a_0}{P} (1 - n) \left[\log \frac{P}{a_0} - (n - 1) \right] + \log \left[n(1 - (n-1)\frac{a_0}{P}) \right]$$

onde P é o valor agregado do atributo e n é o número de indivíduos envolvidos no cálculo.

Os dados do quadro 03 permitem análises em

Quadro 03

Valores do Índice de Theil por Subsetor para os Anos de 1973 e 1977

j	T_j^1 {2}	T_j^2 {3}	f_j^1 {4}	f_j^2 {5}	T_j^1 {6}	Posição {7}	T_j^2 {8}	Posição {9}	Variação(x10 ²) {10}
01	0,2405	0,2084	3,0908	3,2631	0,0777	46	0,0639	46	- 1,38
02	0,4910	0,5139	2,7729	3,1054	0,1771	33	0,1655	35	- 1,16
03	0,6087	1,0417	1,9848	2,4734	0,3067	11	0,4212	05	+11,45
04	0,6384	0,8282	1,6422	1,7413	0,3887	06	0,4756	03	+ 8,69
05	1,0400	1,2323	3,4572	3,8847	0,3008	12	0,3172	09	+ 1,64
06	0,7703	0,8951	2,8738	3,5253	0,2680	13	0,2692	15	+ 0,12
07	0,5562	0,4745	3,6056	4,0301	0,1543	37	0,1177	42	- 3,66
08	0,7014	0,7359	3,5164	4,2688	0,1995	22	0,1724	34	- 2,71
09	0,3995	0,6261	1,7311	2,3206	0,2308	19	0,2698	13	+ 3,90
10	0,6449	0,7728	2,7491	3,3729	0,2346	18	0,2291	25	- 0,55
11	0,7065	0,8124	3,1325	3,4615	0,2255	20	0,2347	23	+ 0,92
12	0,5043	1,0529	1,9804	2,6075	0,2546	15	0,4038	07	+14,92
13	0,5427	0,3582	1,7124	1,8886	0,3169	09	0,1897	32	-12,72
14	0,4812	0,7157	2,0195	2,4770	0,2383	17	0,2889	11	+ 5,06
15	0,5928	0,6923	3,2096	3,5707	0,1847	28	0,1939	28	+ 0,92
16	0,7526	0,3016	2,3960	2,4899	0,3141	10	0,1211	41	-19,30
17	0,5467	0,7114	1,4312	1,7208	0,3820	07	0,4134	06	+ 3,14
18	0,2278	0,1763	1,4390	1,7638	0,1583	35	0,1000	45	- 5,83
19	0,2984	0,5069	2,6774	3,2676	0,1115	42	0,1551	37	- 4,36
20	0,6268	1,0018	3,3621	3,8061	0,1864	26	0,2632	16	+ 7,68
21	0,7255	0,8836	4,0710	4,6452	0,1782	32	0,1902	30	+ 1,20
22	0,4692	0,6006	2,5464	3,0667	0,1843	29	0,1958	27	+ 1,15
23	2,1307	2,2367	3,1366	3,0062	0,6793	02	0,7440	01	+ 6,47

cont.

j	T_j^1	T_j^2	f_j^1	f_j^2	T_j^1	Posição	T_j^2	Posição	Varição ($\times 10^2$)
{1}	$\{2\}$	$\{3\}$	$\{4\}$	$\{5\}$	$\{6\}$	$\{7\}$	$\{8\}$	$\{9\}$	$\{10\}$
24	0,2603	0,4853	1,7877	1,8023	0,1456	39	0,2693	14	+12,37
25	0,3483	0,5040	2,5284	2,6300	0,1378	41	0,1916	29	+ 5,38
* 26	0,7611	0,4918	1,9173	1,9212	0,3970	05	0,2560	19	-14,10
* 27	0,2511	0,2468	1,4970	1,9058	0,1677	34	0,1295	39	- 3,82
28	1,0820	1,1988	2,6019	2,8109	0,4158	04	0,4265	04	+ 1,07
* 29	0,5136	0,5067	3,6801	4,0071	0,1396	40	0,1265	40	- 1,31
30	0,2476	0,4458	2,2736	2,5695	0,1089	43	0,1735	33	+ 6,46
31	0,6570	1,1101	2,5944	3,2281	0,2532	16	0,3439	08	+ 9,07
* 32	0,4714	0,4271	2,4955	2,8084	0,1889	25	0,1521	38	- 3,68
* 33	0,6504	0,6010	2,5374	3,0006	0,2563	14	0,2003	26	- 5,60
* 34	0,3143	0,2944	2,0224	1,8612	0,1554	36	0,1582	36	+ 0,28
* 35	0,9703	0,4825	2,0421	1,9690	0,4751	03	0,2450	21	-23,01
36	0,5971	1,0181	3,3158	3,9748	0,1801	30	0,2561	18	+ 7,60
37	0,1654	0,4758	1,5552	2,0278	0,1064	44	0,2346	24	+12,82
38	0,4118	0,7094	2,2163	2,8581	0,1858	27	0,2482	20	+ 6,24
39	0,6501	3,9510	3,0490	3,6735	0,2132	21	0,2589	17	+ 4,57
* 40	1,1303	1,0219	3,0970	3,5400	0,3650	08	0,3060	10	- 5,90
* 41	1,2656	1,2315	1,6584	1,6860	0,7631	01	0,7304	02	- 3,27
42	0,3235	0,4559	1,6731	1,8747	0,1934	24	0,2432	22	+ 4,98
* 43	0,4769	0,4327	2,3409	2,2759	0,2037	23	0,1901	31	- 1,36
44	0,1770	0,2448	1,8844	-2,1292	0,0939	45	0,1150	44	+ 2,11
45	0,1896	0,2392	1,2730	2,0565	0,1489	38	0,1163	43	- 3,26
46	0,3208	0,5929	1,7876	2,0648	0,1795	31	0,2871	12	+10,76

Obs.: a) Os superescritos 1 ou 2 indicam que o conteúdo diz respeito aos anos de 1973 e 1977.

b) T_j , f_j e T_j^1 designam o índice de Theil, o fator de normalização e o índice normalizado do subsetor j.

c) A coluna "variação" mede a variação ($T_j^2 - T_j^1$).

várias direções. É interessante lembrar aqui que, mesmo trabalhando-se com o índice de Theil normalizado, o índice não-normalizado traz uma informação bastante significativa, derivada do conceito de entropia⁴. Em estudos de concentração industrial, esse conceito cresce de significado, na medida em que um aumento do número de empresas que disputam o mercado reflete uma tendência oposta àquela da oligopolização.

Desta forma, índice de Theil pode crescer ou decrescer entre dois períodos, refletindo a entrada ou saída de firmas no mercado. Portanto, o índice de Theil pode captar os efeitos do que se denomina de concentração absoluta (abandono do mercado por empresas, favorecendo a tendência oligopolística) em contraste à concentração relativa (diferenciação das taxas de crescimento entre empresas).

De forma geral, o índice de Theil, na forma não-normalizada, diminui tanto com o aumento de empresas no mercado como com a tendência a uma distribuição mais equânime do atributo considerado. Por outro lado, o índice normalizado verifica somente o nível de equidade na distribuição, indiferentemente ao número de elementos na população (amostra) considerada.

Assim, os dados do quadro 03 servem ao exame "cross-section" da estrutura industrial brasileira em dois períodos diversos e permitem, também, uma análise da evolução da estrutura de cada sub-setor entre 1973 e 1977. O estudo pode ser feito de formas a considerar o problema da concentração relativa somente; através do índice de Theil normalizado, ou visualizando a questão mais amplamente, através do índice não-normalizado.

⁴ Ver seção 2.5 para uma interpretação do índice de Theil à luz do conceito de entropia.

Analisando-se os dados da segunda coluna, por exemplo, verifica-se que dos dez subsetores mais concentrados (subsetores 23, 41, 40, 28, 05, 35, 06, 26, 16 e 21, em ordem decrescente da medida de concentração), seis (os destacados) revelam presença preponderante de empresas estatais ou sob controle acionário de grupos estrangeiros⁵. Considerando apenas os cinco subsetores mais concentrados (petróleo, fumo, bebidas, borracha e siderurgia, nessa ordem), somente o terceiro deles apresenta dominância da empresa nacional privada⁶. Os valores da terceira coluna confirmam a presença maciça de empresas estatais ou multinacionais nos subsectores mais concentrados: aqui, dos dez maiores índices de concentração (subsectores 23, 05, 41, 28, 31, 12, 03, 40, 36, 20, em ordem decrescente), cinco, em especial os quatro primeiros, pertencem a subsectores onde a empresa nacional privada não é predominante.

