

de Codes, Ana Luiza Machado

Working Paper

Métodos de estudo sobre pobreza: Abordagens tradicionais e a modelagem de equações estruturais

Texto para Discussão, No. 1345

Provided in Cooperation with:

Institute of Applied Economic Research (ipea), Brasília

Suggested Citation: de Codes, Ana Luiza Machado (2008) : Métodos de estudo sobre pobreza: Abordagens tradicionais e a modelagem de equações estruturais, Texto para Discussão, No. 1345, Instituto de Pesquisa Econômica Aplicada (IPEA), Brasília

This Version is available at:

<https://hdl.handle.net/10419/91142>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

TEXTO PARA DISCUSSÃO N° 1345

MÉTODOS DE ESTUDO SOBRE POBREZA: ABORDAGENS TRADICIONAIS E A MODELAGEM DE EQUAÇÕES ESTRUTURAIS

Ana Luiza Machado de Codes

Rio de Janeiro, julho de 2008

TEXTO PARA DISCUSSÃO N° 1345

MÉTODOS DE ESTUDO SOBRE POBREZA: ABORDAGENS TRADICIONAIS E A MODELAGEM DE EQUAÇÕES ESTRUTURAIS*

Ana Luiza Machado de Codes**

Rio de Janeiro, julho de 2008

* Agradeço a Inaiá Carvalho e a Robert Verhine, da Universidade Federal da Bahia, por me incentivarem a escrever este texto.

** Especialista em Políticas Públicas e Gestão Governamental, alocada na Diretoria de Estudos Sociais do Ipea.

Governo Federal

**Ministro de Estado Extraordinário de
Assuntos Estratégicos** – Roberto Mangabeira Unger

**Secretaria de Assuntos Estratégicos
da Presidência da República**

ipea Instituto de Pesquisa
Econômica Aplicada

Fundação pública vinculada à Secretaria de Assuntos Estratégicos da Presidência da República, o Ipea fornece suporte técnico e institucional às ações governamentais, possibilitando a formulação de inúmeras políticas públicas e programas de desenvolvimento brasileiro, e disponibiliza, para a sociedade, pesquisas e estudos realizados por seus técnicos.

Presidente

Marcio Pochmann

Diretor de Administração e Finanças

Fernando Ferreira

Diretor de Estudos Macroeconômicos

João Sicsú

Diretor de Estudos Sociais

Jorge Abrahão de Castro

Diretora de Estudos Regionais e Urbanos

Liana Maria da Frota Carleial

Diretor de Estudos Setoriais

Márcio Wohlers de Almeida

Diretor de Cooperação e Desenvolvimento

Mário Lisboa Theodoro

Chefe de Gabinete

Persio Marco Antonio Davison

Assessor-Chefe de Comunicação

Estanislau Maria de Freitas Júnior

URL: <http://www.ipea.gov.br>

Ouvidoria: <http://www.ipea.gov.br/ouvidoria>

ISSN 1415-4765

JEL: 132

TEXTO PARA DISCUSSÃO

Publicação cujo objetivo é divulgar resultados de estudos desenvolvidos pelo Ipea, os quais, por sua relevância, levam informações para profissionais especializados e estabelecem um espaço para sugestões.

As opiniões emitidas nesta publicação são de exclusiva e inteira responsabilidade do(s) autor(es), não exprimindo, necessariamente, o ponto de vista do Instituto de Pesquisa Econômica Aplicada ou da Secretaria de Assuntos Estratégicos da Presidência da República.

É permitida a reprodução deste texto e dos dados nele contidos, desde que citada a fonte. Reproduções para fins comerciais são proibidas.

SUMÁRIO

SINOPSE

ABSTRACT

1 INTRODUÇÃO	7
2 CARACTERÍSTICAS DESEJÁVEIS EM UM MÉTODO DE ABORDAGEM RELACIONAL DA POBREZA	9
3 METODOLOGIAS TRADICIONALMENTE UTILIZADAS NOS ESTUDOS SOBRE POBREZA	14
4 A MEE COMO MÉTODO PARA A ANÁLISE RELACIONAL DA POBREZA	47
5 CONSIDERAÇÕES FINAIS	65
REFERÊNCIAS	67

SINOPSE

Este trabalho se propõe a contribuir para o estudo da pobreza ao avaliar o potencial de métodos de pesquisa tradicionalmente utilizados e da Modelagem de Equações Estruturais (MEE) como instrumentos de análise “relacional” dessa questão social. Considerando-se que a evolução das discussões teóricas sobre a pobreza tem convergido para a noção de que se trata de um problema multidimensional e complexo, em que vários fatores socioeconômicos estão relacionados, reforçando-se mutuamente, coloca-se a necessidade de que tal complexidade seja refletida nas pesquisas empíricas sobre o tema. Dentro desta perspectiva, os principais métodos utilizados nos estudos sobre pobreza – ou seja, os índices sintéticos, os sistemas de indicadores sociais e as análises de regressão – mostram-se insuficientes para abordarem o fenômeno “relacionalmente”. A MEE, por sua vez, destaca-se em meio ao estado da arte das metodologias quantitativas de pesquisa como um método promissor para ajudar a pensar objetos sociais relacionalmente. Ela é considerada como a técnica mais avançada do ponto de vista da capacidade de articulação de variáveis sociais, uma vez que permite calcular simultaneamente todas as relações entre os fatores associados a um fenômeno. Neste trabalho, a contemplação das potencialidades analíticas oferecidas pela MEE enquanto método de análise relacional acontece em nível conceitual – em que se discutem também os potenciais dos outros métodos, de modo a se detectarem teoricamente as limitações e o alcance de cada um deles. Os resultados do estudo apontam que, de fato, a MEE apresenta-se como um instrumento mais propício à abordagem relacional da pobreza do que os métodos quantitativos tradicionais, uma vez que é capaz de refletir, na esfera das investigações empíricas, a complexidade do fenômeno, em afinidade com as formulações teóricas mais recentes.

ABSTRACT

The present research aims to contribute to the area of poverty studies by appraising the potentiality of traditional research methods and the Structural Equation Modeling (SEM) as proper instruments for analyzing poverty situations according to a “relational” perspective. The theoretical concepts of poverty have gradually developed towards the idea that it is a complex and multidimensional problem, which involves interrelated social and economic factors. Empirical research methods about the subject, however, do not appear to reflect such complexity. Under this point of view, one can say that the main methodologies usually supporting poverty studies – such as synthetic indices, social-indicator systems and regression analysis – are not satisfactory ways of treating the phenomenon “relationally”. Considering the state of the art of quantitative methodologies for social research, SEM is regarded as an innovative method that can provide helpful ways of approaching social phenomena relationally. SEM is seen as the most advanced technique in terms of its ability to articulate social variables, as it allows for a simultaneous estimation of the relationships among all factors associated with a phenomenon. This paper examines SEM’s analytical features as a relational method by using a conceptual approach, in which the traditional methods are also discussed, in order to theoretically appraise their limitations and potentialities as relational methodologies. The findings show that, indeed, SEM is a more powerful instrument for carrying out a relational analysis of poverty than the traditional quantitative methodologies, since unlike those, SEM is able to empirically reflect the articulate complexity of recent conceptual formulations.

1 INTRODUÇÃO

TEORIA E MÉTODO NO ESTUDO DA POBREZA: AS DUAS INSTÂNCIAS DE CONSTRUÇÃO DO OBJETO DE PESQUISA SOCIAL

Ao refletir sobre a natureza do trabalho de pesquisa social, Bourdieu (1998, p. 23-24) coloca a importância de que todas as instâncias de uma investigação sejam pensadas como componentes interligados de um conjunto. Ele chama a atenção para o fato de que os elementos “teoria” e “metodologia” não devem ser tomados como se fossem duas instâncias separadas, mas afirma que, dentro do trabalho de pesquisa, “as opções técnicas mais ‘empíricas’ são inseparáveis das opções mais ‘teóricas’ de construção do objecto” (BOURDIEU, 1998, p. 24).

No campo de estudos sobre pobreza, as discussões teóricas mais recentes tendem a reconhecer a complexidade e a multidimensionalidade desse fenômeno que envolve vários fatores que se articulam e se influenciam mutuamente. A questão da multidimensionalidade da pobreza é preponderante no debate atual, constituindo-se no foco de discussão de diversos autores (SALAMA; DESTREMAU, 2001; WHITE; KILLICK, 2001; ROCHA, 2003; SCHWARTZMAN, 2004, entre outros).

Ante a perspectiva apresentada por Bourdieu, é relevante que as pesquisas desenvolvidas sobre a pobreza disponham de procedimentos metodológicos capazes de refletir, na esfera das investigações empíricas, a complexidade que costuma marcar as discussões teórico-conceituais. Diversos esforços vêm sendo empreendidos nessa direção. São análises que se caracterizam por levarem em conta vários fatores, tais como renda, acesso a bens duráveis e a saneamento básico, educação, saúde, posição na ocupação no mercado de trabalho etc.

Contudo, a maioria dessas investigações não chega a refletir, de modo satisfatório, a natureza complexa e multidimensional do problema, caracterizada pela articulação dos diversos fatores associados ao fenômeno. Isso se deve ao fato de que aqueles trabalhos estão pautados em técnicas de abordagem que possuem potencial limitado para captar a referida articulação. Tal limitação pode ser notada, por exemplo, na utilização de indicadores sintéticos como modo de analisar a pobreza. Ao sintetizarem a informação social, tais indicadores exprimem um fenômeno multidimensional através de um número ou coeficiente, incorrendo assim em perda de transparência sobre a complexidade do processo. Com relação ao procedimento de cruzar distintos indicadores sociais, observa-se que só é possível articular, no máximo, indicadores relativos a três dimensões do problema. Essa restrição está atrelada ao fato de que esse tipo de cruzamento só se faz inteligível até os limites da exposição dos resultados numa tabela de três entradas. As análises de regressão, por sua vez, atêm-se à abordagem das relações entre os fatores associados ao fenômeno e as medidas de pobreza, não contemplando o modo como ocorre a articulação entre aqueles fatores.

Nesse sentido, uma técnica denominada Modelagem de Equações Estruturais (MEE) destaca-se como um método mais avançado do ponto de vista da sua capacidade de articulação de variáveis sociais (BOOMSMA, 2000; EDWARDS; BAGOZZI, 2000; KLEM, 2000; MCDONALD; RINGO HO, 2002; VINOKUR; SCHUL, 1997, 2002). Em linhas gerais, essa técnica de análise quantitativa multivariada consiste em testar a plausibilidade de um modelo construído com base

em uma teoria que sustenta o fenômeno estudado. Como tal metodologia caracteriza-se por permitir que diversas variáveis sejam correlacionadas simultaneamente, de modo que as múltiplas relações possíveis sejam apreciadas, ela coloca-se como alternativa promissora, capaz de delinear como se articulam os conceitos sociais envolvidos no fenômeno da pobreza. O método pode propiciar uma visão da forma como os fatores associados à pobreza se interligam, indicando quais são os seus impactos sobre as condições de vida dos indivíduos pobres, tanto direta quanto indiretamente, e ponderando a relevância de cada uma das inter-relações sob análise (KLINE, 1998; BENTLER; WU, 2002).

Frente ao exposto, o objetivo deste trabalho é analisar os métodos tradicionalmente utilizados nos estudos sobre pobreza – quais sejam, os indicadores sintéticos, os sistemas de indicadores sociais e as análises de regressão –, no que se refere às suas capacidades de captarem e refletirem a complexidade do fenômeno, tal como colocam as formulações teóricas recentes. Além disso, o presente texto se propõe a contribuir para a área de estudos sobre pobreza, ao apresentar a MEE e discutir o seu potencial como instrumento capaz de refletir, na esfera das investigações empíricas, o caráter complexo e multidimensional dessa questão social. A intenção é que as principais características desse método sejam explanadas de modo simples e acessível aos pesquisadores sociais, em vez de se desenvolver uma abordagem que privilegie os detalhes técnicos de sua utilização. Com isso, espera-se contribuir para a compreensão das potencialidades desse procedimento relativamente novo e ainda pouco utilizado nos estudos sociais no Brasil e, conseqüentemente, para o avanço no uso de métodos quantitativos mais sofisticados na área.

Como a necessidade de reduzir a pobreza representa hoje um consenso nos meios acadêmicos, governamentais, sociais e na comunidade internacional, estudos que ajudem a fornecer uma melhor compreensão do fenômeno justificam-se pela relevância que esse tema assume nas sociedades contemporâneas. Em termos mais específicos, a grave situação de pobreza em que vive uma parcela significativa da população no Brasil requer a realização de estudos que forneçam substratos para intervenções sociais. Neste sentido, este trabalho constitui uma contribuição aos esforços de análise do fenômeno da pobreza, oferecendo elementos adicionais para uma compreensão “relacional” desse processo social, no qual fatores socioeconômicos articulam-se entre si e influem, de modo conjugado, na tendência de um indivíduo viver em condições de carências e privações.

Para discutir a potencialidade de cada um dos métodos aqui abordados, este trabalho se estrutura em quatro seções, além desta introdução. Inicialmente, apresenta-se uma revisão da literatura que discute métodos de pesquisa sobre pobreza. Com base nessa revisão, evidenciam-se algumas características tidas como “desejáveis” para se abordar a questão dentro de uma perspectiva relacional de análise. Na seção seguinte, as metodologias tradicionalmente utilizadas nos estudos sobre pobreza são descritas e analisadas. Discute-se em que medida cada uma delas apresenta os referidos atributos desejáveis, revelando-se assim as potencialidades analíticas e limitações de cada método. Em seguida, é colocada uma visão geral da MEE, em que se expõem seus elementos constitutivos básicos, além de alguns pontos mais específicos, tais como as partes componentes de um modelo, aspectos relativos à aferição do grau de ajuste aos dados e a forma como a MEE lida com os erros de

mensuração das variáveis. A partir disso, examina-se a correspondência que a MEE apresenta em relação àquelas características. Por fim, são retomados os principais pontos da discussão sobre os potenciais e limitações de cada método em refletir o caráter relacional da pobreza.

2 CARACTERÍSTICAS DESEJÁVEIS EM UM MÉTODO DE ABORDAGEM RELACIONAL DA POBREZA

As formulações teóricas recentes sobre a pobreza convergem para a noção de que se trata de um fenômeno multidimensional e complexo. Os diversos fatores associados ao fenômeno manifestam-se simultaneamente, articulando-se entre si e influenciando nas condições de vida dos indivíduos. Essa perspectiva aproxima-se da noção de “campo”, proposta por Bourdieu (1998, p. 27-28), segundo a qual as realidades sociais são vistas como “espaços de relações”. Isso significa que o objeto de pesquisa social não está isolado, mas situa-se num conjunto de relações que lhe afetam e que condicionam suas propriedades. Sob esse enfoque, é desejável que a construção do objeto, em uma pesquisa social, se dê conforme tal concepção relacional, tanto em seus aspectos teóricos como metodológicos.

As formulações teóricas recentes sobre a questão da pobreza encontram ressonância nessa perspectiva de que as realidades sociais são de natureza “relacional”. Tendo em vista a necessidade de que as escolhas metodológicas sejam capazes de refletir, na etapa de análise dos dados empíricos, a construção teórica do objeto, faz-se necessário pontuar quais atributos conferem a uma metodologia a capacidade de analisar a pobreza de acordo com tal enfoque relacional. Esses atributos devem permitir a apreensão de como se articulam os fatores socioeconômicos envolvidos no fenômeno da pobreza – tais como a estabilidade laboral, a idade e a qualificação do trabalhador, o equilíbrio da família e seu patrimônio acumulado –, conforme postulado pela teoria adotada (FITOUSSI; ROSANVALLON, 1998, p. 87).

