

Decomposição funcional do índice de Gini com dados de renda do Brasil

José W. Rossi*

Neste artigo analisamos as propriedades da decomposição funcional do índice de Gini, aplicando-as, em seguida, a dados de renda tanto do censo de 1980 como relativos às declarações das pessoas físicas para o imposto de renda em 1979. Apesar de essas aplicações terem caráter sobretudo ilustrativo, os resultados obtidos permitiram constatar, entre outras coisas, o quanto, em média, é maior a parcela dos rendimentos do trabalho (nos rendimentos totais) das famílias de renda mais baixa *vis-à-vis* aquelas de renda mais elevada, e também o quanto maior é a fração dos rendimentos do tipo "aluguéis" e "distribuídos por pessoas jurídicas", por exemplo, para as famílias mais ricas quando comparadas com as mais pobres. Considerações são também apresentadas sobre como tendências porventura existentes na decomposição do índice de Gini podem ser úteis na análise do processo dinâmico do desenvolvimento econômico de um país.

1. Introdução; 2. Decomposição funcional do índice de Gini; 3. Aplicações; 4. Considerações finais.

1. Introdução

Se uma população for desagregada em vários subgrupos e o índice de desigualdade das rendas totais puder ser decomposto nos componentes "desigualdades dentro dos subgrupos" (geralmente uma média ponderada das desigualdades dentro dos subgrupos) e "desigualdade entre os subgrupos" (em geral alguma função das diferenças das rendas médias dos subgrupos), então diz-se que tal índice apresenta a propriedade da decomposição aditiva.¹ O índice de redundância de Theil e a variância logarítmica, por exemplo, possuem a propriedade da decomposição aditiva, mas não o índice de Gini.² De qualquer maneira, apesar dos atrativos óbvios da propriedade da decomposição, ela por si só não assegura a atratividade de um

*Da Coppe/UFRJ.

¹ Para discussões sobre a decomposição aditiva de um índice de desigualdade, ver Love & Wolfson (1976), Bourguignon (1979) e Shorrocks (1980).

² Para uma discussão detalhada sobre a decomposição do índice de Gini, ver Pyatt (1976).

índice de desigualdade: basta mencionar que a variância da renda e a variância logarítmica da renda são decomponíveis; todavia, enquanto a primeira destas medidas é sensível a mudanças na escala em que são medidas as rendas, na segunda temos a possibilidade de violação do princípio da transferência de Pigou-Dalton — o qual estabelece que a transferência de renda de uma pessoa rica para uma menos rica, sem que haja alteração no posto (*ranking*) da renda de uma com relação à da outra, deve contribuir para uma redução no índice de desigualdade.

O índice de Gini não apresenta as deficiências que acabamos de mencionar e se não é, por um lado, decomponível no sentido antes definido, é, por outro lado, dotado da propriedade de ser decomposto funcionalmente como a média ponderada dos chamados índices de concentração dos fatores componentes da renda total; esses índices de concentração guardam alguma semelhança com os índices de Gini dos fatores, mas não são sinônimos daqueles, conforme veremos adiante.

No item 2 discutimos a decomposição funcional do índice de Gini em detalhes, para no item 3 apresentarmos as aplicações a dados de renda no Brasil. No item 4 temos os comentários finais.

2. Decomposição funcional do índice de Gini³

A decomposição funcional do índice de Gini apresenta algumas peculiaridades conforme seja aplicada a dados individuais ou dados agregados. Por essa razão, apresentaremos tais decomposições separadamente.

2.1. Decomposição para dados individuais

Como ponto de partida, suponha-se que as rendas totais dos indivíduos, y , sejam desagregadas segundo os seus fatores componentes (por exemplo, rendas do trabalho, rendas do capital, etc.). Nestas circunstâncias, o índice de Gini para as rendas totais, $G(y)$, seria dado por:⁴

$$G(y) = \sum_{k=1}^m \gamma_k C_k \quad (1)$$

onde γ_k é a renda do fator k como proporção da renda total e C_k é o índice de concentração para o fator k com relação à renda total; esse índice de concentração é obtido a partir da curva de concentração da renda do fator, cujo conceito é apresentado a seguir.

³ Este item baseia-se, essencialmente, no trabalho de Pyatt, Chen & Fei (1980); adotaremos, aqui, a mesma notação daqueles autores para facilitar a exposição, já que nos referimos com frequência a certos resultados por eles derivados.