Por outro lado, através das colunas 7 e 9 do quadro 03, pode-se verificar que, à exceção dos subsectores 21 e 25, os demais que apresentam participação da empresa nacional privada abaixo de 50% do patrimônio líquido do subsector estão acima da posição média do "ranqueamento".

A análise da evolução do índice por subsector pelos dados das colunas 02 e 03 do quadro 03 mostra que, dos 46 subsectores, apenas 15 apresentam redução do índice de Theil entre 1973 e 1977 (estão assinalados por asterisco à esquerda do código numérico na primeira coluna). Destes, apenas os subsectores 16 e 41 (tra-

⁵ Os dados sobre presença de estatais e multinacionais originam-se do "Quem é quem" da Editora Visão LTDA.

⁶ Os termos empresa estatal, nacional e multinacional, dizem respeito ao fato de o controle acionário da empresa pertencer a capital estatal, nacional privado ou estrangeiro.

tores, máquinas de terraplenagem, implementos e fumo) revelam - se não dominados pela empresa nacional privada. Utilizando-se o índice normalizado (dados das colunas 6, 8 e 9), verifica-se que, do ponto de vista da concentração relativa, dos 14 (quatorze) subsectores onde predominam o capital estrangeiro ou as empresas estatais, apenas 3 (três) registram decréscimo do valor do índice. Enquanto isso, os 36 (trinta e seis) subsectores nos quais predominam a empresa privada nacional dividem-se igualmente entre variações positivas e negativas.

A evidência estatística parece indicar que a presença da empresa multinacional ou estatal na estrutura industrial brasileira está, em geral, associada a níveis de concentração acima da média. Esta hipótese será testada, em um outro contexto metodológico, no capítulo seguinte.

No quadro 04, utilizando dados dos quadros anteriores, são mostrados os valores do índice de Theil para um outro nível de agregação, obtidos segundo a equação (08). Aí, verifica-se que, globalmente, o índice de Theil, tanto normalizado como não, evoluiu significativamente no período analisado, crescendo em 15,6% na aferição não-normalizada e em 6,8% na aferição normalizada. Outro dado significativo é que, na apuração dos índices não-normalizados, constata-se que 91,16% do seu valor em 1973 e 91,39% em 1977, deve-se à parcela proveniente das desigualdades internas dos setores. Obtendo o índice global a partir da decomposição em subsectores, as desigualdades internas representam 57,83% e 60,3% do total (respectivamente em 1973 e 1977). Verifica-se, então, que apesar de o processo de crescimento industrial no Brasil ser bastante desequilibrado, acelerando alguns setores prioritários em detrimento de outros, os desequilíbrios intra-setoriais são mais re-

Quadro 04

Valores do Índice de Theil por Setor e Global (1973 e 1977)^a

Setor	T_i^1	T_i^2	f_i^1	f_i^2	$T_i'^1$	$T_i'^2$
01	0,4750	0,5965	3,7990	4,1413	0,1250	0,1375
02	1,5720	1,5227	4,8368	5,2585	0,3250	0,2896
03	0,6760	0,7331	3,5982	4,3666	0,1879	0,1679
04	0,6967	0,8223	3,5529	4,0560	0,1961	0,2027
05	1,3165	1,2231	4,3383	4,6360	0,3035	0,2638
06	0,5362	0,8596	3,6755	4,2340	0,1459	0,2030
07	2,3084	2,7347	5,1029	5,5098	0,4524	0,4963
08	0,9565	0,9796	2,7864	3,0407	0,3433	0,3222
09	0,5133	0,6568	3,9131	4,3332	0,1312	0,1516
10	0,5919	0,8327	4,2062	4,8061	0,1407	0,1733
11	1,4766	1,4103	3,4546	3,6531	0,4274	0,3861
12	0,4538	0,4629	2,6069	2,6288	0,1741	0,1761
13	1,5628	1,5623	3,6669	3,9269	0,4262	0,3978
GLOBAL	1,4416	1,6669	6,3824	6,9075	0,2259	0,2413

Obs.: a) Ver quadro 03 para descrição das variáveis das colunas.

levantes e continuam crescendo de importância no desequilíbrio global. Esta constatação pode receber mais ênfase, repetindo-se que o critério metodológico usado neste trabalho exclui as empresas de pequeno porte. A inclusão destas evidenciaria o papel das desigualdades internas dos (sub)setores, aumentando a participação desta parcela na desigualdade da indústria de transformação nacional.

4.4. Aplicação do Coeficiente de Variação e Resultados Obtidos

Com o intuito de obter uma confirmação para os resultados da seção anterior, submetemos os dados já utilizados a uma nova apreciação, utilizando-se como medida de concentração o coeficiente de variação de Pearson. As características principais deste índice de desigualdade foram previamente discutidas na seção 2.4. A fim de possibilitar a decomposição do índice, faz-se uso, no decorrer da exposição, da forma quadrática do coeficiente, também apresentada na seção 2.4.

O conceito de normalização do índice, aplicado na seção anterior, é necessário também neste caso. Em situações onde o zero é definido como valor mínimo para o atributo sob estudo, o fator de normalização para o coeficiente de variação é $\sqrt{n-1}$, onde n é o total de elementos da população. No caso deste trabalho; onde o índice é aplicado a populações onde o atributo tem a_0 como valor mínimo, o fator de normalização (ou o valor do índice para a situação de máxima desigualdade) é dado por

$$g(\bar{x}, a_0, n) = \left(1 - \frac{a_0}{\bar{x}}\right) \sqrt{n-1}$$

onde \bar{x} é o valor médio do atributo com relação aos n elementos da população.

No quadro 05, estão apresentados os valores obtidos para o coeficiente de variação, aos dois níveis de agregação descritos no quadro 01, para os anos de 1973 e 1977.

Os resultados são, em geral, compatíveis com os obtidos através da aplicação do índice de Theil. Para verificação da coerência entre os índices, foram comparados, para os dois anos, os "rankings" dos índices normalizados, através do teste de correlação de ordem de Spearman⁷: Os valores obtidos para a estatística r_s ,

$$r_s = 1 - \frac{\sum_j d_j^2}{n(n^2 - 1)} \quad j = 1, 2, \dots, n ; n = 46$$

conhecida como coeficiente de correlação de ordem de Spearman, foram superiores a 0,85, enquanto o intervalo de aceitação da hipótese de não-associação entre os dois "rankings", para $\alpha = 0,01$, é $[-0,38; +0,38]$.

Quanto à variação do índice entre os dois anos considerados, ressalta o que foi postulado na seção 2.5 com respeito à grande sensibilidade do índice de Theil às mudanças na distribuição. Em geral, as variações percentuais do coeficiente de

⁷ Ver Anderson & Sclove (1978), pg 599-603.

Valores do Coeficiente de Variação por Subsetor, Setor e Global (1973 e 1977)

i	j	$v_j^{1(b)}$	v_j^2	$g_j^{1(c)}$	g_j^2	$v_j^{1(d)}$	v_j^2	Variação ($\times 10^2$) e
{1}	{2}	{3}	{4}	{5}	{6}	{7}	{8}	{9}
01								
	01	1,1559	1,3852	9,9669	10,9547	0,1160	0,1264	- 1,44
	02	0,7296	0,6425	5,9476	5,9333	0,1227	0,1083	- 0,07
	03	1,2401	1,3816	6,4224	7,1779	0,1931	0,1924	+ 6,12
	04	1,3519	2,0363	3,0056	3,9847	0,4498	0,5110	+ 6,78
	05	1,3321	1,6818	2,5754	2,8749	0,5172	0,5850	
02								
	01	3,4006	3,7700	15,7729	18,6420	0,2156	0,2022	- 1,05
	02	1,8934	2,2776	5,9492	7,3992	0,3183	0,3078	- 3,18
	03	1,5604	1,7041	5,0145	6,0981	0,3112	0,2794	- 2,55
	04	1,3706	1,2760	11,6942	13,9194	0,1172	0,0917	
03								
	01	1,8037	1,9034	10,4650	14,0560	0,1724	0,1354	- 4,00
	02	1,8653	1,9703	10,0264	13,4947	0,1860	0,1460	- 4,75
	03	1,0973	1,2500	2,8611	3,7202	0,3835	0,3360	
04								
	01	1,5917	2,2091	8,6555	10,3528	0,1839	0,2134	- 6,38
	02	1,5159	1,5207	5,6910	7,5044	0,2664	0,2026	+ 0,34
	03	1,5878	1,7522	6,4137	6,9809	0,2476	0,2510	
05								
	01	2,8062	2,5371	11,1510	12,4611	0,2517	0,2036	+ 9,20
	02	1,0159	1,7471	2,8237	3,8672	0,3598	0,4518	-12,61
	03	1,1622	0,8905	2,4498	2,5564	0,4744	0,3483	+ 2,57
	04	0,9419	1,3174	2,6225	3,4228	0,3592	0,3849	- 1,50
	05	1,5092	1,5908	7,5212	8,4735	0,2007	0,1877	-18,17
	06	1,5779	0,8243	4,4638	4,7987	0,3535	0,1718	+ 4,95
	07	1,2184	1,4396	2,1574	2,3434	0,5648	0,6143	
06								
	01	1,4628	2,2339	11,4403	14,0078	0,1279	0,1595	- 9,99
	02	0,7327	0,6468	2,9060	4,2500	0,2521	0,1522	+ 2,64
	03	1,0303	1,5069	7,1903	8,8783	0,1433	0,1697	+ 5,28
	04	1,4722	2,1707	7,9377	9,1110	0,1855	0,2383	
07								
	01	7,9310	9,6588	17,7764	18,8611	0,4462	0,5121	- 1,44
	02	1,7987	1,9400	11,1327	13,1822	0,1616	0,1472	+ 2,20
	03	1,1535	1,6582	5,5219	7,2989	0,2089	0,2309	+ 5,13
	04	3,7721	3,6910	4,8516	4,4532	0,7775	0,8288	+12,12
	05	0,7006	1,0553	2,5984	2,7004	0,2696	0,3908	+ 7,53
	06	0,9307	1,3620	5,2893	5,4207	0,1760	0,2513	-17,34
	07	1,8251	1,2977	3,5057	3,7374	0,5206	0,3472	