A adequação entre as formulações teóricas sobre pobreza e a forma de fazê-las transitar para a esfera empírica, por meio da aplicação de uma metodologia de pesquisa, é alvo de discussão de diversos trabalhos.¹ A maioria dessas discussões aponta as dificuldades metodológicas de se abordar a natureza complexa desse fenômeno social. As dificuldades derivam da existência de uma multiplicidade de conceitos, definições e classificações, assim como de uma pluralidade de variáveis que podem ser utilizadas para mensurar certos aspectos sociais. A análise da pobreza complica-se também pelo fato de que a grande quantidade de fatores envolvidos no fenômeno se inter-relacionam e sofrem constantes variações (SLIWIANY, 1997, p. 19).

Para exemplificar a complexidade a ser enfrentada nos esforços de investigação empírica sobre a pobreza, basta lembrar que as pesquisas costumam envolver uma série de temas, tais como características da população, emprego e remuneração do trabalho, proteção ao trabalhador, composições familiares, educação e treinamento de mão-de-obra, saúde, nutrição, saneamento básico e habitação. O aprofundamento de

1. A esse respeito ver, por exemplo, Pnud (1997), Salama e Valier (1997), Schwartzman (1997), Sliwiany (1997), Betti, D'Agostino e Neri (2000) e Salama e Destremau (2001).

cada um desses temas, por sua vez, ocorre por meio de seus desdobramentos em diversos “subtemas”. Isso aumenta ainda mais o número de conceitos e variáveis a serem considerados e, conseqüentemente, incrementa a dificuldade de manipulá-los articuladamente e de forma inteligível. Todos esses aspectos também abrem espaço para a existência de ambigüidades e de um alto grau de variabilidade entre os resultados de distintas pesquisas que buscam conhecer a população e suas condições de vida (SLIWIANY, 1997, p. 14).

Apesar de todas essas dificuldades, é inquestionável a necessidade de se desenvolverem investigações baseadas em métodos estatísticos. Dentro desse contexto, a demanda por ferramentas mais refinadas, que possam produzir “melhores” informações sobre aquelas realidades sociais, tem efetivamente guiado as inovações metodológicas para o estudo desse tema. Nesse sentido, são de grande valor os avanços que têm sido alcançados na área dos métodos quantitativos de pesquisa social e dos *softwares* de modelagem estatística (McGEE; BROCK, 2001, p. 4).

O debate atual acerca dos métodos de investigação sobre a pobreza converge para a noção de que é necessário ir além dos esforços de distinção entre pobres e não-pobres. O empenho de se identificar a população pobre é tido como um estágio inicial dos trabalhos de pesquisa, que devem prosseguir em direção à descrição e à análise daquela população. Tais etapas de descrição e análise são enaltecidas como o cerne das pesquisas sociais, uma vez que elas permitem o estudo das circunstâncias que podem afetar a probabilidade de as pessoas viverem em situações de carências e privações. Com isso, abre-se espaço para que se tenha acesso ao conhecimento de determinadas feições do fenômeno, como a estimação da distribuição dos graus de pobreza entre os desfavorecidos e a realização de comparações entre as condições de vida entre os próprios pobres, o que leva em consideração a heterogeneidade existente dentro dessa parcela da população. Esse é o ponto de vista defendido por autores como Betti, D’Agostino e Neri (2000, p. 2), que afirmam a necessidade de se refletir, nas investigações empíricas, o fato de que o bem-estar relativo de indivíduos e famílias é uma questão de grau, sendo também relevante o estudo dos fatores que influenciam esse aspecto.

Como se disse, tendo em vista a necessidade de que a construção teórica da pobreza enquanto “espaço de relações” transite para a etapa de análise empírica, é desejável que a metodologia a ser utilizada espelhe as principais propriedades do fenômeno. De modo sintético, Sliwiany (1997, p. 32) afirma que para analisar a pobreza em sintonia com as formulações teóricas recentes são necessários métodos capazes de prover modelos que ofereçam uma representação gráfica daquele fenômeno social. Esses modelos devem consistir num conjunto de equações capazes de designar as relações básicas entre os diferentes aspectos envolvidos no problema, exprimindo a configuração das “linhas de força” que evidenciam a forma como se articulam os diversos fatores.

Tendo em vista tais considerações, cabe alertar que nenhum modelo construído com dados empíricos conseguirá dar conta de todas as ligações existentes na realidade. Não se é de esperar que a complexidade de tal fenômeno seja esgotada em um modelo empírico, cuja construção, muitas vezes, encontra-se condicionada à disponibilidade de dados estatísticos. Segundo Sliwiany (1997, p. 32), a idéia é que sejam priorizadas

apenas as análises das relações constantes, relevantes e não casuais entre os fatores envolvidos, as quais já se constituem em grandes contribuições para o conhecimento das realidades sociais de pobreza. Note-se que, mesmo tendo em mente o fato de que os modelos empíricos devem abranger os fatores associados à pobreza com parcimônia, é importante que se analise uma multiplicidade de aspectos associados à condição de baixa renda. Uma vez que as questões envolvidas na problemática da pobreza sobrepõem-se, estudos que desconsiderem essa complexidade correm o risco de chegar a resultados supersimplificados sobre a natureza do fenômeno.

Como se sabe, na maioria das vezes os fatores que afetam as condições de vida dos indivíduos estão correlacionados entre si, a exemplo do que ocorre no caso de educação, raça e tipo de inserção no mercado de trabalho. Por isso, Székely (1996, p. 14) chama a atenção para o fato de ser desejável que se conheçam não apenas os efeitos diretos que cada um dos fatores exerce sobre as condições de vida,² mas também os seus efeitos indiretos, que ocorrem por meio da correlação com outros fatores. Nesse sentido, procedimentos metodológicos que estabeleçam associações simultâneas entre as medidas de pobreza – tais como renda, acesso a serviços públicos e a bens duráveis – e indicadores referentes a outros fatores relacionados ao fenômeno são bem-vindos. Tais procedimentos propiciam o aumento das possibilidades de explicar a pobreza ou, mais especificamente, de explicar a variabilidade da renda e de outros indicadores das condições de vida da parcela pobre da população (SZÉKELY, 1996, p. 34).

Uma vez estabelecidas as referidas associações, a intenção é que se possam determinar quais fatores ou variáveis são mais relevantes em relação aos demais, em um dado contexto de pobreza. A idéia é que se conheçam quais deles exercem maior influência ou têm maior peso sobre as condições de vida dos indivíduos, tanto direta como indiretamente. A importância dessa abordagem reside em que, com ela, abre-se a possibilidade da compreensão de quais grupos, assumindo certas características, estão mais sujeitos a vivenciarem situações de pobreza. Deve-se notar ainda que esse tipo de análise tem a potencialidade de evidenciar situações em que alguns fatores mostram-se influentes em vários setores da população, enquanto outros podem se revelar interferentes apenas no que tange a grupos específicos (SZÉKELY, 1996, p. 34).

Por fornecer informações sobre os fatores que aumentam a probabilidade do indivíduo viver em pobreza e sobre os pesos relativos entre os diversos aspectos relacionados ao fenômeno, esse tipo de análise pode ajudar a lidar com o problema, servindo de subsídio à elaboração e à avaliação de políticas públicas. Székely (1996, p. 13-14) comenta que, sob o ponto de vista da formulação de políticas, é preferível que se identifiquem um ou alguns fatores que expliquem uma parcela significativa do fenômeno do que um rol excessivo de variáveis. A valorização desse tipo de resultado parcimonioso advém do fato de que é mais factível intervir sobre os aspectos proeminentes em uma realidade social, que ter de atuar sobre uma extensa lista de variáveis indicadas por certos estudos e que, muitas vezes, são apresentadas de modo desconexo e fragmentado. Sem uma visão conjunta de como se articulam os fatores associados à pobreza, as possibilidades de atuação sobre a realidade tornam-se

2. Ou, no jargão familiar à estatística, a “capacidade de explicação” daquele fator ou característica da população sobre a variação das condições de vida.

estreitas, podendo mesmo dificultar a atuação junto a certos subgrupos populacionais específicos.

Todos os aspectos ora mencionados mostram que, para que uma pesquisa empírica possa apreender uma situação de pobreza, é importante utilizar uma técnica correspondente à complexidade do fenômeno. Frente a tal constatação, sem se referir a uma metodologia específica, Wilber (1975, p. 6) sugere que se estabeleça uma “estrutura de procedimento” para que se possa desenvolver esse tipo de investigação. Em princípio, o autor enfatiza que o estudo deve partir de um sistema de proposições teóricas inter-relacionadas, referentes às dimensões associadas ao problema, das quais podem ser deduzidas hipóteses testáveis. É desejável que as relações sejam estabelecidas de forma clara e que os conceitos discutidos na teoria sejam trazidos à abordagem empírica de maneira apropriada, mediante a utilização de variáveis passíveis de mensuração. Em seguida, a idéia é que não apenas as pressuposições postuladas na teoria sejam testadas, mas também as chamadas “hipóteses rivais”.³ A partir disso, algumas tentativas de conclusões devem ser delineadas, abrindo-se espaço para que as formulações iniciais sejam revistas, caso necessário. Os testes de hipóteses devem então continuar sucessivamente, até que se chegue a um resultado considerado satisfatório.

De modo mais detalhado, pode-se dizer que o procedimento sugerido por Wilber (1975, p. 166) consiste em quatro etapas. Inicialmente, deve-se formular um diagrama esquemático baseado na teoria, constituído pelo agrupamento de fatores e variáveis que, supostamente, influenciam a variável dependente. Em seguida, deve-se proceder a um “estreitamento de foco”. Tal estreitamento refere-se à já comentada impossibilidade de se incluírem em um só estudo todos os fatores e variáveis apontados na teoria e que, de acordo com o procedimento ora descrito, seriam levados a constar do diagrama esquemático. A decisão sobre quais aspectos devem ser mantidos no diagrama esquemático deve se basear em razões substantivas, lastreadas na consideração de normas sociais ou na potencialidade de cada aspecto servir de base para intervenções de governo (WILBER, 1975, p. 167).

A terceira etapa compõe-se de formulações e testes de modelos alternativos, podendo ser vista como um desdobramento do estágio anterior. Partindo do pressuposto de que o potencial de inclusão de variáveis em um modelo é limitado, emerge a necessidade de elaboração e de teste de composições alternativas, que compreendam aspectos não inseridos em outras versões. Para fazer a escolha entre elas, deve-se questionar a habilidade relativa de uma representação específica para explicar a variabilidade do fenômeno. Somente por meio do teste de versões alternativas pode-se ter segurança sobre um modelo particular. Por fim, deve-se partir para a revisão do diagrama esquemático; é esperado que, como consequência dos

3. Em um procedimento como o sugerido por Wilber (1975), as hipóteses principais do estudo costumam ser estabelecidas a partir do arcabouço teórico que o ilumina. São chamadas “hipóteses rivais” aquelas que não coincidem com as suposições inicialmente declaradas, podendo diferir pela inclusão de aspectos inexistentes naquelas proposições originais, pela supressão de alguma de suas partes ou mesmo pelo estabelecimento de um cenário inconsistente em relação àquele delineado inicialmente. Tais divergências podem emergir de outras teorias sobre o assunto estudado, de conhecimentos e inspirações vindas de áreas distintas ou ainda de outras indicações que podem ser consideradas relevantes por parte do pesquisador.

testes feitos com várias composições alternativas, o esquema original sofra modificações (WILBER, 1975, p. 168-169).

Nesse ponto, é necessário comentar que, mesmo com a prescrição de uma estrutura de procedimento, ainda há dificuldades a serem enfrentadas nas pesquisas sobre pobreza. Dentre as dificuldades, podem-se mencionar as ligações do fenômeno com outros processos sociais abstratos, também chamados de “latentes”, que não são diretamente mensuráveis por intermédio de variáveis observáveis. Pode ocorrer também que as bases de dados disponíveis não forneçam a mensuração de certas variáveis pertinentes ao estudo. Nesses casos, estariam excluídos do modelo alguns fatores mencionados pela teoria. Além disso, não se pode descartar a possibilidade de que haja fatores associados às realidades de pobreza que nem mesmo são considerados no arcabouço teórico.

Há ainda um outro aspecto a ser notado, que diz respeito à dificuldade de lidar com erros de mensuração. Apesar de todos os cuidados que podem ser tomados, sabe-se que o processo de medição de variáveis sociais é árduo e propenso a imprecisões, o que torna praticamente inevitável a existência de erros nas estimativas produzidas por métodos quantitativos de pesquisa. Em termos mais específicos, ao se tentar aferir a informação social por meio da coleta de dados, dificilmente o pesquisador consegue a informação almejada em sua forma “pura” ou “100% verdadeira”; os dados coletados embutem conteúdos de duas naturezas: a parte “verdadeira”, relativa ao tópico social que se quer aferir; e uma outra parcela de caráter aleatório, a que se chama de “erro de mensuração”. Assim, o valor assumido por uma variável consiste na soma desses dois componentes.

Para a análise estatística, a existência de erros de mensuração atrela-se à questão da qualidade das medições. Em geral, a qualidade de uma medida é indicada por intermédio de critérios relativos à sua exatidão e à sua precisão. A questão da exatidão diz respeito à *validade* da medida, uma consideração técnica que “se refere ao grau com que uma medida empírica reflete adequadamente o *significado real* do conceito considerado”⁴ (BABBIE, 1999, p. 196).

A questão da precisão, por seu turno, é que se encontra propriamente ligada ao problema dos erros de medição. Aqui, está-se falando da *confiabilidade* do dado estatístico, que indica a consistência da mensuração. Conforme descreve Babbie (1999, p. 195), “a *confiabilidade* é o problema de uma determinada técnica, ao ser aplicada repetidamente a um mesmo objeto, produzir, a cada vez, os mesmos resultados”. Metodologicamente, a confiabilidade de uma variável é definida como a proporção da variância observada que é uma variância “verdadeira”, ou ainda, como a parte de uma medida que está livre do erro de mensuração. Assim, uma medida é considerada confiável se ela reflete majoritariamente a informação verdadeira (BOLLEN, 1989; KLINE, 1998).

Na maioria das situações, as confiabilidades das medidas são desconhecidas. Contudo, é possível estimar a confiabilidade ou a variância do erro de mensuração de

4. Para um melhor detalhamento da discussão sobre “validade” e sobre os tipos de validade, ver Babbie (1999), Bollen (1989) e Kline (1998).

algumas variáveis, havendo diversas técnicas para isso.⁵ No entanto, mesmo em situações em que a confiabilidade é conhecida – ou seja, tem seu valor estimado –, ela não é incorporada nos cálculos realizados pela maioria das metodologias tradicionalmente utilizadas nas pesquisas sociais. Com isso, os erros de mensuração são incluídos nas investigações e acabam funcionando como “ruidos” ou “desordenações interferentes” no processo de análise de dados (WISNIK, 1989, p. 33).

De fato, na prática, não se podem expurgar tais erros dos dados coletados. Por isso, os resultados gerados por uma dada metodologia são afetados por eles. Uma vez que o processamento dos dados acontece não apenas sobre a parte “real” da informação social, mas também sobre a parcela de erro embutida na medida, os resultados das análises distanciam-se, de alguma forma, daquilo que seria verdadeiro.