⁴ Esta decomposição é devida a Rao (1969), podendo ainda ser encontrada em Kakwani (1980) e Pyatt, Chen & Fei (1980).

Considere-se que os indivíduos sejam ordenados segundo os níveis crescentes das suas rendas totais. Chame-se $F(y)$ as percentagens acumuladas das pessoas com renda até o nível y e $F_1(y)$ as correspondentes percentagens acumuladas das rendas recebidas por tais pessoas. Como é sabido, da conexão dos pontos representados pelas coordenadas $F(y)$ e $F_1(y)$ tem-se a curva de Lorenz para a distribuição das rendas totais, y , da qual obtém-se o índice de Gini como um menos duas vezes a área sob a curva. Sejam agora as rendas totais desagregadas entre os seus fatores componentes. Se denominarmos $F_{1k}(y)$ como as percentagens acumuladas das rendas do fator componente k , correspondentes às mesmas percentagens acumuladas das pessoas $F(y)$ que acabamos de definir, então da conexão dos pontos $F(y)$ e $F_{1k}(y)$ temos a chamada curva de concentração para as rendas do fator k com relação às rendas totais, cujo índice de concentração é dado, de maneira semelhante ao índice de Gini, como um menos duas vezes a área sob a curva de concentração das rendas do fator k .

Com essas considerações, depreende-se que o índice de concentração para o fator k não é a mesma coisa que o índice de Gini para aquele fator.⁵ Enquanto o cálculo do índice de Gini para o fator k requer primeiramente uma ordenação das rendas desse fator, x_k , segundo os valores crescentes dessa variável, o índice de concentração para o fator k utiliza a ordenação das rendas estabelecida pelos níveis crescentes das rendas totais dos indivíduos (isto é, y). Existe, entretanto, uma relação formal entre essas medidas. Mais precisamente, chame $G(x_k)$ o índice de Gini para as rendas do fator k , e seja $C(x_k/y)$ o seu índice de concentração. Segue-se então que:

$$C(x_k/y) = R \cdot G(x_k) \quad (2)$$

onde

$$R = \text{Cov}[x_k, r(y)] \div \text{Cov}[x_k, r(x_k)] \quad (3)$$

com $\text{Cov}(a,b)$ sendo a covariância entre a e b ; x_k é o nível da renda do fator k ; e $r(x_k)$ e $r(y)$ são, respectivamente, os postos (*rank*) das rendas conforme estabelecidos por níveis crescentes das rendas do fator k (isto é, x_k) ou das rendas totais y .⁶

Note-se que R em (3) teria no máximo o valor de uma unidade, o argumento para isso sendo, nas próprias palavras de Pyatt, Chen & Fei (1980, p. 457-8):

⁵ De maneira genérica se x for a renda e $g(x)$ uma função qualquer da renda, temos que "a curva de concentração para $g(x)$ não é a mesma coisa que a curva de Lorenz para $g(x)$. Ambas só serão idênticas se $g(x)$ for estritamente monotônica e possuir derivada contínua $g'(x) > 0$ para todo x " (Kakwani, 1977, p. 721).

⁶ Para uma prova desse resultado ver Kakwani (1980, p. 173-4), ou ainda Pyatt, Chen & Fei (1980, p. 454-8).

“Para obter $G(x_k)$ as famílias são ordenadas segundo o valor de x_{ik} . Portanto, o ponto de ordem r na curva de concentração depende das famílias nas quais os x_{ik} forem os menores de todos. Uma ordenação distinta é utilizada na obtenção de $C(x_k/y)$, qual seja: a ordenação das famílias segundo o tamanho de y_i . A soma de x_{ik} não poderá ser menor para as primeiras r famílias de acordo com esta última ordenação. Portanto, a curva de concentração que produz $C(x_k/y)$ não poderá estar abaixo daquela que produz $G(x_k)$, e então $C(x_k/y)/G(x_k) = R \leq 1$.”

2.2 Decomposição para dados agregados

O índice de Gini para dados agregados será aqui denominado $G(\hat{y})$ para distingui-lo do índice para dados individuais ao qual já denominamos $G(y)$, no subitem anterior. Note-se que, agora, conhecemos apenas as rendas médias das famílias nas várias faixas de renda total (isto é, \hat{y}_i), mas não os valores individuais das rendas.