Quadro 05 - (Cont.)

i	j	V_j^1	V_j^2	g_j^1	g_j^2	V_j^1	V_j^2	Varição ($\times 10^2$)
{1}	{2}	{3}	{4}	{5}	{6}	{7}	{8}	{9}
08		2,5016	2,3901	6,3223	7,1987	0,3640	0,3320	- 4,89
27		0,8472	0,8380	3,3635	4,1270	0,2519	0,2030	- 1,43
28		2,2129	2,4388	4,9258	5,6083	0,4492	0,4349	
09		1,3644	1,9954	14,7114	16,8698	0,0927	0,1183	
29		1,2962	1,2999	12,2948	13,2275	0,1054	0,0983	- 0,71
30		0,8179	1,1567	4,7609	5,7525	0,1718	0,2011	+ 2,93
31		1,8386	2,9830	6,4125	8,6429	0,2867	0,3451	+ 5,84
10		1,6814	2,2621	16,4642	19,6842	0,1021	0,1149	
32		1,2440	1,1263	5,8937	5,8037	0,2407	0,1941	- 4,66
33		1,5818	1,4240	5,6392	7,0777	0,2805	0,2012	- 7,93
34		0,9175	0,9684	4,1848	4,1808	0,2192	0,2316	+ 1,24
35		2,0034	1,2377	3,3401	3,9566	0,5998	0,3128	-28,70
36		1,9232	2,7701	9,4544	10,9078	0,2034	0,2540	+ 5,06
37		0,5885	1,0945	2,3388	3,7101	0,2516	0,2950	+ 4,34
38		1,3048	1,6414	4,7506	5,8513	0,2747	0,2805	+ 0,58
39		1,6827	2,1000	7,5928	9,5928	0,2197	0,2189	- 0,08
11		3,0228	3,0169	7,2954	7,7758	0,4143	0,3880	
40		2,4729	2,3800	6,6026	7,1408	0,3745	0,3333	- 4,12
41		1,8933	1,8464	2,1810	2,1941	0,8681	0,8415	- 2,66
12		1,2713	1,2635	6,1412	6,4219	0,2070	0,1967	
42		0,8309	1,1108	2,8107	3,2459	0,2956	0,3242	+ 2,86
43		1,5258	1,2828	5,6208	5,3930	0,2715	0,2379	+ 3,36
13		2,7920	2,9262	7,7159	8,9800	0,3618	0,3259	
44		0,5529	0,7506	2,4263	2,8007	0,2279	0,2609	+ 3,30
45		0,7059	0,7603	2,8377	4,0219	0,2488	0,1890	- 5,98
46		0,9984	1,7946	4,0913	4,9414	0,2440	0,3632	+11,92
GLOBAL		5,9237	7,9301	43,3812	50,5696	0,1366	0,1568	

- Obs.: a) Os superescritos 1 e 2 referem-se respectivamente a 1973 e 1977.
b) V_j é o coeficiente de variação não-normalizado do (sub)setor $j(i)$.
c) g_j é o fator de normalização do (sub)setor $j(i)$.
d) V_j^i é o coeficiente de variação normalizado do (sub)setor $j(i)$.
e) A coluna variação mede $(V_j^2 - V_j^1)$.

variação normalizado são inferiores às variações do índice de Theil normalizado, chegando mesmo, em cinco subsetores, a registrar diminuição do grau de concentração enquanto o Theil representa variação positiva, devendo este fato ser creditado à maior amplitude do conceito de desigualdade implícito na formulação do índice de Theil.

Sobre os dados do quadro 05, verifica-se, ao nível de agregação a 13 setores, que aqueles setores identificados mais facilmente como produtores de bens de consumo imediato (têxtil e vestuário, produtos alimentícios, bebidas e fumo), apresentam, nos dois anos, coeficientes de variação normalizados médios inferiores à média dos demais setores (0,2030 e 0,2071 contra 0,2447 e 0,2407, em 1973 e 1977, respectivamente), mesmo contando com o setor de bebidas e fumo, que registra o segundo valor para o índice de desigualdade em ambos os anos.

Também a nível de setores, foram verificadas as hipóteses de associação entre rapidez de crescimento e variação do grau de concentração e entre importância do setor e nível de concentração. Foi utilizado o teste não paramétrico de Spearman e os resultados indicaram a não-rejeição da hipótese de não-associação entre as variáveis.

Estas hipóteses, a primeira das quais já verificada para o período de crescimento rápido da economia brasileira entre 1968 e 1973⁸, serão testadas no capítulo seguinte, através de uma investigação econométrica que inclui análise de regressão tanto para os graus de concentração dos setores como para as suas variações.

⁸ Ver Tavares & Façanha (1977).

5. UMA ANÁLISE ESTRUTURAL DA CONCENTRAÇÃO INDUSTRIAL NO BRASIL

5.1. Introdução

No capítulo anterior, os valores obtidos para as medidas dos graus de concentração do patrimônio líquido dos (sub)setores da indústria da transformação nacional¹, nos anos de 1973 e 1977, permitiram a verificação de algumas evidências estatísticas na estrutura industrial brasileira. A distorção dos padrões normais de concentração pela presença dominante de empresas multinacionais e/ou estatais na estrutura (sub)setorial, as diferenças entre os graus médios de concentração dos setores produtos de bens de consumo imediato e outros fatos notáveis ao nível do estudo estatístico dos quadros 03, 04 e 05, carecem, pela sua importância, de uma investigação econométrica que possa avaliar globalmente o poder das variáveis envolvidas na "explicação", ao menos parcial, dos níveis de concentração observados.

O objetivo final deste capítulo é a elaboração e verificação empírica de modelos econométricos relacionados ao problema da concentração na Indústria de Transformação nacional em um período recente. Entretanto, sendo também proposta fundamental deste trabalho a generalização do uso de modelos de distribuição de renda à repartição de outros atributos, o contexto metodológico em que a análise econométrica se realiza engloba o ajustamento de uma função previamente discutida (a função de distribuição log-logísti

¹ Classificação do "Quem é quem" da Editora Visão LTDA. Ver quadro 01.

ca biparamétrica²) a dados de distribuição da receita total dos setores da Indústria de Transformação, segundo classificação da Secretaria da Receita Federal.

Observe-se que o atributo usado neste capítulo, a receita total das empresas, não tem sido normalmente utilizado em estudos de concentração industrial. Tem sido dada prioridade a atributos como o capital social (ou o patrimônio líquido), número de empregados, consumo de energia elétrica, valor da produção ou mesmo o faturamento. Cada um destes, assim como a receita total, representa de certo modo o "poder" da empresa no seu mercado e, certamente, apresentam entre si um significativo grau de correlação, sendo o faturamento e a receita total, praticamente idênticos.