Como se pode observar, todos esses aspectos confluem para que seja reforçada a perspectiva de que a complexidade de uma realidade social é fluida e inesgotável. Isso evita, de antemão, que se tenha a pretensão de apreendê-la plenamente em um modelo empírico. Entretanto, tanto os aspectos desejáveis em uma metodologia para o estudo da pobreza quanto a realização do procedimento sugerido por Wilber têm se tornado cada vez mais factíveis. Para tanto, têm contribuído os avanços dos métodos quantitativos de pesquisa social e dos programas computacionais estatísticos. Além disso, a evolução dos procedimentos científicos e tecnológicos tem proporcionado um refinamento da apreensão da complexidade do fenômeno nos estudos empíricos, por também dar conta de algumas das limitações já mencionadas. Nesse sentido, as novas técnicas têm propiciado: a ampliação do número de fatores que podem ser introduzidos simultaneamente em um modelo empírico; a inclusão, nos cálculos estatísticos, de conceitos sociais abstratos não diretamente observáveis na realidade; e a avaliação de erros de mensuração, uma vez que existem métodos que possibilitam o ajuste dos resultados em relação a esses erros, por meio da modelagem e da provisão de estimativas explícitas desses parâmetros.

3 METODOLOGIAS TRADICIONALMENTE UTILIZADAS NOS ESTUDOS SOBRE POBREZA

Desde o advento do capitalismo, a pobreza tem atraído os interesses intelectual e político de diversos agentes da sociedade. Várias tentativas de se mensurar sua dimensão, de se identificar quem são os pobres e de se compreender suas determinações têm sido empreendidas, de acordo com os entendimentos e definições acerca do fenômeno vigente em cada época. Para tanto, têm sido utilizadas metodologias de sistematização e de análise de dados empíricos, que possam propiciar tais conhecimentos sobre as realidades de pobreza.

Em princípio, as formulações conceituais sobre o tema priorizavam a noção de que os pobres eram aqueles que não obtinham renda suficiente para garantir suas sobrevivências – ou seja, quando predominava o ideário da “subsistência”. Naquele contexto, os esforços de investigação empírica voltavam-se tão-somente para a contagem ou quantificação da parcela da população que era considerada pobre. Os

5. A esse respeito, ver Bollen (1989).

métodos utilizados para balizar essas pesquisas consistiam na confecção de indicadores unidimensionais, calcados no valor da renda necessária para que a sobrevivência da família fosse assegurada. Os estudos limitavam-se então a enumerar quantos eram os pobres, aqueles cuja renda fosse inferior a esse patamar, em uma sociedade.

Em épocas posteriores, houve a ampliação do conceito em direção à multidimensionalidade, ocorrida a partir da formulação da satisfação de diversas “necessidades básicas”, para que o ser humano pudesse ter condições mínimas de vida. Mesmo assim, os estudos empíricos permaneceram, por algum tempo, voltados para a questão da quantificação da população pobre. No entanto, em vez de se basearem exclusivamente no critério monetário, as contagens dos pobres passaram a se calcar em uma série de indicadores que levavam em conta os parâmetros relativos à satisfação das necessidades consideradas mínimas em uma sociedade. As diversas dimensões da pobreza podiam também ser abarcadas e sumarizadas sob a forma de um único número ou índice, constituído a partir da aglutinação das várias dimensões do fenômeno, o qual se costuma designar de “indicador sintético” (JANNUZZI, 2001, p. 22).

Com a evolução das formulações teóricas, consolidou-se a perspectiva multidimensional e emergiu a noção de que a pobreza é fruto de processos socioeconômicos complexos. Com isso, à preocupação de se quantificar a parcela pobre da sociedade adicionou-se a intenção de se identificar “quem são eles”. Os trabalhos interessaram-se então por traçar o perfil dessa fatia da população, detectando quais as principais características daqueles atingidos pelas condições de carências e privações. A idéia era que fossem indicados os atributos que faziam com que alguns grupos de indivíduos estivessem mais sujeitos à pobreza do que outros. Posteriormente, as investigações enveredaram na busca pelo conhecimento sobre o potencial de explicação de cada uma das referidas características a respeito das condições de pobreza. Dessa forma, os métodos empregados passaram a almejar o estabelecimento de associações entre atributos individuais e indicadores de pobreza, de modo que se pudesse definir quais fatores revelavam-se mais influentes e quais as intensidades de tais influências.

Os procedimentos metodológicos mencionados até aqui, que foram se impondo gradativamente de modo que as investigações empíricas pudessem acompanhar a evolução das percepções sobre a pobreza ao longo do tempo, predominam até os dias de hoje. Por meio deles, realizam-se diversos estudos acerca das situações de pobreza nas sociedades atuais, que são desenvolvidos com base nas formulações conceituais mais recentes sobre o tema. Entretanto, observa-se que a complexidade da pobreza é captada por esses métodos tradicionais de forma parcial e fragmentada, particularmente no que se refere à articulação entre os múltiplos fatores envolvidos no fenômeno.

A insuficiência desses métodos em lidar com a complexidade inerente às formulações atuais sobre o tema liga-se ao fato de que a discussão conceitual tem evoluído na direção de se incluírem as abordagens dos diversos processos sociais envolvidos no fenômeno. Mais especificamente, observa-se que, no âmbito teórico, tem-se enfatizado a contemplação de como vários aspectos sociais relacionam-se entre si e com as condições de pobreza. Essa ênfase faz com que indicadores unidimensionais não

bastem para sustentar o desenvolvimento de estudos empíricos, uma vez que não espelham a complexidade da questão. Pelo mesmo motivo, também não se trata de agregar as múltiplas dimensões da pobreza em um número ou índice, com vistas à quantificação do problema em uma sociedade. Como se vê, a discussão tem caminhado para uma perspectiva que não pode ser retratada apenas por métodos baseados em agregações de dados, já que ressalta aspectos como as inter-relações das dimensões do fenômeno e as especificidades relativas às distintas realidades sociais, as chamadas “síndromes de pobreza”.⁶

Pelo exposto, observa-se a necessidade de que a discussão metodológica acompanhe a trajetória do debate conceitual. É desejável que o arsenal técnico utilizado nas pesquisas esteja atualizado e sintonizado com os avanços teórico-conceituais, de maneira que ele possa cumprir o seu papel de “ponte” que une os conteúdos elaborados na teoria e as pesquisas sobre realidades empíricas.

Guiados por esse entendimento, Betti, D’Agostino e Neri (2000, p. 2) realizaram uma revisão da literatura sobre metodologias quantitativas de pesquisa social. Os autores observam que o estágio atual das investigações sobre pobreza não contempla o vasto potencial oferecido por essa área do conhecimento. Constatou-se que a visão multidimensional do problema, como difundida nas últimas décadas, não tem sido devidamente apreendida pela maioria dos estudos empíricos, os quais têm se calcado em modos tradicionais de analisar a questão, tais como indicadores unidimensionais de renda, róis de índices socioeconômicos ou mesmo indicadores que sintetizam, em um único número, as múltiplas dimensões da pobreza. Por isso, os autores afirmam a necessidade de se utilizarem novos métodos, capazes de refletir as feições do fenômeno abordadas na teoria.

Fitoussi e Rosanvallon (1998, p. 85) também corroboram a opinião de que a aplicação de métodos estatísticos deve buscar adequar-se à evolução da discussão teórico-conceitual, evolução essa que espelha as mudanças ocorridas na realidade social. Segundo os autores, a estatística social – destinada a avaliar aspectos como a renda familiar, suas comodidades materiais e a inserção profissional do chefe – era um exercício de pouca complexidade em contextos anteriores, quando a maioria das famílias seguia uma “composição-padrão”, caracterizada pela presença de um homem assalariado, com uma inserção estável no mercado de trabalho, empregado em tempo integral e remunerado em função da categoria de seu posto e de seu tempo de serviço. Os autores observam que, contrariamente a esse panorama, os tempos atuais são marcados pela complexidade dos fenômenos, a qual deve ser açambarcada pelos procedimentos metodológicos das pesquisas. Como se disse, os arranjos familiares são hoje mais diversificados e instáveis, podendo haver a presença de duas ou mais pessoas ativas, cujos riscos de exposição à desocupação são heterogêneos. Assim, as múltiplas variações que marcam a vida social na atualidade fazem com que as formas

6. Schwartzman (1997) propõe que as reflexões acerca do tema se elaborem em termos de “síndromes de pobreza”. Essas síndromes, diz o autor, referem-se à abordagem de condições complexas – relativas a linguagem, etnia, cultura, localização, instituições econômicas, educação e história –, específicas dos grupos afetados e de seus ambientes sociais. Essa perspectiva é reforçada por Rocha (2003). Segundo a autora, a diversidade de níveis de desenvolvimento entre países e suas distinções culturais demandam a adoção de conceitos de pobreza que considerem as especificidades de cada situação. Por isso, a definição de um conceito de pobreza e a escolha de procedimentos de mensuração pertinentes resultam de uma análise cuidadosa de cada realidade social específica, de modo que se possam identificar os traços essenciais da pobreza em determinada sociedade.

tradicionais de abordagem das realidades empíricas não dão conta da diversidade das situações existentes.

Fitoussi e Rosanvallon (1998, p. 87) reconhecem que, por várias razões, a captação desses fenômenos é uma tarefa complexa, que revela a necessidade de se instaurar novos procedimentos e categorias de análise. Salama e Destremau (2001, p. 131) compartilham desse ponto de vista e comentam que, na França, têm emergido abordagens inovadoras que articulam e refinam medidas estatísticas da pobreza, adaptando-as a contextos heterogêneos e móveis. No bojo dessas abordagens, nota-se o desejo de que as investigações acerca de um ponto no tempo sejam ampliadas para estudos longitudinais, que são aqueles que acompanham os indivíduos ao longo do tempo e, portanto, são capazes de detectar o modo como evoluem suas condições de vida. A idéia é que se possam perceber o ritmo e a forma das mudanças nos padrões de vida das pessoas, de maneira que seja possível compreender as dinâmicas de certos fenômenos, como o desemprego e a inserção laboral. A esse respeito, deve-se comentar que, apesar de valiosos em termos dos conhecimentos que revelam, tais estudos são mais densos, custosos e complexos. Por isso, pesquisas longitudinais sobre pobreza têm viabilidade complicada, sendo ainda raras em todo o mundo.

De todo modo, o anseio de ampliar as investigações em direção à análise dos tempos e dinâmicas dos fenômenos indica que o debate metodológico segue rumo à incorporação da complexidade da pobreza nas abordagens empíricas. Conseqüentemente, os objetivos de medida e de construção de indicadores deixam de constituir o foco de atenção principal. Vale notar que, frente a tal redirecionamento, atributos tradicionalmente abordados, como educação, renda, desemprego e moradia, continuam sendo destacados como fatores relevantes dentro das novas análises (SALAMA; DESTREMAU, 2001, p. 132).

Com relação às investigações realizadas nos países em desenvolvimento, Sliwiany (1997, p. 16) afirma que os estímulos às pesquisas estatísticas sobre a população são menos intensos que nos desenvolvidos. Os diferenciais entre as produções científicas entre esses grupos de países apresentam-se tanto em termos de quantidade dos trabalhos realizados, como da qualidade das técnicas utilizadas. Mesmo assim, estudos dessa natureza são empreendidos nos países em desenvolvimento, voltando-se para análises de realidades sociais e contribuindo para que a estatística social ganhe impulso e difusão nessas sociedades.

No âmbito brasileiro, pode-se notar que as pesquisas recentes sobre pobreza seguem a tendência ora mencionada de se utilizarem metodologias tradicionais de investigação. Buscando refletir a evolução do campo teórico-conceitual, os trabalhos empíricos têm almejado capturar a complexidade e a multidimensionalidade do fenômeno por meio da elaboração de novos indicadores sintéticos e do cruzamento de diversos indicadores tradicionais. Há também algumas tentativas de se conhecerem as relações entre fatores socioeconômicos e medidas de pobreza, mediante estudos baseados em análises de regressão. Entretanto, como a maioria de tais investigações baseia-se em métodos que se mostram limitados para apreender a articulação dos fatores associados ao problema, tais abordagens sobre pobreza acabam não refletindo o teor das formulações teóricas mais recentes, que a concebem como “um espaço de relações”.

A partir dessa discussão acerca das metodologias quantitativas de análise, são examinados a seguir os principais procedimentos que têm sido aplicados à abordagem da pobreza enquanto fenômeno complexo: os indicadores sintéticos, os sistemas de indicadores sociais e as análises de regressão. A explanação sobre cada método consiste em uma breve descrição do procedimento e menciona alguns estudos recentes sobre o fenômeno que tenham sido baseados nas respectivas técnicas. São também notados os pontos fortes inerentes a cada método, além de serem colocadas algumas críticas sobre seus potenciais de refletirem, durante análise dos dados empíricos, a percepção relacional que vigora atualmente no debate teórico sobre a pobreza.

3.1 ÍNDICES SINTÉTICOS

Ao discutir a questão da adequação entre teoria e método, Székely (1996, p. 2) faz uma revisão da literatura a respeito das metodologias de pesquisa aplicadas ao estudo da pobreza e percebe o delineamento de uma tendência nas últimas décadas. O autor nota que, entre os trabalhos recentes, assiste-se ao surgimento de um grande número de índices sintéticos, que sumarizam as características de populações de diversos contextos. Os estudos que aplicam tal metodologia buscam captar a multidimensionalidade do fenômeno a partir da conjugação das diversas facetas que o compõem. A idéia é que se possa expressar, por meio de um número ou coeficiente, uma estimativa do nível das carências e privações existentes em uma determinada situação.

Os indicadores sintéticos ou “índices sociais” são criados a partir da aglutinação de dois ou mais indicadores simples, relativos a distintas dimensões da realidade social. Conforme Jannuzzi (2001, p. 22), indicadores de natureza composta têm o mérito de serem simples e de oferecerem alto poder de síntese. São, portanto, vistos como propícios a situações em que se deseja ter uma avaliação geral sobre um conceito que está sendo investigado em determinados grupos sociais. O autor entende que, quando se trata de uma questão complexa, em que se relacionam múltiplos fatores, existem vantagens de se trabalhar com um indicador sintético, em vez de se dispor de sistemas constituídos por uma série de indicadores separados. A razão para isso é que o índice sintético é capaz de dar uma idéia genérica sobre o fenômeno estudado, de forma mais rápida e objetiva.

Além disso, indicadores sintéticos oferecem a possibilidade de que sejam realizados confrontos e classificações entre situações sociais distintas, assim como propiciam comparações ao longo do tempo (HAQ, 2003, p. 110). No que se refere ao estudo da pobreza, Ravallion (1994, p. 1) ressalta que as comparações visam a esclarecer questões que podem ser relevantes para o conhecimento das nuances que marcam realidades diversas, tais como: a pobreza tem crescido ao longo do tempo? É mais alta em um lugar do que em outro? Há mais pobreza após alguma mudança de política?

Como se disse, diversos estudos sobre pobreza têm sido desenvolvidos nos últimos tempos com base nessa metodologia, tanto no âmbito internacional como no nacional. Internacionalmente, o exemplo mais famoso desse tipo de indicador é o Índice de Desenvolvimento Humano (IDH), que foi proposto pelo Programa das Nações Unidas para o Desenvolvimento (Pnud) no começo da década de 1990.

Criado por Mahbub ul Haq em colaboração com Amartya Sen, o IDH apresenta-se como medida geral e sumária do desenvolvimento humano, ainda que não ambicione abranger todos os aspectos do desenvolvimento – uma vez que tal conceito é mais rico e complexo do que um índice pode esperar capturar.

Em épocas anteriores, a avaliação do bem-estar de uma população e, por conseguinte, a classificação de países ou regiões, se dava pelo tamanho de seu Produto Interno Bruto (PIB) *per capita*. Entretanto, com a evolução da noção de que o progresso humano e a melhoria das condições de vida das pessoas não podem ser medidos apenas por sua dimensão econômica, emergiu a idéia de se criar um indicador mais abrangente, que transcendesse o aspecto econômico da pobreza, englobando outros componentes do fenômeno. Instituiu-se assim o IDH, cujo objetivo é oferecer um contraponto ao PIB *per capita* enquanto indicador de condições de vida. Sob essa ótica, diz-se que uma vantagem apresentada pelo indicador sintético está no fato de que ele se oferece como uma medida mais robusta e abrangente do conceito desenvolvimento humano, já que é formado por múltiplas facetas consideradas relevantes nessa discussão, sendo a dimensão da renda apenas uma delas. Assim, pode-se fornecer uma média sobre o nível de desenvolvimento humano de um país ou região que seja mais significativa do que aquela calculada unicamente a partir do PIB *per capita*.