É sabido que $G(\hat{y})$ como estimado na prática, através da curva de Lorenz formada pela conexão em linha reta dos pontos cujas coordenadas são $F(\hat{y})$ e $F_1(\hat{y})$, subestima o verdadeiro valor do índice de Gini $G(y)$ que prevaleceria caso conheçêssemos os dados individuais. Esta questão fora extensivamente investigada por Gastwirth (1972), que estabelece limites inferiores e superiores para $G(y)$ a partir das informações de renda fornecidas pelos dados agrupados dessa variável. Conforme mostrado por Gastwirth (1972), a diferença entre essas duas medidas é em geral pequena, particularmente quando se dispõe de um número razoável (acima de 20) de intervalos de classe para a renda. De qualquer maneira, não nos ocuparemos com essa questão aqui. O interesse agora é estabelecer uma decomposição funcional exata para $G(\hat{y})$, qualquer que seja a sua subestimação do verdadeiro índice de Gini a prevalecer para os dados individuais da renda. Com esse objetivo, suponha-se inicialmente que os dados das rendas pessoais sejam agrupados segundo as várias faixas de renda total e que para cada uma dessas faixas tenhamos informações sobre o número de recipientes, as suas respectivas rendas totais, bem como a distribuição dessas últimas entre os diversos fatores que as compõem. Nestas circunstâncias, a renda média na faixa i com todos os tipos de renda seria, admitindo-se m fatores:

$$\hat{y}_i = \sum_{k=1}^m \bar{x}_{ik} \quad (4)$$

com \bar{x}_{ik} sendo o valor médio das rendas do fator k para todas as famílias situadas na faixa de renda total i . Vale dizer, \bar{x}_{ik} seria o montante das rendas com o fator k para as famílias na faixa de renda total i , dividido pelo total de famílias situadas naquela faixa, independentemente do fato de possuírem ou não renda com o fator.

Com essas considerações, pode ser demonstrado (por exemplo, Pyatt, Chen & Fei, 1980) que:

$$G(\hat{y}) = \sum_{k=1}^m \gamma_k C(\bar{x}_k/\hat{y}) \quad (5)$$

onde γ_k é a proporção, nas rendas totais, das rendas do fator k . Temos, pois, aqui, uma decomposição semelhante àquela apresentada na expressão em (1) no contexto de dados individuais.

Observe-se que $C(\bar{x}_k/\hat{y})$ é o índice de concentração para o fator k quando os dados são ordenados segundo os níveis da renda total e não deve, portanto, ser confundido com o índice de Gini para aquele fator, que requer a ordenação das rendas por níveis crescentes das rendas do fator k . Na verdade, Pyatt, Chen & Fei (1980) demonstram que:

$$C(\bar{x}_k/\hat{y}) \leq G(\hat{x}_k) \quad (6)$$

onde $G(\hat{x}_k)$ seria o índice de Gini para o fator k (isto é, quando os agrupamentos são estabelecidos de acordo com os valores de x_{ik}). Note-se que, enquanto o índice de Gini para o fator $G(\hat{x}_k)$ só poderia ser obtido conhecendo-se os dados individuais (pois para seu cálculo seria necessário primeiramente estabelecer os agrupamentos segundo os níveis das rendas do fator), o índice de concentração do fator k com relação às rendas totais poderá ser calculado, sem qualquer dificuldade, a partir dos dados agrupados segundo as rendas totais — isto é, enquanto a variável \bar{x}_{ik} poderá ser calculada dos agrupamentos de y_i , a variável \hat{x}_{ik} não o poderia. Portanto, parece claro agora por que o índice de concentração do fator não pode ser confundido com o índice de Gini do fator. Na realidade poderá haver diferença apreciável entre essas duas medidas. Se, por um lado, tal diferença não pode ser determinada, em vista da impossibilidade de se obter o índice de Gini para o fator (note-se que necessitaríamos primeiramente dos dados individuais para calcular esse índice), por outro lado, uma idéia da magnitude dessa diferença poderá ser estabelecida a partir de uma terceira medida, calculável com dados do tipo aqui considerado, a qual se situaria entre aquelas duas outras e que é denominada quase-fator Gini por Pyatt, Chen & Fei (1980). Vejamos, pois, no que consiste tal medida.