A opção pela receita total prende-se à disponibilidade de dados para o período focado e, também, ao fato de ser esse atributo o reflexo final da atividade da empresa enquanto agente econômico. Os atributos número de empregados e consumo de energia elétrica revelam-se deficientes na medida em que devem minimizar o grau de concentração devido à maior eficácia na utilização desses recursos pelas empresas de maior porte. Por outro lado, o atributo lucro apresenta inúmeros problemas, desde a questão do lucro negativo até a da tributação diferenciada.

A receita total (ou o faturamento) é, sem embargo, o atributo que mais adequadamente avalia a posição da empresa com relação a seu mercado, por refletir conjuntamente os efeitos tamanho (em termos de capital) e eficiência da empresa, sendo também sensível apenas a mudanças conjunturais objetivas, além de só registrar a participação da empresa quando esta se encontra efetivamente disputando o mercado.

² Ver seção 3.6 deste trabalho.

5.2. Aspectos Metodológicos Preliminares

Conhecido o atributo com relação ao qual é de finido, neste capítulo, o grau de concentração, cumpre destacar a obtenção dos dados básicos para o cálculo do índice de concentração e a forma como foi calculado, bem como discutir a representatividade destes dados.

Os dados sobre distribuição da receita total das firmas provêm da publicação "Imposto de Renda - Pessoa Jurídica" da Secretaria da Receita Federal do Ministério da Fazenda. Foram utilizados os volumes referentes aos exercícios de 1974 a 1978 (anos-base de 1973 a 1977) no que concerne aos "resumos das declarações de rendimentos segundo classes de receita total". A classificação a dois dígitos da SRF foi adotada como subdivisão da Indústria de Transformação, representada pelos setores de atividade econômica codificados pelos números de 10 (dez) a 30 (trinta). O quadro 06 apresenta, sob a codificação vigente neste trabalho, esses 21 (vinte e um) setores bem como a correspondência entre a classificação da SRF e aquela utilizada no capítulo anterior.

Assim, para os 21 (vinte e um) setores do quadro 06 em cada um dos 5 (cinco) anos considerados, foi ajustada uma função de distribuição log-logística. O parâmetro b da função ajustada dá origem ao índice de concentração utilizado, ou seja,

$$I = 1/b$$

Como ficou patenteado na seção 3.6, o índice I é equivalente à razão de concentração de Gini para a função log-logística, quando $b > 1$. Para $0 < b \leq 1$, a função densidade log-lo

Quadro 06

SUBDIVISÃO DA INDÚSTRIA DE TRANSFORMAÇÃO EM SETORES, SEGUNDO A SECRETARIA DA RECEITA FEDERAL (E CORRESPONDÊNCIA COM CLASSIFICAÇÃO DO "QUEM É QUEM")

Código	Descrição da Atividade	Subsetores correspondentes no "Quem é quem" (Quadro 01)
01	Produtos de Minerais não-Metálicos	01, 02, 03, 04
02	Metalúrgica	05, 06, 07
03	Mecânica	08, 09
04	Material Elétrico e de Comunicações	10, 11
05	Material de Transporte	12 a 17
06	Madeira	18
07	Mobiliário	19
08	Papel e Papelão	20
09	Borracha	28
10	Couros e Peles	27
11	Química	21, 23, 24
12	Produtos Farmacêuticos e Veterinários	25
13	Perfumaria, Sabões e Velas	26
14	Produtos de Matérias Plásticas	22
15	Têxtil	29
16	Vestuário, Calçados e Artefatos de Tecidos	30, 31
17	Produtos Alimentares	32 a 39
18	Bebidas	40
19	Fumo	41
20	Editorial e Gráfica	42, 43
21	Diversas	44, 45, 46

gística é zero-modal e a relação entre Gini e o parâmetro b não é explicitável. Mesmo assim, o parâmetro b pode ser considerado como "avaliador do grau de igualdade entre os membros da população", variando no intervalo $(0; \infty)$. O índice I , portanto, varia no intervalo $(0; \infty)$, sendo tanto maior quanto menor for b e, portanto, quanto maior for o grau de desigualdade na população.

O parâmetro b da população (conjunto de empresas operando num setor da atividade econômica em um determinado ano) é estimado por ajustamento de curva, pelo método dos mínimos quadrados (como especificado na seção 3.6.), ao histograma de frequência da amostra de empresas correspondente àquelas pessoas jurídicas que apresentaram suas declarações de rendimentos através dos formulários I ou III³.

Não é possível tecer maiores considerações sobre a representatividade da amostra, a não ser reconhecer que o número de empresas amostradas é alto o suficiente para os objetivos do trabalho. Entretanto, não há indicações para que se faça um juízo preciso das implicações da escolha dessa amostra. Como já foi anteriormente reconhecido, a utilização de estatísticas levantadas com outros propósitos é uma das limitações deste trabalho e, portanto, as conclusões a serem tiradas devem ser sempre vinculadas à amostra utilizada.

5.3. Obtenção dos Parâmetros da Função Log-Logística

Para que se desenvolva a investigação econométrica objeto deste capítulo, faz-se mister ajustar aos dados de

³ Ver publicações da Secretaria da Receita Federal citadas nas referências bibliográficas.

distribuição (histograma de receita total) curvas log-logísticas para a obtenção do parâmetro b , que determinará o valor do índice de concentração I a ser utilizado.

As estimações lineares necessárias foram realizadas em máquinas calculadoras de mão, dispensando-se o uso de computador. No quadro 07, estão apresentados os resultados obtidos para os parâmetros a e b da função de distribuição logística bi-paramétrica para cada um dos 21 setores em cada um dos 5 anos abordados, bem como o correspondente valor do coeficiente de determinação R^2 da regressão. Observe-se que o ajustamento é, na maioria dos casos, bastante aceitável, não ficando comprometida a adequação da curva usada para a descrição dos dados observados. Saliencia-se que as estimações se prenderam rigorosamente aos dados, utilizando-se todas as classes de receita total (dezessete, em todos os casos), conforme a fonte, excluindo-se apenas as empresas com receita total nula.

Sabe-se que o parâmetro b não é afetado pela unidade monetária utilizada, o mesmo não acontecendo com o parâmetro de escala a . Para viabilizar uma análise da variação de a , já que a quantidade $1/(1 + a)$ corresponde ao percentual da população que percebe renda inferior à unidade monetária utilizada, foi aplicado aos dados o deflator implícito no índice de preços para o total da Indústria de Transformação (coluna 53 da Conjuntura Econômica) e a unidade utilizada foi o milhão de cruzeiros de 1973.

Pode-se observar, assim, analisando-se os dados do quadro 07 que, em geral, os setores considerados apresentam redução tanto no grau de concentração quanto no grau de "pobreza" (avaliado este pelo índice $1/(1 + a)$) entre 1973 e 1974 para, a partir daí, revelarem tendência ao aumento das desigualdades e da "po

PARÂMETROS DA CURVA LOG-LOGÍSTICA AJUSTADA POR SETOR E ANO

ANO	Setor Parâmetro	01	02	03	04	05	06	07	08	09	10
1973	a	0,0830	0,3072	0,3530	0,5473	0,5569	0,0994	0,0792	0,7420	0,2866	0,2050
	b	0,8310	0,8849	0,8849	0,7737	0,7526	0,9941	1,1004	0,8690	0,8414	0,9032
	r ²	0,9964	0,9926	0,9926	0,9930	0,9936	0,9853	0,9846	0,9857	0,9940	0,9796
1974	a	0,2298	0,6364	0,6500	0,9330	0,9885	0,2456	0,1675	1,2350	1,3246	0,3667
	b	1,0073	0,9585	0,9920	0,8329	0,8518	1,1522	1,2075	0,9011	0,7927	1,0455
	r ²	0,9892	0,9930	0,9880	0,9950	0,9900	0,9940	0,9882	0,9942	0,9745	0,9923
1975	a	0,2629	0,5602	0,6863	0,8128	0,8685	0,1959	0,2018	0,9871	0,6142	0,3845
	b	0,9887	0,9419	0,9446	0,8068	0,7779	1,1731	1,2175	0,8710	0,9975	1,1008
	r ²	0,9912	0,9924	0,9908	0,9918	0,9892	0,9944	0,9896	0,9924	0,9979	0,9878
1976	a	0,2606	0,5331	0,6701	0,8253	1,0313	0,2016	0,1997	1,1434	0,5849	0,4249
	b	0,9471	0,9226	0,8736	0,7706	0,8227	1,1437	1,2333	0,8480	0,9312	1,0416
	r ²	0,9894	0,9878	0,9859	0,9936	0,9825	0,9874	0,9851	0,9880	0,9827	0,9867
1977	a	0,3106	0,5548	0,6345	0,8658	0,9256	0,1944	0,1467	1,0496	0,6143	0,3870
	b	0,8353	0,8227	0,7666	0,7153	0,7139	1,0017	1,1066	0,7162	0,8182	0,8932
	r ²	0,9872	0,9914	0,9839	0,9912	0,9853	0,9879	0,9734	0,9862	0,9826	0,9749

cont.