O IDH é composto pela combinação de três elementos básicos do desenvolvimento humano. O primeiro é a longevidade, medida pela esperança de vida ao nascer, que prima por refletir, entre outros aspectos, as condições de saúde da população. O segundo refere-se à educação, medida pela união da taxa de alfabetização de adultos e a taxa combinada de matrícula nos níveis de ensino fundamental, médio e superior. Por fim, tem-se a renda, que é medida pelo poder de compra da população, baseado no PIB *per capita* ajustado ao custo de vida local, de modo que se possa torná-lo comparável entre países e regiões.⁷

Indicadores de natureza sintética podem sofrer vários tipos de decomposições, de modo a exibirem as informações implícitas nas medidas globais. Um indicador como o IDH, por exemplo, pode ser decomposto no nível das dimensões que o constituem. Assim, podem ser produzidos subíndices ou “IDHs parciais”, relativos à longevidade, à educação e à renda. Além disso, as decomposições podem ocorrer segundo outros critérios que se mostrem relevantes. É possível a produção de IDHs separados por sexo, grupo étnico, região geográfica ou por outros aspectos que espelhem a sociedade. Com isso, as disparidades existentes entre distintos subgrupos da população de um país podem ser delineadas (HAQ, 2003, p. 106).

Um outro exemplo de indicador sintético de aplicação internacional é o Índice de Pobreza Humana (IPH), apresentado também pelo Pnud, em seu Relatório de Desenvolvimento Humano de 1997. Este índice pode ser visto como um desdobramento do IDH e surge da noção de que as situações de pobreza que se apresentam nos países em desenvolvimento não são comparáveis àquelas vigentes nos países desenvolvidos, tanto em termos de profundidade como de especificidades. Assim, o IPH propõe-se a capturar

7. Tal ajuste se dá através da metodologia conhecida como Paridade do Poder de Compra (PPC), que elimina as diferenças de custo de vida entre os países, baseando-se no poder de compra da moeda de cada nação.

elementos da pobreza humana nos países em desenvolvimento, a partir da inclusão de três componentes: longevidade, conhecimento e padrão de vida. A longevidade é inserida sob a forma da porcentagem de pessoas cuja expectativa de vida é inferior à idade de 40 anos. O conhecimento é incluído no cálculo do indicador através da porcentagem de adultos que são analfabetos. O padrão de vida, por sua vez, é representado como uma composição de três variáveis: porcentagem de pessoas que têm acesso a serviços de saúde; porcentagem de pessoas que usufruem de água potável; e porcentagem de crianças menores de cinco anos que estão subnutridas.

Entre a produção recente dos pesquisadores sociais brasileiros, observa-se também uma tendência de produção de indicadores sintéticos sobre pobreza. Essas tentativas que vêm sendo empreendidas demonstram uma preocupação relativa à caracterização da pobreza em áreas específicas. Trata-se, eminentemente, de trabalhos que aliam a criação de novos indicadores sintéticos a técnicas de geoprocessamento de informações, confeccionando os chamados “mapas de exclusão social”. Esses mapas registram as configurações espaciais de variáveis e índices de pobreza elaborados pelos próprios pesquisadores.

Entre essas iniciativas, pode-se citar o Mapa da exclusão/inclusão social da cidade de São Paulo: dinâmica social dos anos 90 de Sposati *et al.* (2000), um trabalho que foi iniciado em 1995 e que, com base em sua segunda edição lançada no ano 2000, teve a intenção de indicar tendências espaciais e temporais dos fenômenos analisados na cidade. Inspirando-se no IDH e defendendo a necessidade da criação de um índice que oferecesse um “modo novo de pensar a cidade para propor soluções” (SPOSATI *et al.*, 2000, p. 5), o Índice de Exclusão/Inclusão (IEX) foi apresentado como um “conjunto de ferramentas” capaz de auxiliar a análise da cidade. Conceitualmente, a criação do IEX partiu do entendimento de que a inclusão social pode ser aferida por meio de noções abstratas, como autonomia, desenvolvimento humano, qualidade de vida e equidade. Com base nesse entendimento, cada um desses aspectos veio a se constituir nas dimensões a serem incluídas e aglutinadas no cálculo do IEX.⁸

Uma outra iniciativa de criação de indicadores sintéticos apresenta-se em Pochmann *et al.*, 2003, que é um trabalho de natureza semelhante à do recém-mencionado. Ele apresenta abordagens nos níveis estadual e municipal, a partir da utilização dos dados do Censo de 2000 do Instituto Brasileiro de Geografia e Estatística (IBGE). Esse indicador erige-se também a partir da consideração de conceitos abstratos, que devem ser conjugados para o cálculo do índice: padrão de

8. A dimensão da autonomia foi representada por intermédio de indicadores relativos a emprego, renda familiar e contagem da população em situação de indigência. O desenvolvimento humano foi incorporado através de informações sobre o grau de instrução dos chefes de família, a longevidade, os anos potenciais de vida perdidos (definidos como o acúmulo de anos não vividos entre a idade de morte e a expectativa de vida até os 70 anos) e a taxa de homicídios. A qualidade de vida, por sua vez, foi indicada por aspectos como densidade habitacional, verticalização distrital (considerada como a taxa anual de lançamento de edifícios, o que indica o interesse imobiliário em regiões da cidade, que muda o perfil populacional e atrai novos serviços coletivos), oferta de serviços para crianças de 0 a 4 anos, escolas de educação infantil, escolas de educação fundamental e saúde. Por fim, a equidade foi incluída no cálculo do indicador por meio de duas variáveis: o número distrital de mulheres chefes de família e o número distrital de mulheres chefes de família analfabetas.

vida digno, conhecimento e risco juvenil.⁹ Para tais dimensões, foram calculados subíndices específicos, que foram posteriormente agregados em uma única medida global, o Índice de Exclusão Social (IES).

Os trabalhos ora mencionados ilustram uma tendência que vem se delineando entre pesquisadores estrangeiros e brasileiros de que se elaborem novos indicadores sintéticos, que priorizem abarcar a multiplicidade de fatores envolvidos na discussão conceitual sobre a pobreza. Entretanto, ainda que a elaboração desses indicadores seja enriquecida pela aplicação de técnicas de geoprocessamento para a confecção de mapas de exclusão social, esse procedimento metodológico exhibe propriedades eminentemente enunciativas, produzindo estatísticas descritivas, e não trabalhando a dinâmica das relações entre as dimensões associadas ao fenômeno.

Apesar da intensa produção de trabalhos que têm se baseado nessa metodologia e das já referidas potencialidades dos indicadores sintéticos para a análise de situações sociais, são apontadas algumas limitações inerentes aos índices dessa natureza. Jannuzzi (2001, p. 22) coloca que um dos problemas intrínsecos aos indicadores sintéticos está no fato de que eles reduzem a multidimensionalidade de um conceito complexo a um único número ou coeficiente. Coloca-se que a operação de sumarizar os conteúdos relativos às diversas facetas do conceito em um só índice pode incorrer em perda de transparência de informações para seus potenciais usuários. Essa crítica é reforçada por Conde (1999, p. 144), quando acrescenta que esse método inclui cada dimensão de maneira independente, não permitindo que se identifiquem e analisem as inter-relações entre elas.

Tal limitação pode ser notada nas análises sobre pobreza que vêm sendo realizadas com base na utilização de indicadores sintéticos. Ao sumarizarem informações socioeconômicas a respeito dos indivíduos sob a forma de um algarismo, tais indicadores oferecem uma visão opaca sobre tal fenômeno multidimensional, uma vez que não revelam a estrutura de articulação existente entre os fatores envolvidos no processo sob análise. Isso pode ser observado em qualquer dos exemplos de indicadores sintéticos sobre pobreza recém-mencionados. Tomando-se como ilustração o caso do IES (POCHMANN *et al.*, 2003), percebe-se que tal indicador nada esclarece sobre a maneira como os seus componentes – padrão de vida digno, conhecimento e risco juvenil – encontram-se relacionados nas realidades analisadas.

Ainda discutindo esse assunto, Strobel (1996, p. 207, *apud* SALAMA e DESTREMAU, 2001, p. 114) acrescenta que o procedimento de se criarem índices sintéticos para a análise da pobreza está imbuído de uma contradição entre método e teoria, já que, “partindo de uma hipótese inicial da multidimensionalidade da pobreza, chega-se, *in fine*, a uma classificação unidimensional”. Esse aspecto assume relevância dentro do presente estudo, uma vez que se busca analisar aqui procedimentos quantitativos de pesquisa sob o ponto de vista da capacidade de

9. A dimensão do padrão de vida digno foi representada por indicadores relativos à pobreza dos chefes de família, pela taxa definida a partir da divisão entre o número de empregos formais pela população em idade ativa (PIA) e por uma *proxy* da desigualdade de renda. O conhecimento foi incluído no indicador por intermédio da taxa de alfabetização de pessoas acima de cinco anos de idade e do número médio de anos de estudo do chefe de domicílio. O risco juvenil foi aferido por meio da porcentagem de jovens da população e do número de homicídios por 100 mil habitantes.

apreensão da natureza relacional dos fenômenos sociais, particularmente do caráter articulado da questão da pobreza.

Essas limitações referentes à utilização de indicadores sintéticos para o estudo da pobreza encontram-se relacionadas ao princípio de agregação de dados, que é criticado por reduzir a quantidade de informação disponível (RAWORTH; STEWART, 2003, p. 148). As críticas relacionadas a tal princípio aplicam-se de maneira mais contundente quando se efetua a agregação de dados relativos a grandes áreas. Especificamente no que concerne aos estudos sobre pobreza, considerando-se que situações de carências e privações caracterizam-se por assumirem particularidades em seus diversos contextos – fato que levou Schwartzman (1997) a cunhar a expressão “síndromes de pobreza” para designar as peculiaridades existentes em situações distintas –, há que se ter cautela ao se utilizarem métodos de análise que se calcem em agregações de dados, como é o caso dos indicadores sintéticos. Ao se efetuarem essas agregações, principalmente em áreas extensas, podem ser perdidas especificidades relativas a cada situação. Essa é uma das críticas dirigidas, por exemplo, ao cálculo de IDHs para os países. Sabe-se que, em uma mesma nação, é possível que se encontrem diversos níveis de desenvolvimento humano; a produção de um único indicador que represente todo o conjunto pode mascarar a existência de tal diversidade.

Em geral, os procedimentos de agregação de dados partem da assunção de que as informações são agrupadas por áreas onde pode haver homogeneidade interna, em termos de características socioeconômicas e demográficas. Contudo, nem sempre essa premissa se faz verdadeira. Nada garante que a distribuição do evento pesquisado seja homogênea dentro de tais áreas, já que as unidades de levantamento de dados são normalmente estabelecidas segundo critérios operacionais (setores censitários) ou políticos (municípios). A possibilidade de existência de heterogeneidade dentro de uma mesma área é ainda mais intensa em casos de países com grandes contrastes sociais, como acontece com o Brasil. Nessas situações, é usual que grupos sociais distintos – favelas e áreas nobres – estejam agregados em uma mesma região de coleta, dando origem a indicadores que representam a média entre populações diferentes (DIAS *et al.*, 2002, p. 91).

Ainda apontando as fragilidades analíticas que podem advir da operação de agregação de dados, Dias *et al.* (2002, p. 92) assinalam que, para uma mesma população estudada, “a definição espacial das fronteiras das áreas afeta os resultados obtidos”. Em outras palavras, as estimativas feitas a partir de certo sistema de unidades de área podem variar, a depender da forma como se agrupem as áreas a serem agregadas. Com isso, pode-se chegar a resultados inconsistentes entre si, já que a modificação das fronteiras entre essas áreas é capaz de alterá-los.

A questão da agregação é também posta por Anand e Sen (2003b, p. 114-115). Os autores comentam o caso do IDH, em que as medidas usuais de desempenho de grupo (como o PIB e a expectativa de vida ao nascer) tendem a ser indicadores agregados, calculados com base na média das circunstâncias individuais. Eles reconhecem que tal procedimento envolve a perda de informações detalhadas, por usar um número agregado (um escalar) para aglutinar um conjunto de números que representam circunstâncias individuais (um vetor). Para esclarecer essa questão, os

autores apresentam o exemplo de uma situação em que três pessoas têm, respectivamente, os níveis de renda nos valores de (1, 9, 11). De acordo com os procedimentos de agregação, utilizados no cálculo do indicador sintético, tal situação parece ser similar a outra, em que três pessoas têm rendas nos valores de (7, 7, 7), embora essas duas situações sociais dificilmente possam ser tidas como equivalentes em termos de preocupações e valores.

Casos como esse apontam a necessidade de que se elaborem indicadores “sensíveis à distribuição”, em lugar dos usuais indicadores baseados em médias simples. Isso se deve ao fato de que o procedimento de se tirar uma média deixa escapar o verdadeiro padrão de distribuição de uma coletividade de números, concentrando-se apenas em seus valores intermediários. Ainda que se busquem tais indicadores “sensíveis à distribuição”, dizem aqueles autores, eles continuariam a incorrer em perda de informação, já que não há como capturar e representar a riqueza de conhecimento fornecida por um grupo de números através de um só algarismo (ANAND; SEN, 2003b, p. 115).

Ravallion (1994, p. 76) conduz a discussão crítica sobre os indicadores sintéticos para a questão da escolha de dimensões que os irão compor. Particularmente no que se refere ao tema da pobreza, o autor chama a atenção para o fato de que nem sempre é possível conjugar, de modo preciso, as múltiplas dimensões do bem-estar. O problema reside na arbitrariedade envolvida na escolha dessas dimensões; é possível que as diversas dimensões escolhidas conceitualmente para compor o indicador representem fenômenos logicamente correlacionados, de modo que haja redundância de informações, o que vem a adicionar pouca luz à discussão (SALAMA; DESTREMAU, 2001, p. 114; RAWORTH; STEWART, 2003, p. 149). Neste caso, estar-se-ia medindo apenas uma faceta do fenômeno e, conseqüentemente, pouco se estaria captando em termos da sua multidimensionalidade. Por outro lado, a existência de uma situação oposta, em que se tenham aspectos de naturezas divergentes a serem congregados em uma única medida, pode comprometer a estabilidade do índice que se quer calcular (RAWORTH; STEWART, 2003, p. 144).

Outra questão referente à escolha das dimensões componentes do índice diz respeito à relevância conceitual de cada uma delas. Sabe-se que a capacidade de incorporação de distintas dimensões em um indicador sintético é limitada. Sendo a pobreza um fenômeno multifacetado, recomenda-se que sejam incluídos apenas os aspectos mais significativos para a análise. Sobre esse assunto, Haq (2003, p. 111) argumenta que a inserção de um número maior de dimensões não garante a melhoria das propriedades de um indicador sintético. Pode acontecer que elas venham a adicionar elementos supérfluos à análise, tornando confuso o panorama que se deseja obter e dificultando o delineamento das tendências principais. Assim, há vantagens em se manter a parcimônia durante o processo de construção de indicador sintético.

Há que se comentar também sobre a possibilidade de aferição empírica das dimensões escolhidas. Além da relevância conceitual, os componentes de um indicador sintético devem ser escolhidos em função de suas viabilidades empíricas. A idéia é que os aspectos conceituais possam ser captados em realidades sociais, a partir da disponibilidade e do acesso a dados confiáveis que possam refleti-los adequadamente.