Conforme já ressaltado, o índice de concentração para o fator k é obtido da curva de concentração, que é definida a partir das coordenadas $F(\hat{y})$ e $F_{1k}(\hat{y})$. Isto é, as rendas médias do fator k na faixa i , \bar{x}_{ik} , são ordenadas de acordo com os níveis das rendas totais, y_i , e não por níveis de x_{ik} . Como o agrupamento estabelecido pelos níveis de y_i independe do nível do fator k , não existe nenhuma razão para que a ordenação de \bar{x}_{ik} seja a mesma daquela estabelecida para \hat{y}_i . Isto é, não há razão para que os valores médios das rendas com um dado fator cresçam monotonicamente quando nos deslocamos ao longo das várias faixas de renda definidas

de acordo com os níveis das rendas totais. Como o quase-fator Gini $G(\bar{x}_k)$ baseia-se na ordenação de \bar{x}_{ik} , segue-se então que:

$$C(\bar{x}_k/\bar{y}) \leq G(\bar{x}_k) \quad (7)$$

A diferença entre as medidas em (7) poderá ser esclarecida ainda mais com o seguinte argumento: "A curva de concentração que produz o quase-fator Gini $G(\bar{x}_k)$ ordena as famílias de acordo com $r(\bar{x}_{ik})$, isto é, de acordo com o nível médio de x_{ik} para grupos que resultam da estratificação das famílias por níveis da renda total. (. . .) isto implicará o mesmo agrupamento que $r(\bar{y}_i)$. A única diferença que poderá surgir entre $C(\bar{x}_k/\bar{y})$ e $G(\bar{x}_k) = C(\bar{x}_k/\bar{x}_k)$ é, portanto, na ordenação destes grupos na derivação da curva de concentração. A ordenação para $G(\bar{x}_k)$ é aquela que minimiza a ordenada da curva de concentração para cada valor da abscissa" (Pyatt, Chen & Fei, 1980 p. 463). Vale dizer, a curva que produz o quase-fator Gini jamais poderia estar acima da curva de concentração do fator, situando-se muito freqüentemente abaixo dela.

Quanto à comparação entre o quase-fator Gini, $G(\bar{x}_k)$, e o índice de Gini, $G(\hat{x}_k)$, temos que "ambos $G(\bar{x}_k)$ e $G(\hat{x}_k)$ são coeficientes de Gini calculados a partir de dados agrupados. A diferença entre eles reside então no critério de agrupamento. (. . .) Para \bar{x}_{ik} , o critério de agrupamento é o tamanho de y_i ; enquanto para \hat{x}_{ik} o critério é o tamanho de x_{ik} . Portanto, o critério de agrupamento para computar \hat{x}_{ik} é aquele que minimiza a ordenada da curva de concentração (isto é, a curva de Lorenz neste caso) para cada valor da abscissa. Portanto, a curva de concentração (Lorenz) que produz $G(\bar{x}_k)$ nunca poderá estar abaixo daquela que produz $G(\hat{x}_k)$ " (Pyatt, Chen & Fei, 1980, p. 463).

Resumindo, temos a seguinte relação entre as três medidas consideradas:

$$C(\bar{x}_k/\bar{y}) \leq G(\bar{x}_k) \leq G(\hat{x}_k) \quad (8)$$

Portanto, uma vez mais, e conforme salientado por Pyatt, Chen & Fei (1980), é incorreto interpretar o índice de concentração do fator como o seu índice de Gini, como efetuado por Fei, Ranis & Kwo (1978), pois, do exposto, depreende-se que poderá haver diferença apreciável entre essas duas medidas; aliás, o índice de concentração poderá até ser negativo, conforme explicado adiante.

É inapropriado, pois, considerar o índice de Gini total como uma média ponderada dos índices de Gini dos fatores que compõem a renda total. Neste sentido, Pyatt, Chen & Fei (1980) demonstram que o índice de Gini para dados individuais poderá ser decomposto como:

$$G(y) = \sum_{k=1}^m \gamma_k G(x_k) + \sum_{k=1}^m \gamma_k \epsilon_k + \epsilon \quad (9)$$

onde

$$\epsilon_k = C(\bar{x}_k/\bar{y}) - G(\bar{x}_k), \text{ para } k = 1, \dots, m \quad (10)$$

e

$$\epsilon = [G(\bar{y}) - G(\hat{y})] - \sum_{k=1}^m \gamma_k [G(x_k) - G(\bar{x}_k)] \quad (11)$$