Quadro 07 - (Cont.)

ANO	Setor Parâmetro	11	12	13	14	15	16	17	18	19	20	21
1973	a	0,5807	0,9680	0,2521	0,3089	0,3531	0,1322	0,1993	0,1019	0,5760	0,1845	0,1306
	b	0,7334	0,8126	0,7650	0,8941	0,9020	1,0467	0,8453	0,7403	0,6712	1,0427	0,9834
	r ²	0,9716	0,9716	0,9900	0,9867	0,9619	0,9851	0,9748	0,9793	0,9882	0,9966	0,9875
1974	a	1,1581	1,1440	0,4618	0,5214	0,6618	0,2779	0,5708	0,4759	0,9999	0,3488	0,2914
	b	0,7598	0,7979	0,8578	0,9692	0,9069	1,1679	0,9747	0,9509	0,7073	1,0738	1,0283
	r ²	0,9898	0,9910	0,9932	0,9886	0,9783	0,9964	0,9936	0,9964	0,9912	0,9970	0,9956
1975	a	1,1971	1,2678	0,5375	0,4986	0,6798	0,2260	0,6185	0,5362	0,9218	0,3139	0,2741
	b	0,7558	0,8611	0,8785	0,9628	0,8953	1,2040	0,9618	0,9759	0,6759	1,0953	1,0411
	r ²	0,9964	0,9861	0,9936	0,9864	0,9817	0,9922	0,9920	0,9918	0,9813	0,9960	0,9930
1976	a	1,1993	1,5453	0,5663	0,5729	0,8487	0,2942	0,6336	0,5145	0,9982	0,3206	0,3492
	b	0,7075	0,8880	0,8095	0,9391	0,8429	1,1637	0,9227	0,8669	0,7054	1,1697	1,0286
	r ²	0,9878	0,9882	0,9902	0,9890	0,9857	0,9956	0,9888	0,9888	0,9730	0,9948	0,9930
1977	a	1,1487	1,2853	0,6872	0,5807	0,7933	0,2511	0,7351	0,5979	0,9940	0,2563	0,3208
	b	0,6396	0,6831	0,8292	0,7968	0,8182	0,9882	0,8355	0,7491	0,6794	0,9006	0,8931
	r ²	0,9797	0,9920	0,9888	0,9830	0,9861	0,9854	0,9851	0,9815	0,9756	0,9843	0,9872

Obs.: Setores classificados segundo quadro 01. A fonte dos dados básicas é a publicação da SRF citada nas referências bibliográficas.

breza". Observa-se que, embora lenta, a retomada desta tendência é tal que já em 1977, para a quase totalidade dos setores, a situação observada é mais desigual do que a de 1973.

5.4. Preliminares dos Modelos Testados

A rigor, a análise de regressão que aqui se propõe, pretende verificar quais fatores econômicos conjunturais "explicam" as diferenças inter-setoriais e inter-temporais quanto ao grau de concentração e às variações deste grau de concentração ao longo do período enfocado. Desta forma, basicamente dois modelos foram testados. No primeiro, que será objeto de análise na seção seguinte, procura-se identificar quais as variáveis que desempenham papel significativo na evolução do índice de concentração dos setores. A variável dependente do estudo é, então, a variação do índice de concentração I , ano a ano, por setor. O modelo é, por conseguinte, "pooled". O segundo modelo, destacado na seção 5.6, tem como variável dependente o grau de concentração setorial, medido pelo índice I . Este tipo de análise, bem como aquela inerente ao primeiro modelo, tem sido feita através de estudo de "cross-section", o que, no caso de presente trabalho, foi julgado inadequado vez que assim teríamos grande limitação no número de graus de liberdade da regressão, já que o número de observações ficaria restrito ao total de setores. Foi adotado, então, um modelo "pooled" onde se procura "explicar" as diferenças inter-setoriais e inter-temporais quanto aos graus de concentração observados.

Em ambos os modelos, as variáveis independentes são aquelas evidenciadas na literatura econômica e mais as indicadas especificamente para o caso brasileiro. Nas seções seguintes,

anteriormente à apresentação e discussão dos modelos e dos resultados, o papel dessas variáveis é brevemente discutido.

5.5. Análise da Evolução do Grau de Concentração Setorial

As questões fundamentais associadas a mudanças no grau de concentração de uma determinada população entre dois pontos no tempo prendem-se às variações no tamanho da população (concentração ou desconcentração absolutas) e à diferenciação entre os indivíduos da população das taxas de crescimento do atributo no período. Quando o indicador utilizado para o grau de concentração é obtido através da abordagem probabilística do problema da distribuição de renda, faz-se necessário considerar a população suficientemente grande para tratar o atributo como contínuo. Portanto, o ajustamento de funções densidade ou de distribuição fornece índices de concentração não diretamente afetados por variações no tamanho da população. No caso de estudos de concentração industrial, entretanto, a eliminação ou a entrada de firmas em um mercado se dá prioritariamente na base da pirâmide e, logo, afetará indiretamente o cálculo do índice.

No âmbito deste estudo, o efeito (des)concentração absoluta não será tratado de forma direta. É evidente que esse efeito se registra como reação da estrutura industrial ao comportamento da conjuntura econômica, traduzido principalmente pela taxa de crescimento (atratividade) do mercado e pela existência maior ou menor de barreiras à entrada de novos competidores (ou de facilidades, como no caso de incentivos fiscais, subsídios ao crédito e outros instrumentos de política econômica). No caso brasileiro, e para o período considerado, o nível das barreiras à entra

da pode ser interpretado em associação à presença de empresas de capital estrangeiro ou de estatais no setor, por serem esses investimentos, em geral, voltados prioritariamente para os setores com grande intensidade de capital⁴ (setores de bens de produção) e/ou, por outro lado, com avançada tecnologia de produção (setores de bens de consumo duráveis).

Com base nessa reflexão e, também, nas conclusões do capítulo anterior, as variáveis consideradas neste trabalho para a investigação dos determinantes da evolução do grau de concentração setorial são: (1) a taxa de crescimento da receita setorial, (2) a presença no setor de empresas sob controle acionário de grupos estatais ou estrangeiros, (3) o departamento⁵ em que está incluído o setor, (4) a intensidade de capital no setor e (5) o crescimento da importação e exportação de produtos do setor.

Espera-se que os sinais das derivadas parciais da variável dependente com respeito à presença de multinacionais e estatais no setor sejam positivos, caracterizando o efeito concentrador desses investimentos o departamento onde está incluído o setor é definido com relação a atividade mais importante do setor, de acordo com o quadro 6, e da origem a variáveis "dummy" que buscam identificar efeitos específicos da presença de multinacionais e estatais em cada departamento.

⁴ O conceito de intensidade de capital é aqui entendido como a relação entre (capital + reservas) e receita total (ou faturamento).

⁵ A classificação origina-se em Marx, tendo sido ampliado por Kalecki. No âmbito deste estudo, a divisão vigente é a seguinte: Departamento I (bens de produção), Departamento II (bens de consumo não-duráveis) e Departamento III (bens de consumo duráveis).

Três formas funcionais foram testadas:

a) a forma log-linear, ou seja;

$$\ln Y = \sum_{i=1}^n \beta_i \ln X_i ,$$

b) a forma mista, ou seja,

$$Y = \sum_{i=1}^m \beta_i X_i + \sum_{i=m+1}^n \beta_i \ln X_i \quad e$$

c) a forma linear, ou seja,

$$Y = \sum_{i=1}^n \beta_i X_i .$$

A primeira delas não produziu quaisquer resultados empíricos relevantes; as duas seguintes, embora revelando um poder de "explicação" apenas razoável, produziram alguns resultados dignos de nota⁶, como se verifica no quadro 08.

⁶ Foram excluídas do quadro 08, as equações que apenas duplicavam informações já existentes nas outras equações.