Ligada à discussão das dimensões componentes de um indicador sintético encontra-se a questão da ponderação. A construção de um índice numérico, que se proponha a contemplar a multidimensionalidade de um fenômeno, requer que sejam tomadas decisões também acerca dos pesos que serão atribuídos a cada uma das dimensões incluídas no indicador. A discussão sobre a ponderação é tida como uma etapa inevitável do processo de criação de um índice composto (BECCARIA; FERES; SÁINZ, 1999, p. 17; SALAMA; DESTREMAU, 2001, p. 114; ANAND; SEN, 2003a, p. 207).

O exercício de ponderação é criticado como sendo uma etapa arbitrária no processo de criação do índice sintético (MESTRUM, 2002, p. 37). Isso se deve ao fato de que não há consenso ou regra sobre a necessidade de atribuição de distintos pesos a cada faceta incluída no indicador. Da mesma forma, não há uma convergência em torno dos procedimentos segundo os quais os pesos devem ser imputados (MEJÍA; VOS, 1999, p. 176).

Um exemplo dessa controvérsia pode ser encontrado na discussão existente em torno do IDH. No caso desse índice, pesos iguais são arrogados a todas as dimensões. Com isso, está-se afirmando que todas as escolhas são igualmente importantes (HAQ, 2003, p. 104). Contudo, há controvérsias quanto a esse posicionamento. Críticos afirmam que a atribuição de ponderações iguais para os três componentes daquele indicador é um ato arbitrário; tal desaprovação baseia-se no entendimento de que, conceitualmente, a renda deveria receber um peso maior por ser considerada como a dimensão mais importante do IDH (RAWORTH; STEWART, 2003, p. 148).

Apesar das críticas voltadas para o potencial de reflexão de conceitos complexos e multidimensionais por parte dos índices sintéticos, esses índices quantitativos são vistos como relevantes para a elaboração de políticas públicas. Isso decorre do fato de que ações antipobreza precisam se balizar em instrumentos que ajudem a delimitar a população visada, a estimar seus custos e a avaliar seus efeitos (SALAMA; DESTREMAU, 2001, p. 105). Dessa forma, aquelas pesquisas justificam-se por produzirem indicadores que se propõem a auxiliar a formulação de políticas voltadas para a melhoria das condições de vida das pessoas pobres em uma sociedade.

Nesse contexto, Székely (1996, p. 2) coloca que, para que se possam delinear ações voltadas para esses fins, são necessários conhecimentos sobre “quantos são os pobres”, “qual é o perfil do grupo” e “por que aquelas pessoas são pobres”. Aliando as formulações teóricas recentes aos propósitos de formulação de políticas públicas, o autor afirma que, dentre os aspectos recém-mencionados, deve-se priorizar a compreensão de “por que eles são pobres”. Entende-se que é esse conhecimento que aumenta as possibilidades de se lidar com o problema.

Sob esse prisma, os índices sintéticos de pobreza são de pouca ajuda para fornecer tal tipo de ciência acerca da realidade. Em vez de ajudarem a compreender “por que eles são pobres”, seu maior potencial está no fato de que eles oferecem uma avaliação genérica sobre as condições de privação em um contexto, além de servirem de ferramenta para a realização de comparações de pobreza entre distintas realidades. Essas características, como se disse, associam-se ao procedimento da agregação de dados que, por sua vez, traz os já comentados inconvenientes para a análise das situações de carências e privações.

A essa altura, vale a pena mencionar o balanço realizado por Anand e Sen (2003a, p. 205) a respeito da discussão sobre o potencial analítico oferecido pelos indicadores sintéticos para o estudo da pobreza. De modo sucinto, após analisarem os prós e os contras referentes a tal metodologia, os autores afirmam que a capacidade de se refletir conceitos amplos e complexos – como o de desenvolvimento humano ou mesmo a perspectiva da privação – em um índice útil e não complicado, é uma meta que ainda não foi devidamente atingida.

Tendo em vista todas essas considerações, pode-se dizer que a abordagem metodológica dos indicadores sintéticos não reflete, de modo satisfatório, as elaborações teórico-conceituais recentes acerca do fenômeno, segundo as quais o objeto social é construído relacionalmente. Frente a isso, Székely (1996, p. 2) lamenta o fato de que, ainda hoje, a maior parte da literatura dedique-se a quantificar os pobres através do cálculo de índices. Em consonância com esse posicionamento, Ravallion (1994, p. 2) ressalta a necessidade de que se pensem em novas formas de extrair informações adicionais a partir de dados que já existem.

3.2 SISTEMAS DE INDICADORES SOCIAIS

Outra abordagem metodológica bastante difundida entre os estudos de pobreza consiste na elaboração dos chamados “sistemas de indicadores sociais”. Esses sistemas são formados por conjuntos de indicadores relativos a um fenômeno social que se deseja analisar. Conforme explica Jannuzzi (2001, p. 17) “À semelhança do espaço geométrico euclidiano, em que é necessário um sistema de coordenadas para definir um ponto, o espaço social requer um sistema de indicadores referidos a múltiplas dimensões, para caracterizar um fenômeno social”.

Para que se elabore um sistema de indicadores sociais, deve-se proceder a uma série de decisões metodológicas. Assim como acontece na formulação de indicadores sintéticos, é necessário que se tenha uma definição operacional do conceito abstrato ou tema que deve ser reportado pelo sistema em questão. A partir daí, as distintas dimensões que compõem esse conceito são especificadas, de modo que possam ser representadas por meio de variáveis observáveis. Uma vez que se esteja de posse dos dados empíricos que reflitam as dimensões conceituais escolhidas, a elaboração do sistema de indicadores sociais acontecerá a partir de procedimentos de agregação de dados e da organização das estatísticas disponíveis (JANNUZZI, 2001, p. 18).

Diferentemente dos indicadores sintéticos, que primam por condensar as múltiplas dimensões de um conceito social complexo em um único número, os sistemas de indicadores sociais têm por objetivo representar as várias dimensões do conceito separadamente. Cada uma delas deve ser refletida através de indicadores unidimensionais. A idéia é que o sistema seja composto por róis de indicadores mais simples que, sendo analisados de forma conjunta, se propõem a oferecer uma visão abrangente do fenômeno estudado. Por outro lado, por exibir indicadores separados para cada dimensão formadora do conceito, essa abordagem propicia também que os aspectos componentes do fenômeno sejam estudados individualmente.

A aplicação desse procedimento aos estudos de pobreza encontra motivação na intenção de que se delinee a composição do grupo pobre. Tendo em vista esse objetivo é que se empreende um esforço de “decomposição” dessa fatia da população

de acordo com seus traços distintivos. Mais especificamente, tal procedimento consiste na divisão da parcela pobre em subgrupos, definidos segundo certas características que os identificam (nível educacional, região de residência, ocupação, idade, sexo etc.). A idéia é que, a partir disso, seja determinada a proporção de pobreza que se concentra em cada conjunto. Como resultado desse procedimento, são elaboradas tabelas que exibem vários indicadores sociais referentes aos fatores envolvidos no fenômeno.

Székely (1996, p. 4) entende que o maior mérito dos sistemas de indicadores sociais para a análise de situações de pobreza reside justamente no fato de que eles permitem decompor o fenômeno por subgrupos de população. A partir de critérios de classificação demográfica, tais como região, sexo e idade, são construídos os chamados “perfis de pobreza”: panoramas descritivos sobre quem são aqueles que compõem a população pobre. Como se disse, esses perfis são considerados importantes, pois propiciam que se avalie a concentração da pobreza em cada subgrupo.

Ainda que reconheça o valor do caráter “decomposto” da informação oferecida pelos sistemas de indicadores para o estudo do fenômeno, Székely (1996, p. 5) entende que tal procedimento tem alcance limitado quando se busca responder à questão de “por que aquelas pessoas são pobres”, já que não agrega conhecimentos sobre quais são as características que se destacam por oferecerem maior potencial de explicação do fenômeno.

Ao refletir sobre o potencial analítico oferecido pela metodologia dos sistemas de indicadores sociais, Ravallion (1994, p. 59) parte para a realização de um contraponto com as potencialidades oferecidas pelo uso dos indicadores sintéticos. Apesar de admitir a relevância da forma desmembrada como a informação é oferecida pelos sistemas de indicadores, o autor considera que esse aspecto não torna seu uso preferível ao de um índice sintético. Para ele, as mesmas potencialidades analíticas podem ser oferecidas pelo índice composto, uma vez que ele pode ser desmembrado em um perfil de pobreza. Além disso, Ravallion (1994, p. 60) aponta outro aspecto que indica a proximidade entre as duas abordagens. Para o autor, os objetivos almejados por elas são similares: a construção de perfis de populações pobres constitui-se em um caso especial de comparação de pobreza, já que mostra como o fenômeno varia entre subgrupos da sociedade, entre regiões de residência e setores de emprego. Segundo essas considerações, pode-se dizer que ambas as abordagens são essencialmente semelhantes, diferindo na forma como a informação é apresentada. Tal parecer se coaduna com a natureza eminentemente descritiva notada nesses dois procedimentos, que priorizam a análise das médias de determinadas variáveis.

Da mesma forma como foi observado no caso dos indicadores sintéticos, há diversos estudos que se baseiam na elaboração de sistemas de indicadores sociais para a análise da pobreza. Entre os trabalhos mais recentes, pode-se mencionar o projeto “Mapa do Fim da Fome”, realizado pelo Centro de Pesquisas Sociais da Fundação Getúlio Vargas (CPS/FGV, 2001, 2004). Este trabalho caracteriza-se por ser uma iniciativa de grande porte que, baseando-se em microdados do IBGE, apresenta perfis de pobreza de diversos estados e cidades do Brasil, chegando mesmo a exibir resultados em níveis intramunicipais, como distritos e subdistritos. Diversos tipos de

quadros com informações variadas são providos por esse estudo. As informações são exibidas sob a forma de tabelas que apresentam, em cada linha, a área estudada e, em cada coluna, os indicadores sociais de interesse.

As cidades do Estado do Rio de Janeiro, por exemplo, são descritas em termos do tamanho de suas populações totais, da porcentagem de homens e mulheres, do nível educacional das pessoas (medido pelos anos médios de estudo), da média de idade da população, da PIA (pessoas entre 15 e 65 anos de idade), da renda média mensal auferida através do trabalho principal e da jornada média de trabalho (medida pelo valor médio das horas semanais trabalhadas). Há também tabelas em que constam informações sobre como se distribuem as condições de moradia, o acesso a serviços públicos e a bens duráveis entre as populações das áreas estudadas. Além disso, há quadros que indicam a evolução temporal dos indicadores sociais analisados, sempre em termos de valores agregados que designam a média da população (CPS/FGV, 2004). Com relação ao Mapa do Fim da Fome que foi desenvolvido para o Estado da Bahia (CPS/FGV, 2001), há resultados de diversos indicadores sociais que representam a média do estado, assim como de suas regiões econômicas e de seus municípios. São mostradas informações sobre os tamanhos das populações e as porcentagens de pessoas que vivem em situação de indigência.

Outro estudo relevante desenvolvido no âmbito nacional foi realizado por Rocha (2003), que veio a compor o livro *Pobreza no Brasil, afinal, de que se trata?* Nesse trabalho, a autora baseia-se também em microdados do IBGE para traçar uma série de perfis de pobreza. São consideradas distintas situações tais como a totalidade da população brasileira, a metrópole de São Paulo e a zona rural nordestina. Sempre expostos sob os referidos formatos de tabelas, os resultados indicam a porcentagem de pobres e não-pobres em cada contexto, desagregados segundo os critérios de sexo, grupos etários, cor ou raça, proporção de analfabetismo entre jovens e adultos, taxas de frequência escolar e anos de escolaridade da população. Há também perfis dos grupos pobres e não-pobres, em termos de suas condições de inserção no mercado de trabalho. Outros perfis de pobreza são delineados, a exemplo daquele que tem em vista a composição familiar; nesse caso, são levados em conta aspectos como a presença de crianças e a taxa de dependência no domicílio.

Ainda tomando exemplos de trabalhos voltados para a realidade brasileira, o próprio IBGE oferece publicações em que seus dados são organizados sob a forma de sistemas de indicadores sociais. Nessa linha, oferecem-se estudos como a “Síntese de Indicadores Sociais 2000” (IBGE, 2001), que expõe, de modo comparado, indicadores sociais construídos com dados dos anos de 1992 a 1999, numa tentativa de se delinear a evolução das condições sociais do país ao longo da década. Em tal trabalho, as informações são agregadas por estado e região do país. São mostradas taxas e médias de variáveis relativas aos temas dos aspectos demográficos, saúde, educação, trabalho e rendimento, domicílio, família, crianças, adolescentes e jovens, idosos e desigualdades raciais.

Internacionalmente, a Comissão Econômica para a América Latina e o Caribe (Cepal) desenvolve o trabalho intitulado *Anuario Estadístico de América Latina y el Caribe*. Em sua edição relativa ao ano de 2003, a publicação contém uma seleção de indicadores sobre aspectos econômicos e sociais dos países da região, que se dispõem à

realização de comparações entre as distintas situações. Da mesma forma como aconteceu nos exemplos já mencionados, são exibidos diversos quadros estatísticos, compostos por taxas e valores médios de variáveis relativas a um mesmo tema. Assim, há tabelas voltadas para o delineamento do perfil das populações daqueles países em termos demográficos, de condições de inserção laboral, saúde, educação, proteção social, entre outros aspectos. Todas essas informações são apresentadas de forma desagregada segundo diversos critérios, tais como setor de atividade econômica, idade, sexo e nível de rendimento (CEPAL, 2004).

Existem estudos de natureza similar que incluem dados de países de todas as regiões do globo. O Banco Mundial, por exemplo, lançou sua primeira compilação anual de dados sobre desenvolvimento. A publicação chama-se *The World Development Indicators (WDI)*¹⁰ e inclui cerca de 800 indicadores em 87 tabelas, construídas com informações de 152 países. Da mesma forma, são traçados perfis de populações pobres e não-pobres, em que se levam em conta aspectos como sexo, rendimento, idade, educação e saúde (WORLD BANK, 2004).

Seja com a intenção de quantificar os pobres ou de lhes traçar o perfil, ainda que sob o prisma da multidimensionalidade, pode-se notar que abordagens calcadas na exposição de indicadores sociais são eminentemente descritivas, harmonizando-se com as formulações conceituais menos complexas. Tais procedimentos prestam-se bem aos propósitos de ordenar e comparar situações de pobreza. Esses propósitos, contudo, não se coadunam com o estado atual das discussões sobre o tema. Mais especificamente, pode-se dizer que a questão metodológica não consiste mais em se dispor de um modo de indicar “quem são os pobres” ou de contabilizá-los. O ponto principal diz respeito a como se deve proceder, de modo a expressar a multidimensionalidade e a forma de articulação dos fatores envolvidos na questão, evidenciando quais deles são mais relevantes.

Nesse sentido, o procedimento de criação de sistemas de indicadores sociais recebe algumas críticas. Em princípio, podem-se mencionar aquelas que se dirigem à baixa capacidade de articulação das informações por parte de tais sistemas. Por fornecerem uma multiplicidade de indicadores sociais separados, são geradas tabelas – muitas vezes extensas – que produzem uma visão desarticulada dos fatores envolvidos no fenômeno (HAQ, 2003, p. 104). Sliwiany (1997, p. 39) endossa esse ponto de vista, afirmando que uma lista composta por vários indicadores fornecidos de modo disperso dificulta uma análise articulada do problema. Haq (2003, p. 104) acrescenta ainda que o fornecimento de medidas separadas faz com que se perca a sinergia existente entre as diversas facetas que compõem uma questão complexa.