enquanto para dados agregados a correspondente decomposição do índice seria

$$G(\hat{y}) = \sum_{k=1}^m \gamma_k G(\bar{x}_k) + \sum_{k=1}^m \gamma_k \epsilon_k \quad (12)$$

que não envolve o elemento ϵ dado em (9). Considerando-se agora o resultado em (8) temos que ϵ_k em (12) seria:

$$\epsilon_k = \epsilon_{1k} + \epsilon_{2k} \quad (13)$$

onde

$$\epsilon_{1k} = C(\bar{x}_k/\bar{y}) - G(\bar{x}_k) \leq 0 \quad (14)$$

e

$$\epsilon_{2k} = G(\bar{x}_k) - G(\hat{x}_k) \leq 0 \quad (15)$$

Essas relações dão bem uma idéia do erro em que se pode incorrer quando consideramos o índice de Gini como uma média ponderada dos índices de Gini dos fatores componentes. Com relação à expressão em (9), por exemplo, temos que enquanto a quantidade ϵ seria geralmente de pequena monta (ver Pyatt, Chen & Fei, 1980; ou Gastwirth, 1972), ϵ_k poderia ser uma fonte de erro apreciável. No que diz respeito à expressão em (12), por outro lado, temos que “a condição necessária e suficiente para $\epsilon_{1k} = 0$ é que o gráfico de $r(\hat{y}_i)$ versus $r(\bar{x}_{ik})$ seja monotônico não decrescente” (Pyatt, Chen & Fei, 1980, p. 463), ao passo que a condição necessária e suficiente para $\epsilon_{2k} = 0$ é que $r(\bar{x}_{ik}) = r(\hat{x}_{ik})$. Combinando esses resultados, temos que a condição necessária e suficiente para interpretar os índices de concentração $C(\bar{x}_k/\bar{y})$ como índices de Gini do tipo $G(\hat{x}_k)$ seria, de acordo com Pyatt, Chen & Fei (1980): necessária — o gráfico de $r(\hat{y}_i)$ versus $r(\hat{x}_{ik})$ seja monotônico não decrescente; suficiente — $r(\hat{y}_i) = r(\hat{x}_{ik})$ para todo i . Isto é, só em condições muito especiais o índice de Gini, para dados agregados, poderia ser decomposto como uma média ponderada dos índices de Gini dos fatores. As aplicações que se seguem ilustram muitos dos pontos que acabamos de expor.

3. Aplicações

Apresentaremos aqui duas aplicações. Na primeira, os dados utilizados provêm das informações das declarações de renda para o imposto de renda, das pessoas físicas, no Brasil, em 1979 (os dados mais recentes de que se dispõe). As declarações são agrupadas segundo 14 faixas de rendimentos brutos totais, havendo informações sobre as rendas recebidas e o total de declarações nas diversas classes de rendimentos. Adicionalmente, esses rendimentos totais são desagregados segundo as várias fontes de recebimento; mais precisamente, para cada faixa de rendimento bruto, temos a sua distribuição entre as diversas cédulas de declaração.⁷ Esses dados permitem, portanto, a aplicação apropriada da decomposição do índice de Gini mostrado em (5).

Os resultados da decomposição são apresentados na tabela 1.

Tabela 1

Fonte da renda (cédulas)	C_i = índice de concentração	γ_i = proporção dos rendimentos da cédula nos rendimentos totais	$\gamma_i C_i$	$\frac{\gamma_i C_i}{\sum \gamma_i C_i}$
A/B	0,6993	0,0063	0,0044	0,0095
C	0,4174	0,7883	0,3290	0,7034
D	0,6515	0,1304	0,0849	0,1816
E	0,7160	0,0544	0,0389	0,0832
F	0,8326	0,0069	0,0057	0,0123
G	0,3498	0,0081	0,0028	0,0061
H	0,3348	0,0056	0,0019	0,0040
Todas as cédulas	0,4674	1,0000	$\sum_i \gamma_i C_i = 0,4676$	1,0000

Note-se que o índice de Gini de 0,4674, para a distribuição de renda com todas as fontes (cédulas) consideradas (dado na última linha da segunda coluna) é reproduzido quase exatamente⁸ através da média ponderada dos índices de concentração dos vários fatores de rendimento (cédulas de rendimento) que é dada na última linha da quarta coluna (isto é, 0,4676).