As variáveis estão representadas no quadro 08 de acordo com a convenção seguinte:

(a) $Y_{i,t}$ é a variação no grau de concentração do setor i , entre os anos $t-1$ e t , medida por

$$Y_{i,t} = I_{i,t} / I_{i,t-1}$$

onde $I_{i,t} = 1/b_{i,t}$ é o índice de concentração obtido da curva log logística;

(b) $X_{i,t}^1$ é a variação na receita do setor i , entre os anos $t-1$ e t , medida por

$$X_{i,t}^1 = F_{i,t} / F_{i,t-1}$$

onde $F_{i,t}$ é o valor do faturamento agregado, no ano t , das empresas do setor i arroladas pelo "Quem é quem", deflacionado a cruzeros de 1973 pelo índice de preços do setor i (colunas 54 a 105 da "Conjuntura Econômica");

(c) $X_{i,t}^2$ e $X_{i,t}^3$ representam, respectivamente, as presenças das empresas estatais e multinacionais no setor i , ano t , medidas pelos percentuais do patrimônio líquido das

empresas do setor no "Quem é quem" devidos a empresas sob controle acionário de grupos estatais e de grupos estrangeiros.

$$(d) \quad X_{i,t}^5 = X_{i,t}^3 + 1 \quad \text{e} \quad X_{i,t}^4 = X_{i,t}^2 + 1;$$

(e) $X_{i,t}^6$ e $X_{i,t}^7$ representam, respectivamente os crescimentos das exportações e importações de produtos do setor i , entre $t-1$ e t , medidos pelos quocientes dos índices de preços correspondentes da Conjuntura (colunas 144 a 152 e 192 a 200);

(f) $X_{i,t}^8$ mede a relação entre receita média e (capital + reservas) médio no setor i , ano t , sendo os dados originários das publicações da SRF citadas entre as referências bibliográficas;

(g) $X_{i,t}^9 = n_{i,t} / n_{i,t-1}$, onde $n_{i,t}$ representa o número de empresas na amostra sobre a qual foi obtida a estimação da curva log-logística, para o setor i , ano t ;

(h) X_i^{10} é a relação lucros/salários vigente em 1974 para o setor i , sendo os dados originários de [06]

(i) D_i^1 e D_i^3 são variáveis "dummy" que assumem valor 1 caso i pertença, respectivamente, ao Departamento I ou III e valor zero, caso contrário.

Quadro 08

Principais Resultados Empíricos para a Evolução do Grau de Concentração Setorial
(Variável Dependente: $Y_{i,t}$)

Nº de Observações = 76

Equação nº	Constante	X^1	X^2	X^3	D^1X^3	D^3X^2	D^3X^3	D^1X^2	$\ln X^7$	X^6	$\ln X^6$	$\ln X^9$	X^8	X^9	D^3	D^1	X^{10}	R^2	F
01	1,2754	- 0,2252 (2,983) ^a																0,1073	8,8957 ^a
02	1,3791	- 0,2056 (2,998) ^a								-0,1198 (1,764) ^c								0,1438	6,1296 ^a
03	1,2704	- 0,2304 (2,998) ^a																0,1122	3,0343 ^b
04	1,3798	- 0,2021 (2,669) ^a																0,1491	4,2054 ^a
05	1,3700	- 0,1987 (2,605) ^b																0,1536	3,2205 ^b
06	1,3868	- 0,2088 (2,715) ^a																0,1657	2,7806 ^b
07	1,3416	- 0,2137																0,1791	2,5084 ^b
08	1,3471	- 0,2171 (2,802) ^a																0,1842	2,1930 ^b
09	1,3480	- 0,2166 (2,784) ^a																0,1900	1,9643 ^c
10	1,4143	- 2,2205 (2,684) ^a																0,2253	1,2670
11	1,2579	- 0,2029 (2,648) ^b																0,1437	2,9779 ^b

Obs.: Para todas as variáveis, à exceção de X^1 , X^7 e X^8 , os testes realizados foram unilaterais. Os valores entre parêntesis, são as estatísticas "t" de Student. As letras a, b e c indicam, respectivamente, significância estatística dos coeficientes a 1,5 e 10%.

Observa-se que a variável X^1 induz um efeito desacelerador no grau de concentração. Essa parece ser a conclusão mais importante a tirar dos dados do quadro 08, condizente inclusive com os resultados da pesquisa da FINEP sobre a evolução do grau de concentração no período 68-73. Aqui, dado a negatividade do coeficiente, pode-se verificar que altas taxas de crescimento tendem a conter a tendência à concentração enquanto que diminuições no ritmo de negócios do setor têm efeito favorável ao incremento do grau de concentração.

Uma segunda observação interessante é que os setores que mais cresceram em termos de exportação, apresentam evolução mais lenta no grau de concentração.

As outras variáveis, em geral, não apresentam coeficientes significantes. Mesmo assim, cumpre ressaltar a positividade da variável D^1 , com coeficientes significativos em três das seis regressões em que foi considerada. O fato indica constante maior para os setores do departamento I que nos demais, revelando uma tendência à evolução mais rápida do grau de concentração. Em contra-partida, a presença das multinacionais nestes setores (variável $D^1 X^3$) aparece com coeficiente negativo o que parece indicar que, mesmo nos setores de bens de produção, o capital de grupos estrangeiros é mais competitivo com os monopólios estatais que o capital nacional privado.

Não se revelou significativa para o crescimento do índice de concentração a presença de empresas sob controle acionário de grupos estrangeiros. Apesar disso, é interessante verificar, a nível de valores obtidos para os coeficientes, que a presença de empresas estrangeiras nos setores do Departamento III tem efeito suficientemente positivo para anular, inclusive o fato de os setores deste departamento apresentarem constante menor que

os demais (ver D^3 , na equação 10).

A presença de empresas estatais (variável X^2) tem efeito negativo na evolução da concentração apenas no departamento II, onde sua presença é insignificante. Nos demais setores, a empresa estatal influencia positivamente o aumento do grau de concentração, embora sem significância nos coeficientes.

Por fim, vale notar os valores relativamente altos para as estatísticas dos coeficientes da variável X^8 , inclusive com significância na equação 10. O fato indica que setores com maior intensidade de capital apresentam menor velocidade de concentração, o que pode ser atribuído à associação positiva entre intensidade de capital e desempenho do setor, detectada por Helson Braga [04].

As 76 (setenta e seis) observações disponíveis para a estimação das equações do quadro 08 referem-se aos setores 01 a 19 do quadro 06, tendo sido desconsiderados os setores 20 e 21 por não haver disponibilidade de dados para algumas das variáveis dos modelos testados nesta seção. A exclusão destes dois setores, apesar de significar a perda de oito graus de liberdade nas estimações, não acarretou problemas quanto aos aspectos mais propriamente técnicos dos modelos. Por outro lado, foi entendido como desnecessário o uso de correções para heterocedasticidade, dado que o exame dos resíduos não apresentou distorções consideráveis. Tampouco se justificaria, pela pequena variação de t ($t = 1, 2, 3, 4$), a hipótese de auto-correlação dos resíduos, ou seja, $E_{i,t} = \rho E_{i,t-1}$. O baixo grau de correlação entre as variáveis independentes permite assegurar o não-surgimento de problemas de multicolinearidade e, assim, pode-se considerar os modelos testados nesta seção, e também na seguinte, como enquadrados na definição de modelo econométrico básico.

5.6. Determinantes do Grau de Concentração Setorial

Nesta seção, objetiva-se pesquisar, através de análise de regressão, a influência de algumas variáveis na determinação do nível de concentração dos setores do quadro 06. Análises como esta têm, em geral, adotado modelos de "cross-section", considerando pouco importantes as flutuações do nível de concentração em pequenos intervalos de tempo. No âmbito deste trabalho, propõe-se, mais uma vez, um modelo "pooled", com informações da série temporal de 1974 a 1977 para cada setor. Dessa forma, além do número de observações (setenta e seis) ser bastante satisfatório, as mudanças temporais no nível de desigualdades de cada setor, mesmo que pequenas, são levadas em conta e os resultados do modelo são válidos para todo o período considerado.

A variável dependente do modelo desta seção é o valor do índice de concentração I derivado da curva log-logística, por setor e por ano. Em termos gerais, foram considerados como variáveis independentes aquelas referentes ao tamanho do setor, à presença de empresas multinacionais e estatais na estrutura do setor, à intensidade de capital no setor e à posição do setor face à classificação da indústria de transformação em departamentos.

Espera-se que o tamanho do setor apresente coeficiente positivo, caracterizando o padrão de investimentos oligopolizantes nos setores com mercados maiores. Por outro lado, as variáveis correspondentes à presença de investimentos estrangeiros ou estatais devem apresentar sinal positivo, face aos resultados do capítulo anterior. Quanto às variáveis "dummies" departamentais, tomando-se por base o departamento II, cujos setores, em média, são menos concentrados, de acordo com os índices usados no capítulo an

terior, a expectativa é de que apresentem efeitos positivos sobre o grau de concentração.