Por esses motivos, autores como Haq e Sen não se mostram afeitos à utilização de sistemas de indicadores sociais para a análise da pobreza (PNUD, 1999). Eles entendem que um conjunto de tabelas compostas por vários indicadores desconectados tem baixo poder de comunicação da informação. Ainda que tal metodologia apresente a vantagem de não levantar maiores questionamentos sobre ponderação das escolhas feitas ou dos métodos usados para compilar um índice sintético, os autores argumentam que considerações de caráter prático tornam os

10. Os Indicadores de Desenvolvimento Mundial (IDM).

sistemas de indicadores sociais pouco atraentes para a análise do fenômeno: na hora de utilizarem uma informação acerca de uma realidade social sob análise, as pessoas preferem medidas sucintas, que sejam de fácil interpretação. A esse respeito, Haq (2003, p. 112) menciona o caso específico dos formuladores de políticas públicas, que podem encontrar dificuldades em absorver uma lista de indicadores sociais separados, que apontem em diversas direções.

Essa situação é ilustrada por Sliwiany (1997, p. 15), a partir da consideração de um episódio ocorrido no Brasil. A autora menciona que, por ocasião da elaboração do II Plano Nacional de Desenvolvimento (PND), em 1975, os principais órgãos responsáveis pela formulação de políticas sociais propuseram a constituição de um sistema de indicadores sociais, o que envolveria as etapas de identificação, construção e produção de indicadores referentes a certos campos integrantes da área social. No decorrer do processo, foram apresentados nove grandes temas,¹¹ que se desdobraram em 31 indicadores gerais, os quais se multiplicaram ainda em vários outros; só na área da educação, por exemplo, foi proposto um sistema composto por 53 indicadores. Segundo a opinião da autora, “essas reproduções se tornaram rotina, retalhando o social de tal forma que, ao se conhecer ‘os milhares’ de indicadores existentes, percebe-se que não há indivíduo que neles não se perca” (SLIWIANY, 1997, p. 16).

A partir da observação dessa e de outras situações ocorridas no âmbito brasileiro, a autora fornece uma síntese da questão ora discutida no seguinte parecer:

Historicamente [no Brasil], a predominância da dispersão na análise quantitativa da realidade social em milhares de cálculos específicos, isto é, a desenfreada busca de indicadores sociais (...), que podem até ser traduzidos como a grande moda do conhecimento social, resultou no retalhamento integral da realidade social, e, o que é pior, numa pobreza generalizada de métodos quantitativos eficazes quanto à consistência teórico-técnica, que venham a realizar tantas agregações quanto necessárias para representar o social na sua infinidade de variáveis relacionadas (SLIWIANY, 1997, p. 15-16).

Sobre a referida necessidade de se representar o social com base em diversas variáveis que se inter-relacionam, a metodologia dos sistemas de indicadores sociais oferece a possibilidade de que se realizem cruzamentos entres algumas das distintas dimensões de um fenômeno. Como essa metodologia exprime seus achados sob a forma de tabelas, o número de variáveis que podem ser cruzadas simultaneamente é limitado. Para que se produza um resultado inteligível, só é possível articular, no máximo, indicadores relativos a três dimensões do problema – já que esse tipo de cruzamento só se faz compreensível até os limites da exposição dos resultados numa tabela de três “entradas”. Como diversos fenômenos sociais envolvem mais do que três dimensões, essa característica é ainda insuficiente para que a complexidade do problema seja representada de modo adequado.

Além da limitação do número de dimensões que podem ser cruzadas, a realização dos cruzamentos não permite afirmar as significâncias estatísticas¹² das

11. A saber: população, emprego e remuneração do trabalho; proteção ao trabalhador; orçamentos familiares; educação e preparação da mão-de-obra; saúde; nutrição; saneamento básico; habitação.

12. Relação é confiável ou “verdadeira”. Como muitos estudos são feitos com dados amostrais, a significância permite inferir se a relação observada na amostra deve-se ao acaso ou se tal resultado representa o que acontece na população estudada.

relações apresentadas. Em outras palavras, ainda que exista a tentativa de estabelecer relações entre os indicadores das distintas facetas que compõem o fenômeno estudado, os sistemas de indicadores sociais não dispõem de procedimentos técnicos que permitam afirmar, de modo confiável, que aquelas ligações existem (CONDE, 1999, p. 146). Isso tem a ver com o fato de que tais sistemas caracterizam-se por priorizarem análises de outra natureza, que não focalizam a investigação da articulação entre os fatores: eles se voltam para o tratamento da informação agregada, tal como acontece com os indicadores sintéticos (MEJÍA; VOS, 1999, p. 176).

Nos sistemas de indicadores sociais, dados referentes a realidades extensas costumam ser agregados sob a forma de valores médios, taxas, contagens ou porcentagens das variáveis que estão sendo investigadas. Como discutido anteriormente, o procedimento da agregação envolve alguns inconvenientes para a análise de situações sociais. As críticas derivadas do procedimento da agregação, que foram abordadas na seção dedicada aos indicadores sintéticos, podem ser aplicadas aos resultados que se apresentam em sistemas de indicadores sociais.

Há que se mencionar ainda uma crítica dirigida aos sistemas de indicadores sociais, relativa a aspectos conceituais. Tendo em vista o fato de que tais sistemas são criados com o principal objetivo de se traçar o perfil da população pobre, Blum e Rossi (1969, p. 353) questionam quais dimensões devem ser consideradas relevantes para que se possa fornecer esse retrato da pobreza. Os autores entendem que, mesmo que a discussão se limite aos conjuntos de características em torno das quais existe um consenso mínimo, há dificuldades para que se atinja o objetivo de modo satisfatório. Essas dificuldades ligam-se à questão da ponderação dos fatores envolvidos no fenômeno. Ainda que esse não seja um tópico crucial para a metodologia da criação de sistemas de indicadores sociais, existem debates em torno de quais pesos relativos devem ser atribuídos a cada uma das características que virão a compor o almejado perfil dos pobres.

Como se vê, os métodos que traçam um perfil da população pobre também apresentam limitações para retratar, no âmbito empírico, a visão complexa que se tem sobre o fenômeno atualmente.

3.3 ANÁLISES DE REGRESSÃO

Sendo a técnica de regressão mais elaborada do que as duas anteriores, esta subseção disserta com maior profundidade sobre detalhes dessa forma de análise, a fim de fornecer elementos básicos para a compreensão das suas potencialidades e limitações.¹³ Tal detalhamento é considerado relevante para a argumentação desenvolvida neste trabalho, por também subsidiar o entendimento da metodologia que será abordada a seguir, a MEE. Como se pode ver adiante, a MEE deriva da análise de regressão, acrescentando-lhe feições que a aproximam da análise relacional da pobreza.

Nas ciências sociais, procedimentos de regressão são largamente usados como ferramenta de pesquisa. Em geral, pesquisadores utilizam esse método com a intenção

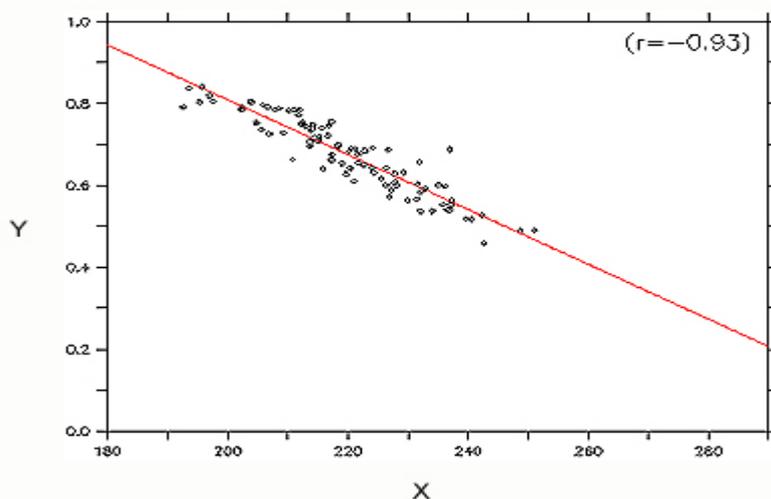
13. Certos trechos desta subseção assumem mesmo um caráter didático. A intenção é instrumentalizar o leitor menos familiarizado com os conceitos da área de métodos quantitativos que são necessários para acompanhar a discussão desenvolvida neste trabalho.

de que se possa responder a perguntas do tipo: “tal aspecto associa-se (ou ajuda a explicar) à variabilidade de tal fenômeno?” Pesquisas educacionais, por exemplo, se interessam em saber quais os fatores que ajudam a melhorar o sucesso escolar dos estudantes. Da mesma maneira, diversos estudos sobre o tema da pobreza também podem se basear na técnica de análise de regressão, no intuito de que se conheçam quais os fatores que exercem impacto sobre as condições de vida dos indivíduos.

A utilização desse procedimento estatístico visa a verificar, com base na manipulação de dados empíricos, se há alguma relação entre duas variáveis. Em caso positivo, estima-se o grau de intensidade da associação entre variáveis, ou seja, o método permite que se avalie se há uma associação fraca ou forte entre elas. A mensuração do grau de ligação entre variáveis indica o efeito que é produzido na variável-resposta, a partir de mudanças ocorridas nos valores das co-variáveis.¹⁴ Assim, o modelo estatístico calcula o valor médio de mudança na variável dependente acarretado pelo acréscimo de uma unidade no valor da variável independente.

Graficamente, a relação entre duas variáveis é representada por uma reta, cuja inclinação expressa o grau de associação entre elas (ver gráfico 1). Sob tal perspectiva, a técnica da regressão prima por ajustar a melhor reta à distribuição dos dados empíricos,¹⁵ fornecendo os dois parâmetros da linha ajustada: intercepto, que designa o valor médio da variável dependente ou, o que é sinônimo, o seu valor quando a covariável assume o valor zero; e o coeficiente de regressão, que denota a inclinação da reta e informa, portanto, o grau de associação entre a variável independente e a variável-resposta; tal parâmetro pode ainda ser interpretado como a magnitude da mudança que ocorre na variável-resposta, quando há o acréscimo de uma unidade na covariável.

GRÁFICO 1
Exemplo de regressão linear simples



14. A variável dependente pode ser também chamada de “variável-resposta”. Da mesma forma, a variável independente pode ser designada pelos termos “co-variável”, “variável explicativa” ou “variável preditiva”.

15. No caso da regressão linear, esse ajuste se dá através do método de estimação dos Mínimos Quadrados. Há, contudo, modelos que requerem que o ajuste aconteça por meio de um outro método de estimação, a dizer, o da Máxima Verossimilhança (como por exemplo, aqueles cuja variável dependente é dicotômica). Para uma abordagem mais aprofundada desse tema, ver Draper e Smith (1998) e Paula (2004).

De uma forma geral, os estudos sobre pobreza utilizam a técnica da “regressão múltipla”, que se caracteriza por estimar relações existentes entre a variável dependente e um conjunto de variáveis independentes. Nesses casos, a intensidade da associação entre a variável-resposta e cada uma das variáveis independentes é calculada, sendo que a idéia central do procedimento é poder medir a capacidade do conjunto de variáveis independentes de explicar a variabilidade da variável-resposta.¹⁶ Para tanto, a regressão múltipla determina uma expressão matemática que descreve a relação funcional entre a variável-resposta e as independentes. Esse procedimento busca definir uma equação que denote a combinação linear entre diversas co-variáveis que melhor possa prever o valor da variável dependente.¹⁷ Em termos genéricos, tal equação pode ser expressa da seguinte forma:

$$Y = a + b_1 * X_1 + b_2 * X_2 + \dots + b_p * X_p + \varepsilon$$

onde:

Y = variável dependente;

X_1, X_2, \dots, X_p = variáveis independentes;

a = intercepto;

b_1, b_2, \dots, b_p = coeficientes de regressão; e

ε = erro aleatório ou resíduo, que designa a distância entre um ponto observado (pertencente aos dados empíricos) e o valor predito pela linha de regressão.

Alguns esclarecimentos devem ser feitos sobre o elemento “resíduo”, apresentado na equação acima. Para isso, há que se retomar a noção de que a técnica da regressão consiste na tentativa de ajustar a melhor reta à distribuição dos dados empíricos. Como se disse, a intenção é que se obtenham as melhores estimativas possíveis. Entretanto, é improvável que tais estimativas sejam exatamente “corretas”, já que existem três fontes de erros inerentes à maioria das aplicações de análises de regressão. Inicialmente, pode-se falar nos erros de mensuração. Conforme já mencionado, muito poucas variáveis podem ser medidas com precisão absoluta, especialmente nas ciências sociais. Segundo, há os erros de amostragem. Em muitos casos, trabalha-se com dados que são apenas uma amostra da população geral e, como se sabe, ela jamais será exatamente igual à população-alvo do estudo. Por fim, modelos de regressão são imbuídos de uma “variação não-controlada”. Esse aspecto advém do fato de que um modelo inclui apenas algumas variáveis independentes, que são consideradas relevantes para explicação da variabilidade do fenômeno estudado; contudo, sempre existem outras variáveis que influenciam a variável-resposta e que não fazem parte do modelo. Por isso, existe uma variabilidade não-

16. A estatística mais utilizada para aferir a capacidade de um conjunto de co-variáveis elucidarem a variabilidade da variável-resposta é chamada de “Coeficiente de Determinação”, mais conhecida por seu símbolo, R^2 . O valor de tal coeficiente indica a porcentagem da variabilidade do fenômeno estudado (representado, dentro do modelo, pela variável-resposta), que é explicada pelo conjunto das co-variáveis inseridas na análise de regressão.

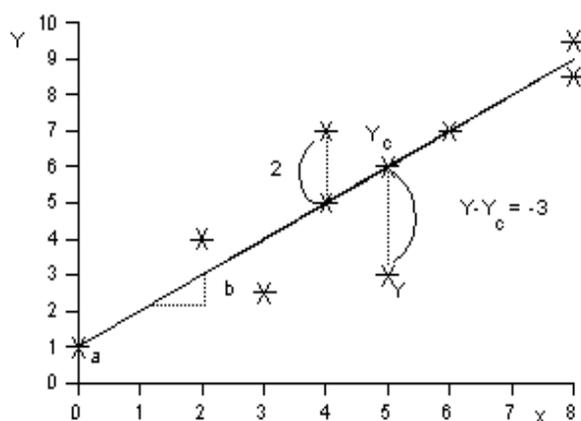
17. Como se trata de uma regressão linear múltipla, a visualização gráfica dessa equação torna-se complicada. Um modelo que inclua p co-variáveis seria representado graficamente em $p + 1$ dimensões. Nessas circunstâncias, o raciocínio requer um alto grau de abstração: seria necessário pensar em termos de hiperespaço (um espaço multidimensional), com a superfície de regressão ajustada através dos pontos (HOWELL, 2002, p. 548).

explicada pelo modelo – ligada às variáveis que lhe são externas – que pode “perturbar” a relação entre a variável dependente e as independentes. Essa variação não-controlada afeta, conseqüentemente, as estimativas dos coeficientes de regressão. Devido a tais motivos, existe uma discrepância entre os dados observados e os valores preditos a partir da equação da regressão – os quais, se representados graficamente, conformam a referida reta. As discrepâncias são os “resíduos” ou “erros aleatórios”.¹⁸

Como se disse, a presença de tais imprecisões em um modelo de regressão gera uma variação aleatória das estimativas dos coeficientes. Em qualquer análise desse tipo, é desejável que se tenha uma estimativa da fidedignidade dos valores calculados dos coeficientes. Portanto, a dimensão e a importância daquela variação devem ser avaliadas. Essa avaliação pode ocorrer através do cálculo de intervalos de confiança ou de testes de hipóteses.