Para uma interpretação apropriada dos valores dos vários índices de concentração obtidos, cabem primeiramente as seguintes observações. Note-se, do que foi exposto, que se o índice de concentração para um fator apresentar valor próximo daquele do índice de Gini para todas as rendas, então tal fato estaria indicando que os rendimentos médios daquele fator, nas várias classes (isto é, os rendimentos do fator na classe divididos pela população com todos os tipos de rendimentos na-

⁷ Esses dados representam um total de 7,3 milhões de declarações. Para detalhes, ver Ministério da Fazenda (1980).

⁸ A reprodução só não é exata devido a erros de arredondamento.

quela classe), guardam proporções aproximadamente fixas com as rendas médias totais; vale dizer, os mais pobres não teriam, nas suas rendas totais, uma proporção maior das rendas com o fator que aquela observada para os níveis de renda mais elevados. Por outro lado, se o índice de concentração para um fator apresentar valor bem superior ao do índice de Gini (como no caso da cédula *F*), isto seria indicação de que o fator apresenta, em média, uma participação maior nas rendas médias (totais) dos ricos, que aquela observada para os pobres; situação inversa ocorreria, naturalmente, quando o índice de concentração fosse menor que o índice de Gini. Finalmente, se um índice de concentração, para um fator, for próximo de zero, então os vários grupos populacionais (classificados de acordo com as rendas totais) estariam recebendo rendas médias daquele fator aproximadamente idênticas (este parece ser o caso dos rendimentos da cédula *G*). Ressalte-se, ainda, que, como mencionamos, um índice de concentração poderia até ser negativo; um resultado desses ocorreria se houvesse uma relação inversa entre as rendas médias do fator e as rendas médias totais, nas diversas classes de renda total.

Com essas considerações, dos resultados da tabela 1 temos que: a) os rendimentos do trabalho assalariado (cédula *C*) apresentam uma ligeira tendência a representar uma proporção maior das rendas totais nas faixas de rendas mais baixas; b) os rendimentos de capital (cédulas *A/B*), do trabalho sem vínculo empregatício (cédula *D*), de aluguéis e *royalties* (cédula *E*), e rendimentos distribuídos por pessoas jurídicas (cédula *F*), particularmente estes últimos, têm, em média, uma participação bem maior nos rendimentos médios dos ricos que aquela observada para as classes mais pobres; c) os rendimentos da agricultura, pecuária e outros (cédula *G*) têm uma participação bem maior nos rendimentos totais das classes mais pobres que nas classes ricas. Esses resultados parecem, em geral, estar de acordo com a expectativa.

É interessante ainda observar dos resultados da tabela 1, que, apesar de os rendimentos da cédula *C* apresentarem índice de concentração menor que o índice de Gini, a sua contribuição para este último é, em vista da sua elevada participação nos rendimentos totais, apreciável: cerca de 70%. Por outro lado, os rendimentos da cédula *F*, que têm índice de concentração bem superior ao índice de Gini, contribuem muito pouco para este, dada a pequena importância dos rendimentos da cédula quando comparados com os rendimentos totais.

Ainda com relação aos dados das declarações, como os rendimentos médios das várias cédulas nas diversas classes de renda total crescem, em geral, de maneira monotônica com as rendas médias totais nas classes, então os quase-fatores Gini das várias cédulas de rendimentos são idênticos aos seus respectivos índices de concentração; a única exceção ocorreria com os rendimentos da cédula *A/B*, cujo índice de concentração fora 0,6993 e o seu quase-fator Gini 0,7006. Ressalte-se, entretanto, que mesmo com o índice de concentração e quase-fator Gini idênticos, ainda assim podemos ter um índice de Gini para o fator sensivelmente superior àqueles; por exemplo, Pyatt, Chen & Fei (1980) obtiveram índice de concentração e quase-fator Gini no valor de 0,26 para as rendas do fator trabalho com dados de