No quadro 09, apresentam-se os principais resultados empíricos obtidos com a análise de regressão. As variáveis independentes estão representadas de acordo com a seguinte convenção:

a) $X_{i,t}^1$ é o tamanho do setor i no ano t medido pelo faturamento do grupo de empresas do setor arroladas no "Quem é quem" no ano t e deflacionado pelo deflator implícito ao índice de preços por atacado do setor (colunas adequadas entre a 54 e a 105 da Conjuntura) ao bilhão de cruzeiros do primeiro semestre de 1969;

b) $X_{i,t}^2$ e $X_{i,t}^3$ como na seção 5.5 ;

c) $X_{i,t}^8$ como na seção 5.5 ;

d) D^1 e D^3 como na seção 5.5 e

e) $Z_{i,t}$ é o índice de concentração extraído da curva log-logística ajustada aos dados de receita total do setor i , no ano t , como em 5.3, ou seja,

$$Z_{i,t} = I_{i,t} = 1/b_{i,t} .$$

Principais Resultados Empíricos para o Nível de Concentração Setorial

(Variável Dependente: $Z_{i,t}$)

Nº de Observações = 76

Equação nº	Constante	D^1	D^3	$\ln(X^{2+1})$	$\ln(X^{3+1})$	X^{2+1}	X^{3+1}	X^1	$D^1 \ln(X^{2+1})$	$D^1 \ln(X^{3+1})$	$D^1(X^{3+1})$	$D^1(X^{2+1})$	$D^3(X^{3+1})$	X^8	R^2	F
01	0,9404			0,6282 (5,604) ^a	0,6081 (8,213) ^a									0,5390	42,679 ^a	
02	0,0289			0,5059 (5,700) ^a	0,4272 (8,102) ^a									0,5351	42,017 ^a	
03	0,9297			0,5019 (3,028) ^a	0,6073 (8,207) ^a	0,1113 (1,033)								0,5458	28,836 ^a	
04	0,9723			0,4934 (2,954) ^a	0,5985 (7,957) ^a	0,1150 (1,130)							-0,0158 (0,734)	0,5492	21,623 ^a	
05	1,1668	-0,3226 (4,750)	-0,1577 (2,566) ^a				0,5958 (4,039) ^a	1,2199 (0,625)	0,2547 (1,445) ^c	0,8668 (4,333) ^a				0,3980	7,603 ^a	
06	1,1668	-0,7832 (3,284) ^a	-2,3030 (1,258)								0,1711 (1,331) ^f	0,4692 (4,078) ^a	1,3443 (0,713)	0,3974	7,583 ^a	
07	0,34023					0,2482 (1,269)	0,4158 (6,399) ^a				-0,2249 (1,956) ^b	0,2365 (1,533) ^c	0,4181 (2,320) ^b	0,6113	18,083 ^a	
08	0,0614	0,3221 (0,421)	-1,1976 (0,724)			0,5124 (0,724)	0,4233 (6,213) ^a				-0,2522 (2,030) ^b	-0,0432 (0,063)	0,8319 (0,495)	0,6181	13,552 ^a	
09	0,0824	0,2878 (0,358)	-1,3658 (0,696)			0,4854 (0,666)	0,4216 (5,757) ^a	0,0213 (0,155)			-0,2559 (1,908) ^b	-0,0036 (0,000)	0,9813 (0,507)	0,6182	10,525 ^a	
10	0,9683	0,0346 (0,518)	-0,1397 (1,936) ^b	0,4221 (0,606)	0,6205 (5,822) ^a			-0,0153 (0,110)	0,1816 (0,230)	0,8916 (0,446)	-0,3753 (1,999) ^b	0,2614 (1,222)		0,6208	10,642 ^a	
11	0,9785	0,0303 (0,518)	-0,1344 (2,111) ^b	0,4526 (0,606)	0,6197 (6,261)			0,1432 (0,190)	0,7673 (0,440)	0,2471 (1,307) ^f			0,0044 (0,167)	0,6206	15,700 ^a	

Obs.: Para todas as variáveis, exceto X^8 , os testes realizados foram unilaterais. Os valores entre parêntesis são as estatísticas "t" de Student. As letras a, b e c indicam, respectivamente, significância estatística dos coeficientes a 1,5 e 10%.

Observam-se, no quadro 09, resultados altamente expressivos. A significância das regressões a 1% indica que a hipótese de nulidade do vetor de coeficientes é rejeitada, sendo portanto as equações bastante representativas.

Quanto às variáveis, nota-se a não-significância em todas as regressões em que foram envolvidas das variáveis tamanho do setor (X^1) e intensidade de capital (X^8). Ambas as variáveis apresentam trocas de sinais dos coeficientes em regressões que diferem somente na forma de medição das demais variáveis. Cabe salientar, entretanto, que nas regressões onde não são consideradas as "dummies" departamentais, o coeficiente de X^1 apresenta valor positivo, com a estatística "t" bastante próxima do ponto crítico a 10%, fato que deve ser creditado à correlação existente entre o departamento e o tamanho do setor⁷, principalmente porque os setores dos departamentos de bens de produção e de bens de consumo duráveis são, em média, maiores que os setores do departamento II. Na verdade, a variável X^1 apresenta correlações de até 0,789 com o grupo de variáveis $D^i \ln(X^Y + 1)$, de até 0,376 com o grupo de variáveis $D^i(X^Y + 1)$ e de 0,204 com a variável "dummy" D^1 .

Uma observação importante deve ser feita quanto ao sinal negativo do coeficiente das variáveis "dummies" D^3 e D^1 (nas equações 05 e 06), indicativos de uma constante menor nas equações dos departamentos I e III, com relação ao departamento II, tomado como base. O fato pode ser interpretado a partir da equação 05, por exemplo, como registrando uma tendência maior à competição equilibrada nos setores dos departamentos I e III dominados pela em

⁷ Sabe-se que a ausência, na regressão, de uma variável positivamente correlacionada com X_i , tende a superestimar o coeficiente de X_i , dependendo do grau de correlação entre a variável não-incluída e as demais variáveis do modelo. Ver Johnston (1972).

presa nacional privada, diminuindo este equilíbrio da competição pela presença mais significativa de empresas estatais, principalmente (valor elevado para o coeficiente de $D^1 \ln(X^2 + 1)$, com alta significância) no departamento I, ou empresas multinacionais, principalmente no departamento III (valor elevado para o coeficiente de $D^3 \ln(X^3 + 1)$, com alta significância).

Em geral, a presença de investimento estatal ou multinacional tem efeito positivo sobre o grau de concentração, como denotam os sinais positivos para os coeficientes das variáveis $\ln(X^2 + 1)$, $\ln(X^3 + 1)$, $X^2 + 1$ e $X^3 + 1$, apresentando significância a 1% na maioria das equações, especialmente nas quatro primeiras.

A partir da 5ª equação é possível verificar os efeitos destes investimentos nos setores de cada departamento, pela inclusão das variáveis "dummies". Observe-se, por exemplo, através das equações 10 e 11, que a presença de capital estrangeiro no departamento I tem efeito menor que nos demais (sinal negativo para o coeficiente da variável $D^1 \ln(X^3 + 1)$, com significância), caracterizando a competição entre o investimento de grupos estrangeiros e o investimento estatal nos setores de bens de produção, confirmando conclusão da seção anterior.

Como esperado, nos setores do departamento III, a presença de empresas multinacionais tem efeito concentrador mais importante que nos demais (sinais positivos para os coeficientes das variáveis $D^3 \ln(X^3 + 1)$ e $D^3(X^3 + 1)$, com significância). É interessante notar que a presença da empresa estatal no Departamento III revela efeito concentrador, apesar de os coeficientes positivos só se revelarem significativos na equação 07. O fato pode ser creditado, apesar do pouco investimento estatal nestes setores,

ã não-competitividade das empresas estatais na produção de bens de consumo duráveis.

Como na seção anterior, o exame dos resíduos das regressões não revelou distorções significativas das premissas do modelo linear básico e os erros podem ser considerados homocedásticos e não auto-correlacionados.

5.7. Considerações Gerais

Com bons resultados, a análise de regressão da última seção confirma as hipóteses comumente feitas quanto à concentração industrial brasileira. Apesar de o estudo envolver um período recente da economia nacional, o que torna difícil a tarefa de pesquisa empírica pela ausência de dados mais adequados, os graus de determinação e a significância das regressões do quadro 09 são bons avais para as diretrizes tomadas neste trabalho. Mesmo os resultados não tão bons do quadro 08 revelam-se importantes na medida em que mostram que os setores mais atingidos pela crise pós-73 apresentam uma evolução mais rápida do grau de concentração, caracterizando distribuição desigual dos efeitos da queda no nível de crescimento.