GRÁFICO 2

Exemplo de representação gráfica dos resíduos em uma reta de regressão



Intervalos de confiança consistem na apresentação de uma faixa de valores possíveis para os coeficientes de regressão. Embora não se possa assegurar que seus valores “verdadeiros” cairão na faixa calculada, pode-se estar razoavelmente confiante disso.¹⁹ O cálculo do intervalo de confiança para um coeficiente pressupõe que se estime o seu erro-padrão.²⁰ Assim, no caso do cálculo de um intervalo de 95% de confiança, seu limite superior é determinado a partir da adição do valor de dois erros-padrão à média estimada para o coeficiente de regressão; de modo análogo, seu limite inferior é calculado quando se subtrai o mesmo montante da média do coeficiente (MORETTIN, 2000, p. 49; HOWELL, 2002, p. 206).

18. Vale esclarecer que existe um arsenal de técnicas próprias para a realização da análise dos resíduos. Tal análise é importante dentro do processo de verificação do ajuste do modelo de regressão aos dados empíricos.

19. Nos estudos sociais, é aceito como razoável o intervalo de 95% de confiança. Isso quer dizer que a probabilidade de que o coeficiente “verdadeiro” situe-se nesse intervalo é de 95%.

20. O erro-padrão mede a precisão de uma estimativa. No contexto em questão, está se estimando o coeficiente de regressão; assim, o seu erro-padrão indica a precisão que pode ser atribuída ao valor que for calculado. Em termos técnicos, o erro-padrão é definido como o desvio-padrão da distribuição amostral de uma “estatística”, ou seja, da estimativa de um parâmetro de um modelo ou de uma população.

Outra forma de avaliar a importância da variação randômica das estimativas dos coeficientes é por meio dos testes de hipóteses. Em uma análise de regressão, parte-se da assunção de que a inexistência de relação entre duas variáveis se expressa através de um coeficiente igual a zero. Entretanto, ao se calcular um modelo dessa natureza, é comum que não se obtenha um coeficiente de valor zero, mas que se encontrem valores muito pequenos. A questão que se coloca, então, é determinar se tais coeficientes designam a inexistência de relação entre as duas variáveis consideradas ou se eles se devem às recém-mencionadas fontes de erros aleatórios.

Nesse contexto, testes de hipóteses são usados para responder à seguinte questão: “essa variável independente realmente afeta a variável-resposta?” Assim, a realização de um teste consiste em negar a hipótese de que o coeficiente de regressão é zero. Novamente, não se pode ter uma resposta definitiva sobre isso, mas é possível calcular a probabilidade de se estar errado. Dito de outro modo, o teste não fornece uma resposta do tipo “sim” ou “não”, mas indica a probabilidade, usualmente denominada “valor-p”, de que aquele coeficiente seja diferente de zero em decorrência de erros aleatórios.²¹ Se o valor-p é pequeno, nega-se a hipótese de que o coeficiente é zero.

Ainda com relação à obtenção dos coeficientes de regressão, que indicam a intensidade das relações entre variáveis, deve-se comentar que no caso da regressão múltipla, os coeficientes são calculados de maneira “controlada” estatisticamente ou, dito de outra maneira, ao calcular a intensidade de associação entre duas variáveis, a regressão múltipla “controla” as demais variáveis simultaneamente. Isso significa que o efeito produzido por uma determinada co-variável sobre a variável dependente é examinado isoladamente das demais co-variáveis, como se elas fossem mantidas constantes. Assim, produz-se apenas uma única estimativa para o efeito de cada co-variável, independentemente do comportamento das demais.²²

Essa propriedade do método da regressão evoca um comentário comparativo com a metodologia dos sistemas de indicadores sociais. É possível que dados sobre uma mesma realidade social, analisados por ambos os métodos, produzam efeitos aparentemente inconsistentes. A razão para isso reside no fato de que a regressão controla os níveis das outras variáveis, o que não acontece com os sistemas de indicadores sociais.

Um exemplo dessa situação é encontrado no estudo realizado por Datt *et al.* (1998), que analisa dados sobre a população pobre de Moçambique a partir da

21. Não se tem a intenção de entrar aqui nos detalhes técnicos para o cálculo do valor-p, mas importa mencionar que ele é obtido a partir da divisão do coeficiente de regressão pelo seu erro-padrão. Gera-se assim um valor denominado “estatística-t”, o qual possui um valor-p associado. Se o erro-padrão for alto, acontece uma diminuição do valor da estatística-t e um conseqüente aumento do valor-p ou da probabilidade de que a relação entre as variáveis seja devida ao acaso. Note-se a relevância do erro-padrão tanto para o teste de hipóteses como para a determinação de intervalos de confiança. Em ambos os casos, a diminuição da sua magnitude contribui para aumentar a precisão das estimativas da análise de regressão. Para mais detalhes, ver Morettin (2000) e Moore (2004).

22. A título de esclarecimento, vale mencionar que o coeficiente produzido para cada co-variável é análogo à média ponderada dos seus efeitos calculados em diferentes subgrupos da amostra. Esses subgrupos são definidos em função das demais co-variáveis que lhe compõem.

construção de sistemas de indicadores sociais e de análises de regressão.²³ Esse estudo aponta que, com base na construção de um perfil de pobreza, as famílias chefiadas por mulheres tendem a ser menos pobres do que as chefiadas por homens, de acordo com certos parâmetros escolhidos pelos autores para medir a pobreza. O resultado da análise de regressão, por sua vez, sugere a situação contrária. Como argumentam os autores, essa contradição deve-se ao fato de que os coeficientes de regressão, que estimam a associação entre sexo do chefe de família e pobreza, são calculados em circunstâncias em que as demais co-variáveis inseridas no modelo são mantidas constantes. Em outras palavras, o efeito calculado do fator sexo sobre a pobreza é “líquido”, por ser estimado independentemente das demais variáveis explicativas do modelo (o número de membros da família, a quantia de terra cultivável, níveis de educação, entre outras).

Ao investigar a relação entre sexo do chefe de família e condições de pobreza, o método do sistema de indicadores sociais, por sua vez, não expurga as influências exercidas pelas demais covariáveis.²⁴ Como as famílias chefiadas por homens e mulheres apresentam características distintas quanto às demais variáveis explicativas consideradas no estudo, os resultados relativos ao fator sexo, fornecidos por tais sistemas, estão imbuídos daquelas outras influências. Dessa forma, tais co-variáveis podem vir a confundir o efeito do sexo do chefe de família, quando se tem em vista a realização de uma análise bivariada da pobreza.

Os autores, entretanto, não descartam a utilidade dos sistemas de indicadores sociais para o propósito do desenvolvimento de políticas públicas. Segundo eles, políticas mais gerais, que visem a assistir famílias pobres independentemente de suas características específicas, podem se pautar nos resultados fornecidos pelo perfil de pobreza, já que os achados dessa metodologia traçam uma visão mais genérica do problema, por não estarem condicionados a aspectos específicos do objeto estudado. Entretanto, se o propósito da intervenção for o de atuar sobre um fator associado a padrões de vida mais baixos, aqueles identificados em uma análise de regressão servem como indicações mais claras e confiáveis para a formulação dessas ações.

Há diversos outros estudos sobre pobreza que utilizam a técnica da análise de regressão, tanto no âmbito internacional como no nacional. Dentre os trabalhos realizados mais recentemente, pode-se citar aquele que foi desenvolvido por Menezes-Filho e Vasconcellos (2004) acerca da realidade brasileira. Esse trabalho faz parte de uma série de investigações sobre pobreza em diversos países do mundo, intitulada *Operationalising Pro-Poor Growth (OPPG)*.²⁵

23. Esse estudo faz parte de uma série de trabalhos sobre condições de pobreza e bem-estar em diversos países do mundo, promovida pelo Instituto Internacional de Pesquisa em Políticas Alimentares (IFPRI) – International Food Policy Research Institute –, órgão sediado em Washington D. C.

24. Devido à ausência desse tipo de controle, a natureza de tal procedimento analítico é conhecida como “incondicional” dentro do jargão metodológico.

25. Operacionalizando o Crescimento a Favor dos Pobres é uma iniciativa conjunta, patrocinada pelos seguintes organismos internacionais: AFD, BMZ (GTZ, KfW Development Bank), DFID, e o World Bank. O programa de trabalho da OPPG declara o objetivo de prover subsídios ao desenvolvimento de políticas públicas que facilitem a participação das pessoas pobres no processo de crescimento. Dentro dessa série de 14 estudos de casos sobre países, que compreende regiões da África, Ásia, América Latina e Europa Oriental, a investigação realizada por Menezes-Filho e Vasconcellos (2004) encontra-se sob o título de *A country case study on Brazil (Um estudo de caso nacional sobre o Brasil)*. O

Tal estudo sobre o Brasil sustenta-se na elaboração de vários modelos de regressão. Em geral, considera-se como variável dependente o fato de a família ser pobre ou não-pobre. O objetivo é que se detectem fatores que aumentem ou diminuam a probabilidade de uma família encontrar-se em situação de pobreza, estimando-se a magnitude de cada associação. As análises revelam que a pobreza no Brasil se mostra associada, entre outros aspectos: à presença de crianças no lar, aos não-brancos, aos níveis mais baixos de educação, à falta de acesso a infra-estrutura, ao desemprego ou ao trabalho informal.

No decorrer do estudo, os pesquisadores consideram que as diversidades existentes entre os contextos rural e urbano justificam a realização de análises separadas para cada uma dessas situações. São então rodadas duas regressões distintas, uma voltada à análise dos dados das zonas rurais e outra baseada em dados de zonas urbanas. Nessa investigação, os fatores educação, acesso a infra-estrutura e setor de atividade destacam-se sobremaneira nas regressões que focalizam as áreas rurais (MENEZES-FILHO; VASCONCELLOS, 2004).

Também utilizando a técnica da regressão para analisar realidades de pobreza, coloca-se a iniciativa do Mapa do Fim da Fome II. A partir de dados sobre o Estado do Rio de Janeiro, o referido trabalho efetua análises de regressão com dados de 91 municípios. Para compor os modelos, considera-se como variável dependente renda familiar *per capita*.²⁶ Em princípio, esse trabalho tenta explicar a variabilidade da renda familiar *per capita* a partir da inclusão de apenas uma variável independente, o município. Com isso, consegue-se explicar 10,61% da variabilidade da renda da população ($R^2 = 0,1061$), ou seja, há a indicação de que 10,61% da variabilidade da renda associam-se à diversidade de contextos municipais. Em seguida, são adicionadas outras variáveis explicativas: tamanho da cidade, nível educacional, idade, sexo e raça do chefe do domicílio. É então construído um novo modelo, que demonstra maior potencial de elucidação do fenômeno, explicando 37,80% da variabilidade da renda familiar *per capita* dos cariocas ($R^2 = 0,3780$). Em tal análise, as variáveis raça, cidades e tamanho das cidades são as que se mostram mais fortemente associadas à variável dependente (CPS/FGV, 2004).

No âmbito desse mesmo estudo, é desenvolvida uma análise semelhante para a cidade do Rio de Janeiro. De modo análogo ao que acontece para o nível estadual, busca-se inicialmente explicar a variabilidade da renda familiar *per capita* em função apenas das

conjunto completo de estudos de casos dos países pode ser encontrado nos *websites* das organizações participantes: BMZ (www.bmz.de), DFID (www.dfid.gov.uk), GTZ (www.gtz.de), KfW Development Bank (www.kfw-entwicklungsbank.de/EN/Fachinformationen) e o Banco Mundial (www.worldbank.org).

26. Note-se que, nessa iniciativa, os pesquisadores buscaram captar os fatores que influenciam a variação da renda de toda a população, e não apenas a distribuição da renda no universo das camadas pobres. Assim, pode-se dizer que o objetivo de estudar os pobres encontra-se diluído, uma vez que os resultados fornecidos não se concentram nos fatores associados à pobreza propriamente, mas se referem às características ligadas à renda da população como um todo. Some-se a isso o fato de que, como se sabe, as camadas mais abastadas da população obtêm parte significativa de suas rendas por meio de outros fatores que não se mostram relevantes em contextos de pobreza, tais como proventos de aluguéis, investimentos financeiros, heranças, entre outros. A esse respeito, Medeiros (2003) analisa a realidade brasileira e encontra indicações de que a riqueza se origina, ou ao menos é perpetuada, por políticas de infra-estrutura, de investimento e políticas sociais que não possuem caráter redistributivo. Assim, como os modelos não contemplam tais fatores, é possível que haja uma redução do potencial explicativo da renda *per capita* da população, uma vez que a análise considera pessoas de todos os níveis socioeconômicos.

32 regiões administrativas da cidade. Nesse caso, é explicada uma proporção de 24,45% da variabilidade do fenômeno ($R^2 = 0,2445$). Quando as características sociais do chefe da família são adicionadas ao modelo, o potencial de explicação aumenta para 43,42% ($R^2 = 0,4342$), sendo que a raça é a variável que se mostra mais relevante, desfavoravelmente aos negros.

Entre os esforços investigativos empreendidos no Mapa do Fim da Fome II, os mesmos cálculos ora mencionados são repetidos, tomando-se como base apenas a PIA, pessoas com idades situadas no intervalo de 15 a 65 anos. Os resultados obtidos nessas novas análises não diferem dos anteriores de maneira pronunciada, nem no nível do estado, nem no de sua capital. Para o Estado do Rio de Janeiro, as regressões rodadas com a PIA apresentam poder de explicação de 10,74% (modelo que considera os municípios como única variável independente) e de 40,72% (quando são acrescentadas as demais variáveis explicativas). No caso da capital, essas novas regressões apresentam capacidade de explicação de 24,24% e de 45,98% para os modelos com uma e com várias co-variáveis, respectivamente, conforme já mencionado.²⁷

A partir da observação dos estudos aqui mencionados, chama a atenção a flexibilidade das análises propiciadas pela técnica de regressão. Isso pode ser constatado na medida em que tal procedimento comporta diversas possibilidades de contemplação de uma mesma realidade social. Em termos mais específicos, essa flexibilidade advém do fato de que se podem construir diferentes modelos que tentem explicar uma mesma variável dependente. As distinções entre eles residem na possibilidade de se usarem diferentes quantidades e tipos de variáveis explicativas. Essa diversidade pode ajudar a acrescentar novas nuances à compreensão do fenômeno estudado.

Pode acontecer ainda que modelos formalmente similares lancem luzes distintas sobre uma mesma realidade sob análise. Isso pode ser feito com base na utilização de dados referentes a subgrupos específicos da população estudada, desde que se tomem os devidos cuidados para que se tenha representatividade amostral.²⁸ Tal possibilidade pode ser notada nos exemplos mencionados, em que foram construídas regressões separadas com dados urbanos e com dados rurais, ou ainda utilizando-se uma faixa etária específica da população estudada.

Apesar de todas as potencialidades oferecidas pelo método da regressão para a análise de fenômenos sociais, há que se ter cautela com algumas características que lhe são inerentes. Em princípio, pode-se comentar acerca da questão do peso atribuído a cada correlação. Segundo tal procedimento, os valores das relações entre variáveis

27. Deve-se esclarecer que os coeficientes de determinação (R^2) não são comparáveis entre modelos formados por diferentes variáveis explicativas. Entretanto, a comparação entre os coeficientes de determinação é válida se estiver se tratando de modelos "aninhados" – diz-se que um modelo está "aninhado" em outro quando ele é composto por um subgrupo de co-variáveis presentes naquele outro modelo (DRAPER; SMITH, 1998, p. 246). Note-se que as comparações entre os R^2 dos distintos modelos calculados pelo Mapa da Fome II foram realizadas dentro dessas circunstâncias.