Formosa, relativos ao ano de 1974, mas o índice de Gini para aquele fator fora de 0,44. Naturalmente, diferenças sensíveis podem também ocorrer entre o índice de concentração de um fator e o seu quase-fator Gini; por exemplo, numa análise semelhante à efetuada aqui, mas utilizando dados das declarações de renda das pessoas físicas em 1974, Rossi (1982) obteve índice de concentração para os rendimentos da cédula *G* de 0,04, mas um quase-fator Gini de 0,28 – como o índice de Gini para o fator seria no mínimo igual ao seu quase-fator Gini, poderíamos ter aqui uma diferença apreciável entre o índice de Gini e o índice de concentração. Este resultado serve para demonstrar que um elevado valor para o índice de Gini é uma condição necessária mas insuficiente para um elevado valor do índice de concentração do fator. Aliás, não é só com relação aos seus níveis que os índices podem diferir, tendências inversas entre o índice de Gini e o índice de concentração podem ser também observadas; este fora o caso obtido por Pyatt, Chen & Fei (1980) com dados de renda do trabalho e da agricultura em Formosa no período de 1964 a 1976 – fato que certamente permite interpretações interessantes sobre o processo de desenvolvimento econômico do país.

Para a segunda aplicação da decomposição funcional do índice de Gini, utilizaremos os dados de renda do censo de 1980, do IBGE, que fornecem informações sobre o número de pessoas economicamente ativas de 10 anos ou mais, ao longo de oito faixas de rendimento médio mensal. Esses dados são desagregados segundo quatro grupos mutuamente exclusivos, quais sejam: somente com rendimento do trabalho, com rendimento do trabalho e outras fontes, somente com rendimentos de outras fontes e, finalmente, aqueles sem rendimento. Com a exclusão deste último grupo, temos uma população economicamente ativa de cerca de 40,35 milhões de pessoas, as quais foram utilizadas para efeito de análise neste estudo.⁹ Para obter os montantes das rendas em cada faixa de renda, multiplicou-se o número de pessoas na faixa pelo ponto médio da faixa: tal procedimento só não foi adotado na primeira e última classes de renda: na primeira, cujo intervalo vai até 1/2 salário mínimo, adotou-se este limite superior como o valor médio da renda; na segunda, que é aberta à direita e tendo como limite inferior 20 salários mínimos, adotou-se como renda média o valor de 30 salários mínimos.

Em vista de os três grupos populacionais descritos serem mutuamente exclusivos (isto é, cada pessoa pertence somente a um grupo), então as faixas de renda dos fatores confundem-se, nesse caso, com as próprias faixas de renda total. Considerando ainda a simplificação adotada de que as pessoas de cada faixa recebem o ponto médio desta, fica claro que, nessas circunstâncias, $r(\hat{y}_j) = r(\bar{x}_{jk}) = r(\hat{x}_{jk})$ e, conseqüentemente, $C(\bar{x}_k/\hat{y}) = G(\bar{x}_k) = G(\hat{x}_k)$ – isto é, temos aqui a conjuntura daquelas condições muito especiais mencionadas no final do item 2.

⁹ Para detalhes adicionais sobre esses dados, ver IBGE (1980).

Após essas simplificações, obteve-se a decomposição funcional para o índice de Gini apresentada na tabela 2.

Tabela 2

Fonte de renda	$C_i =$ índice de concentração = G_i (índice de Gini)	$\gamma_i =$ proporção na renda total	$\gamma_i C_i = \gamma_i G_i$	$\frac{\gamma_i C_i}{\sum_i \gamma_i C_i} = \frac{\gamma_i G_i}{\sum_i \gamma_i G_i}$
Somente rendimento do trabalho	0,4398	0,5928	0,2596	0,4681
Rendimento do trabalho e outras fontes	0,7260	0,4059	0,2947	0,5313
Somente com rendimento de outras fontes	0,2290	0,0013	0,0003	0,0005
Todos os rendimentos	0,5500	1,0000	$\sum_i \gamma_i C_i = 0,5546$	1,0000

Note que o índice de Gini, para todas as rendas (0,5500), dado na última linha da segunda coluna, é aproximadamente igual à média ponderada dos índices de concentração (que são aqui idênticos aos índices de Gini dos fatores), dada na última linha da quarta coluna.

Observe-se que, de maneira similar ao caso dos dados das declarações de renda das pessoas físicas, temos aqui que os rendimentos com “trabalho somente” apresentam índice de concentração (0,44) menor que o índice de Gini total (0,55). Os rendimentos com “trabalho e outras fontes”, por outro lado, têm concentração bem maior que o índice de Gini total, enquanto os rendimentos com “outras fontes somente” apresentam concentração sensivelmente menor que o índice de Gini para todas as rendas. Naturalmente, a interpretação desses índices de concentração segue linha de raciocínio semelhante àquela adotada em relação aos dados das declarações de renda das pessoas físicas e será, pois, omitida, aqui.