Por outro lado, os dados do quadro 07 mostram que, em geral, o uso da função log-logística em problemas de concentração industrial traz bons resultados. A rapidez e a facilidade da estimação, bem como a interpretação imediata dos seus parâmetros e os níveis de aderência aqui obtidos indicam a curva log-logística como um modelo bastante útil na retratação de situações distributivas de atributos ligados à estrutura industrial.

5.8. Recomendações para Futuras Pesquisas

A avaliação das mudanças do grau de concentração da estrutura industrial brasileira ao longo do tempo, bem como a investigação empírica dos mecanismos que orientam essas variações, deve ser uma preocupação constante da pesquisa econômica nacional. As repercussões de um maior ou menor grau de monopólio em determinado setor ou de uma variação brusca da estrutura vertical do setor em prol das firmas maiores são facilmente estendidas a toda a economia via distribuição funcional da renda, nível de falências e desempregos, nível de preços e inflação. A importância, portanto, de estudos nesta área fica bem caracterizada.

Neste contexto, a contribuição deste trabalho, especialmente por focar um período recente da economia nacional, deve ser vista como um esforço inicial no sentido da percepção do funcionamento da estrutura industrial brasileira em um período econômico particularmente crítico.

Como esforço inicial, este trabalho abre perspectivas a outras pesquisas. Por um lado, no que concerne à utilização do instrumental coadjuvante dos capítulos II e III, o autor entende a necessidade de uma investigação mais profunda nas relações entre os conceitos de concentração implícitos a cada um deles. Por outro lado, uma possibilidade interessante vincula-se a uma investigação econométrica similar à deste trabalho, a um nível de desagregação mais amplo da indústria de transformação, por exemplo, a classificação do quadro 01. O maior número de graus de liberdade assim obtido permitiria ao pesquisador uma opção metodológica não explorada neste trabalho, qual seja a aplicação de um mesmo modelo econométrico a períodos distintos de tempo, possibilitan

do, assim, a análise da variações dos efeitos das variáveis consi
deradas.

Outras possibilidades são a verificação das relações entre grau de concentração, preços e relação lucros/salários, a análise da concentração industrial do ponto de vista geográfico e outros temas ligados à questão da participação das peque
nas e médias empresas na estrutura industrial brasileira ou regional.

REFERÊNCIAS BIBLIOGRÁFICAS

01. AMOROSO, Luigi. Ricerche intorno alla curva dei redditi. *Annali di Matematica Pura ed Applicata*. Série 4-21, vol.II, 1925, pp. 123-157.
02. ANDERSON, T.W. & Stanley Clove. *An Introduction to the statistical analysis of data*. Houghton Mifflin Co., Boston, 1978.
03. BARLOW, R. & F. Prochan. *Mathematical theory of reliability*. Wiley, New York, 1965.
04. BRAGA, Helson C.. Determinantes do desempenho da indústria brasileira: Uma investigação econométrica. *Rev. Bras. de Economia*, vol.33, nº 4, out/dez 1979, pp. 501-570.
05. CHAMPERNOWNE, D.G.. A model of income distribution. *Economic Journal*, vol. 63, 1953, pp. 318-351.
06. CONSIDERA, Cláudio. Estrutura e evolução dos lucros e dos salários na indústria da transformação. *Pesquisa e Planejamento Econômico - IPEA*, vol. 10, nº 1, abr/1980, pp. 71-121.
07. COSTA, Ramonaval. Medidas de desigualdade de renda. *Boletim Geográfico*, vol.33, nº 238, fev/1974, pp. 45-72.
08. CRONIN, D.C.. A function for size distribution of income: a further comment. *Econometrica*, vol. 47, nº 3, mai/1979, pp. 773-774.

09. DAGUM, Camilo. El modelo log-logístico y la distribución del ingreso en la Argentina. El Trimestre Económico, vol 44 , nº 176, out/dez 1977, pp. 837-864.
10. DÖELLINGER, Carlos von & Leonardo Cavalcanti. Empresas multinacionais na indústria brasileira. IPEA/INPES, Rio de Janeiro, 1975.
11. EDITORA VISÃO LTDA. "Quem é quem na economia brasileira. Edições de 1974 a 1978.
12. ÉLTETÖ, Ö & E. Frigyes. New income inequality measures as efficient tools for causal analysis and planning. Econometrica, vol. 36, nº 2, abr/1968, pp. 383-396.
13. FISHLOW, Albert. Brazilian size distribution of income. American Economic Review, vol 62, mai/1972, pp. 391-402.
14. FISK, P.R.. The graduation of income distributions. Econometrica, vol. 29, 1961, pp. 171-185.
15. FUNDAÇÃO GETÚLIO VARGAS. Conjuntura econômica. mai/1979.
16. GASTWIRTH, Joseph L.. A general definition of the Lorenz curve. Econometrica, vol. 39, nº 6, nov/1971.
17. _____. Large sample theory of some measures of income inequality. Econometrica, vol 42, nº 1, jan/1974. pp.191-196.

18. _____ & J. Terry Smith. A new goodness-of-fit test. Proceedings of the American Statistical Association, 1972, pp. 320-322.
19. _____. The estimation of the Lorenz curve and Gini index. The Review of Economics and Statistics, vol. 54, ago./1972, pp. 306-316.
20. GIBRAT, R.. Les inegalités économiques. Sirey, Paris, 1931.
21. HOFFMAN, R.. Contribuição à análise da distribuição de renda e da posse de terra no Brasil. Tese de Livre-Docência não publicada, ESAQ, Piracicaba, 1971.
22. JOHNSTON, Jack. Econometric methods. MacGraw Hill, New York, 1972.
23. KAKWANI, Nanak & N. Podder. Efficient estimation of the Lorenz curve and associated inequality measures from grouped observations. Econometrica, vol. 44, n° 1, jan/1976, pp. 137-148.
24. KENDALL, Maurice & Alan Stuart. The advanced theory of statistics. vol. 1, 3^a ed., Charles Griffin & Co., London, 1969.
25. KONDOR, Yaakov. An old-new measure of inequality. Econometrica, vol. 39, n° 6, nov/1971.

26. LANGONI, C.G.. Distribuição de renda e desenvolvimento econômico do Brasil. Expressão e Cultura. Rio de Janeiro, 1973.
27. McDONALD, James B. & Michael Ransom. Functional forms, estimation techniques and the distribution of income. Econometrica, vol. 47, nº 6, nov/1979, pp. 1513-1525.
28. _____ & Bartell Jensen. An analysis of some properties of alternative measures of income inequality based on the gamma distribution function. Journal of the American Statistical Association, vol. 74, nº 368, dez/1979, pp. 856-860.
29. METCALF, C.E.. The size distribution of personal income during the business cycle. American Economic Review, vol. 59, 1969, pp. 657-669.
30. MORGAN, James. The anatomy of income distribution. Review of Economics and Statistics, vol. 44, ago/1962, pp. 270-282.
31. OLIVEIRA, Francisco de. A economia da dependência imperfeita Graal. Rio de Janeiro, 1977.
32. RÖDDER, Wilhelm. Tópicos especiais em manutenção de equipamentos. Notas de aula, UFSC, 1979.
33. SABÓIA, J.L.M.. Uma generalização da lei de Gibrat para o crescimento da firma. Pesquisa e Planejamento Econômico - IPEA, ago/1977.

34. _____ . Um novo índice de concentração: aplicação ao estudo da evolução recente da distribuição de renda no Brasil. Estudos Econômicos - FIEPE/USP, nº 5, 1978, pp.63-112.
35. SALEM, A. & T. Mount. A convenient descriptive model of income distribution: the gamma density. Econometrica, vol. 42, 1974, pp. 1115-1127.
36. SCHUTZ, R.R.. On the measurement of income inequality. American Economic Review, vol. 41, mar/1951, pp. 107-122.
37. SINGH S. & G. Maddala. A function for size distribution of incomes. Econometrica, vol. 44, nº 5, set/1976, pp.963-970.
38. TAVARES, Maria da C.. Acumulação industrial e industrialização no Brasil. Tese de Livre-Docência, UFRJ.
39. _____ & L.O. Façanha. A presença das grandes empresas na estrutura industrial brasileira. Revista da ANPEC, 1977, pp. 44-48.
40. THEIL, Henri. Economics of information theory. North - Holland, Amsterdam, 1967.
41. THURLOW, Lester C.. Analyzing the american income distribution. American Economic Review, vol. 60, 1970, pp.251-269.
42. TOLIPAN, R. & A.C. Tinelli (org). A controvérsia sobre distribuição de renda e desenvolvimento, Zahar, Rio de Janeiro, 1975.