Finalmente, na última coluna da tabela 2, mostramos as importâncias relativas dos índices desses fatores na composição do índice de Gini total.

4. Considerações finais

As aplicações da decomposição do índice de Gini, no item anterior, têm caráter sobretudo ilustrativo. Se, entretanto, dispuséssemos de séries históricas apropriadas, então tal decomposição poderia servir para investigar a própria dinâmica do processo de desenvolvimento econômico do país. A disponibilidade de dados de renda em bases individuais seria particularmente útil para tal propósito, já que permitiria calcular as tendências das razões das correlações por postos – dada em (3) – entre as rendas totais e as rendas dos fatores componentes e então determinar o quanto a ordenação (*ranking*) das famílias pelas rendas dos vários fatores estaria

determinando a ordenação dessas mesmas famílias pelas rendas totais, ao longo dos anos.¹⁰

Naturalmente, a própria tendência do índice de Gini de um fator componente poderia revelar aspectos importantes do processo de desenvolvimento econômico. Por exemplo, um índice de Gini para as rendas da agricultura, com tendência crescente, poderia ser o resultado do fato de que membros das famílias agrícolas situadas nas faixas de renda agrícola mais baixas se estariam transferindo para outros tipos de atividade com melhor remuneração.¹¹

Com os dados de que se dispõe no Brasil, entretanto, não se podem realizar análises desse tipo. À parte a dificuldade em se obterem amostras com dados individuais da renda, temos ainda que, enquanto os vários censos decenais apresentam agrupamentos de renda não comparáveis entre si, alterações na legislação fiscal têm ocasionado mudanças não desprezíveis no número das declarações apresentadas para o imposto de renda ao longo dos anos.¹² Por essas razões, o escopo desta análise foi limitado apenas aos aspectos discutidos.

Referências bibliográficas

- Bourguignon, F. Decomposable income inequality measures. *Econometrica*, 47 (4): 901-20, July 1979.
- Fei, J. C.; Ranis, G. & Kuo, S. W. Growth and the family distribution of income by factor components. *Quarterly Journal of Economics*, 92: 17-53, Feb. 1978.
- Gastwirth, J. L. The estimation of the Lorenz curve and Gini index. *The Review of Economics and Statistics*, 54 (3): 306-16, Aug. 1972.
- IBGE. *Tabulações avançadas do censo demográfico* 1980.
- Kakwani, N. C. Applications of Lorenz curves in economic analysis. *Econometrica*, 45 (3): 719-27, Apr. 1977.
- _____. *Income inequality and poverty*. Washington, D. C., World Bank, 1980.
- Love, R. & Wolfson, M. C. *Income inequality: statistical methodology and Canadian illustrations*. Ottawa, Statistics Canada, 1976.
- Ministério da Fazenda. *Anuário econômico-fiscal*. 1976.
- _____. *Anuário econômico-fiscal*. 1977.
- _____. *Imposto de renda pessoa física*. 1980.
- Pyatt, G. On the interpretation and disaggregation of Gini coefficients. *Economic Journal*, 86: 243-55, June 1976.
- _____; Chen, C. & Fei, J. The distribution of income by factor components. *The Quarterly Journal of Economics*, 94: 451-73, Nov. 1980.
- Rao, V. M. Two decompositions of concentration ratio. *Journal of the Royal Statistical Society*, 132: 418-25, part 3, series A, 1969.
- Rossi, J. W. *Índices de desigualdade de renda e medidas de concentração industrial; aplicação a casos brasileiros*. Rio de Janeiro, Zahar, 1982.
- Shorrocks, A. F. The class of additively decomposable inequality measures. *Econometrica*, 48 (3): 613-25, Apr. 1980.

¹⁰ Para uma aplicação com essa interpretação, ver Pyatt, Chen & Fei (1980).

¹¹ Sobre este ponto, ver Pyatt, Chen & Fei (1980).

¹² Por exemplo, enquanto em 1974 foram apresentadas 13,2 milhões de declarações, em 1975 esse número é reduzido para 4,9 milhões. Ver Ministério da Fazenda (1976; 1977).