28. Essa preocupação emerge do fato de que a constituição de uma amostra baliza-se em parâmetros relativos às variáveis segundo as quais os resultados serão reportados à população que se quer representar. Por isso, tendo-se uma amostra que foi selecionada de acordo com certos propósitos, não é recomendável retalhá-la sem que se considerem os seus critérios de formação, sob pena de que os resultados obtidos a partir disso percam em validade e precisão. Para mais detalhamento sobre os princípios básicos da amostragem, ver Kish (1995) e McClave, Benson e Sincich (2004).

emergem a partir da análise dos dados, ou seja, não há qualquer forma de atribuição externa, por parte do pesquisador, de pesos às relações estabelecidas no modelo. Assim, a força de cada associação reflete aquilo que é constatado na realidade empírica. Esse aspecto exime a técnica da regressão das discussões sobre a imputação arbitrária de ponderações, uma fragilidade apontada nos métodos anteriormente abordados.

Entretanto, há que se ter cuidado com relação ao viés que pode ser produzido nas estimativas dos coeficientes de regressão a partir da análise dos dados. Em condições em que se produza um viés de grande magnitude, pode-se concluir erroneamente que uma co-variável exerce um efeito que, em realidade, ela não tem. Na situação oposta, em que se tenha um viés para menos, está-se sujeito à conclusão de que uma covariável não produz um efeito sobre a variável-resposta, quando ele deveria estar sendo indicado. Na linguagem dos cientistas sociais, diz-se que tais conclusões são “espúrias”.

Com isso, observa-se que, apesar de a técnica de regressão eliminar a preocupação sobre a arbitrariedade da atribuição de pesos, ela comporta a possibilidade de produzir estimativas viesadas. Usualmente, esse problema é contornado com base na seleção criteriosa das variáveis que entrarão no modelo. Isso se deve ao fato de que o viés²⁹ é gerado pela omissão das chamadas “variáveis-controle”, aquelas que têm efeito sobre a variável dependente e que, ao mesmo tempo, estão correlacionadas com as demais co-variáveis. Por causa desses dois tipos de associação, a ausência das variáveis-controle pode induzir à produção de vieses nas estimativas da regressão.³⁰

Como se vê, a elaboração de um modelo de regressão depende da escolha adequada das variáveis que o compõem. É importante que se incluam as variáveis-controle e aquelas que, por indicações teóricas, têm efeito sobre a variável dependente. Essa questão evoca a discussão sobre os conceitos e dimensões a serem contemplados pelo modelo. Da mesma forma como se discutiu nos métodos anteriormente abordados, é necessário que se considerem aspectos substantivamente relevantes para a análise do fenômeno estudado. Para tanto, deve-se ter uma compreensão sobre as relações entre os conceitos que irão participar da análise (HOWELL, 2002, p. 563).

Contudo, no âmbito da técnica da regressão, esse procedimento pode não ser suficiente para que se construa um modelo adequado. A complexidade dos fenômenos estudados pode incluir situações em que existam variáveis que medeiam os efeitos de outras. Esses casos referem-se à existência de variáveis que exercem seus impactos de forma indireta, o que, muitas vezes, pode ser indicado pela teoria. Assim, se as

29. Diz-se que uma estimativa é “viesada” quando o seu valor se afasta sistematicamente da magnitude que, de fato, deveria assumir. A presença de um viés na estimação de um coeficiente faz com que o valor calculado seja maior ou menor do que sua magnitude “real”, de forma que se pode estar superestimando ou subestimando uma característica do objeto estudado.

30. Uma breve explicação sobre o processo da produção de vieses por causa da exclusão de variáveis-controle pode ser dada da seguinte forma: uma vez que a variável-controle associa-se simultaneamente a uma covariável e à variável-resposta, a sua ausência no modelo pode fazer com que o seu efeito sobre a variável-resposta manifeste-se através da outra covariável, a qual passa a ter o coeficiente viesado.

mediadoras forem incluídas em um modelo de regressão, o efeito que deveria ser atribuído àquela que é mediada pode desaparecer. Em outras palavras, os resultados do modelo induziriam à conclusão de que uma determinada variável não exerce efeito sobre a dependente, quando, na verdade, o seu impacto de natureza indireta estaria sendo manifestado pela variável mediadora ou interveniente. Esse problema decorre do fato de que ambas são postas em condições semelhantes no modelo de regressão: as duas são igualmente consideradas como variáveis independentes. Como se vê, tal método apresenta limitações para lidar com situações como essa.

Outra questão inerente à regressão múltipla refere-se às relações existentes entre as co-variáveis que compõem um modelo. Ao se inserirem diversas variáveis explicativas, tem-se a intenção de que cada uma delas traga alguma contribuição para a explicação do fenômeno estudado. Isso significa que é desejável que covariáveis sejam relativamente independentes entre si. Caso contrário, se duas co-variáveis forem altamente correlacionadas, não se estará acrescentando mais luzes à compreensão da situação estudada; sob tais circunstâncias, bastaria incluir apenas uma delas (HOWELL, 2002, p. 563). O problema é que, quando há muitas variáveis envolvidas no fenômeno, as altas correlações podem não ser aparentes de forma imediata e acabam sendo incluídas no modelo.

A questão da alta correlação entre variáveis independentes é denominada multicolinearidade. Além de trazer informações redundantes à análise, multicolinearidade cria complicações à estabilidade do modelo e à precisão das estimativas dos coeficientes. Isso decorre do fato de que altas correlações entre covariáveis aumentam o valor do erro-padrão dos coeficientes.³¹

Conforme já se colocou, quanto mais alto o erro-padrão de um coeficiente, maior será a flutuação do seu valor estimado e, conseqüentemente, menos confiança pode-se ter nessa estimativa. Por isso, é provável que as estimativas dos coeficientes não se mostrem significativas, o que pode conduzir à conclusão de que as co-variáveis não possuem efeito sobre a variável-resposta, quando, em realidade, tal efeito pode existir e ser relevante. Por todos esses motivos, é desejável que se eliminem variáveis do modelo que sejam altamente correlacionadas entre si (HOWELL, 2002, p. 564).

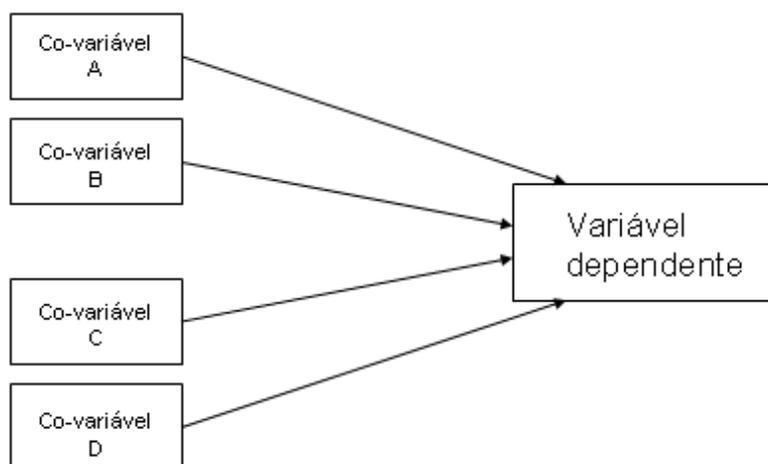
As questões ora mencionadas – a dizer, a necessidade de inserção de variáveis-controle, a existência de efeitos indiretos e a multicolinearidade – são aspectos que se ligam à estrutura das relações existentes entre as variáveis inseridas em um modelo. Como se pode notar a partir das considerações feitas até aqui, no âmbito da técnica da regressão, tais questões constituem-se em problemas a serem enfrentados no decorrer do processo de construção dos modelos e podem comprometer os resultados produzidos. Por apresentar essas dificuldades, pode-se dizer que o método da regressão possui baixo potencial de lidar

31. Como já se explicitou, para estimar a associação entre uma covariável e a variável dependente, a técnica da regressão assume que as demais covariáveis são mantidas constantes, controladas. Se há alta correlação entre covariáveis, não é isso o que acontece: em vez de permanecerem inalteradas, elas também variam. Assim, há um desencontro entre a assunção metodológica e o que, de fato, ocorre com os dados empíricos. Devido a essa falta de correspondência entre os pressupostos do método e o comportamento dos dados, não são calculadas estimativas precisas e estáveis do efeito exercido por cada covariável sobre a variável dependente. Essa falta de precisão e estabilidade reflete-se na elevação do valor do erro-padrão.

com a estrutura das inter-relações existentes entre as variáveis envolvidas em um fenômeno.

Nesse sentido, as limitações são mais acentuadas quando se trata das associações existentes entre as variáveis explicativas. Esse método examina as relações entre cada uma das variáveis independentes e a dependente, mas não se debruça sobre a natureza das associações que se estabelecem entre as co-variáveis. Sua proposta é isolar a influência direta exercida por cada uma delas, controlando para as demais, mas não se intenta capturar a forma como elas se articulam. No esquema do gráfico 3, encontra-se uma ilustração da estrutura de relações entre as variáveis, assumida pela modelagem da regressão.

GRÁFICO 3
Estrutura das relações entre variáveis, assumida pela análise de regressão



Conforme mencionado, o problema reside no fato de que o procedimento da regressão consiste em incluir todos os aspectos na condição indistinta de co-variáveis, a despeito da posição que eles ocupam na estrutura das relações que conformam um fenômeno social. Nota-se então que, por não ter capacidade de refletir a articulação dos fatores envolvidos em um fenômeno complexo, o método da regressão possui baixo potencial de análise relacional de uma situação social.

Há que se comentar ainda sobre um aspecto relativo à questão da agregação de dados. Diferentemente das metodologias recém-abordadas, a análise de regressão se baseia na utilização de microdados, aqueles que apresentam a informação em seu mais baixo nível de desagregação. A vantagem dessa característica é que ela faz da regressão um procedimento sensível às distribuições dos dados. O comportamento de cada variável é considerado na análise, de forma que os já referidos inconvenientes advindos do procedimento de agregação não se apresentam nos resultados fornecidos pela análise de regressão.

Embora não seja intrínseco à técnica da regressão, o procedimento da agregação pode ser encontrado em diversos modelos desse tipo. Alguns fatores confluem para que isso aconteça. Primeiramente, sabe-se que os problemas sociais caracterizam-se por serem complexos e multidimensionais; muitas vezes, não existe uma única variável observável que represente adequadamente um conceito social que se queira

estudar. A metodologia da regressão, por seu turno, admite que se inclua apenas uma variável dependente na construção de cada modelo. Dessa forma, é usual a prática de se criarem “compósitos”, que são medidas únicas que representam um conceito multidimensional. A idéia é que o compósito agregue em si uma gama de variáveis distintas, que se relacionam a um único conceito a ser incluído na análise. Assim, a criação de um compósito acontece por meio de procedimentos de cálculo que aglutinam as informações relativas a cada aspecto que compõe aquele conceito, à semelhança do que ocorre com a elaboração dos índices sintéticos. Em geral, os compósitos são feitos com a intenção de serem incluídos nas modelagens estatísticas, recebendo o mesmo tipo de tratamento dado às variáveis observáveis.

Por ser de natureza semelhante à do indicador sintético, tal procedimento traz à análise os já comentados problemas inerentes àqueles índices. Uma vez que um compósito tende a ser calculado como a média de diversas variáveis, pode-se falar em problemas ligados à opacidade presente nessa medida, à falta de precisão e mesmo à distorção de estimativas finais geradas pelo modelo. Como uma tentativa de expandir as possibilidades de inclusão de conceitos sociais complexos na análise, os compósitos podem também ser encontrados nos lugares das variáveis independentes, apresentando os mesmos tipos de problemas.

Outra instância em que se pode encontrar o procedimento de agregação de dados em análises de regressão refere-se à agregação de dados por área, quando se está tratando de variáveis próprias dos indivíduos. Nesses casos, em vez de as regressões serem calculadas com informações individuais, são utilizadas as médias referentes a uma área, como um bairro, uma cidade etc. Conforme já discutido, o processo de agregação de dados sempre implica a perda de detalhamento das informações, particularmente no que se refere à forma de distribuição dos dados. No caso da análise de regressão, quando se intenta reportar resultados referentes a variáveis próprias do nível do indivíduo e não do das áreas, essa perda de informações pode gerar distorções nos coeficientes de correlação entre variáveis. Trata-se do problema denominado “falácia ecológica”, que se caracteriza por aumentar o valor das correlações calculadas e por diminuir as flutuações estatísticas. Assim, por causa da falácia ecológica, apresentam-se conclusões impróprias para o nível individual, em decorrência do fato de que os cálculos foram feitos com dados agregados por área (DIAS *et al.*, 2002, p. 93). Apesar de não ser recomendável, algumas análises de regressão são realizadas com base em dados agregados por área.

3.4 AS METODOLOGIAS TRADICIONAIS E AS CARACTERÍSTICAS DESEJÁVEIS PARA O ESTUDO RELACIONAL DA POBREZA

As três metodologias expostas até aqui são as mais recorrentes nos estudos de pobreza. Elas caracterizam-se por abarcarem um número de fatores envolvidos no fenômeno, na tentativa de refletirem a multidimensionalidade inerente a essa questão social. Os resultados gerados por cada uma delas apresentam potencialidades próprias, que ajudam a compreender o fenômeno sob determinados pontos de vista. Entretanto, conforme foi abordado, tais técnicas apresentam também limitações analíticas para o estudo relacional da pobreza. Para aprofundar esse ponto de vista, a presente subseção confronta as três técnicas com aqueles atributos que capacitam um método a apreender,

no estágio da análise dos dados empíricos, o caráter relacional e complexo do objeto social. Já tendo sido anteriormente discutidas, tais propriedades desejáveis podem ser sintetizadas da seguinte forma:

1) É necessário *superar os limites da simples quantificação dos aspectos relativos à pobreza*, buscando-se a apreensão da articulação entre as múltiplas facetas que conferem a complexidade ao fenômeno estudado.

2) Deve ser possível *estabelecer as relações entre os diferentes aspectos envolvidos no problema*, tais como antecipadas na teoria. Isso inclui a contemplação não só das relações entre os fatores socioeconômicos e as medidas de pobreza, mas também da forma como aqueles fatores se articulam entre si. Com isso, abre-se espaço para que se possam analisar os efeitos diretos e indiretos de cada fator sobre as condições de vida dos indivíduos.

3) É desejável que se possa *ter uma representação gráfica do fenômeno*. Essa representação consiste em um diagrama que fornece uma visão pictórica das relações ou “linhas de força” que conectam os diversos elementos que compõem o fenômeno. A intenção é que uma representação dessa natureza venha a dar uma visão conjunta do objeto estudado, de modo a ajudar a compreensão do seu caráter relacional.

4) É importante que o método ajude a *determinar que fatores se mostram mais influentes nas condições de vida das pessoas*, com base em indicações das realidades empíricas.

5) Deve-se poder também *valorar a intensidade de tais influências, tanto direta como indiretamente*. Conhecimentos dessa natureza podem ajudar a compreender como a pobreza se distribui entre aqueles por ela atingidos, permitindo também a realização de comparações entre as condições de vida dos próprios pobres.

6) É desejável que a metodologia apresente a habilidade de *explicar a variabilidade do fenômeno*. Em termos menos técnicos, pode-se dizer que esse tipo de conhecimento informa o potencial de explicação do fenômeno da pobreza, por parte de um conjunto de fatores considerados em um estudo. Ou ainda: essa habilidade refere-se ao poder de se determinar qual proporção da variação das condições de vida das pessoas pobres está associada aos fatores incluídos na análise, quando tomados conjuntamente.

7) Um método deve poder *incluir conceitos sociais complexos e abstratos de forma estável*. Como se sabe, uma das dificuldades com que lida o pesquisador sobre pobreza é a existência de ligações desse fenômeno com outros processos sociais abstratos, não passíveis de mensuração através de variáveis diretamente observáveis. Assim, é desejável que um método propicie que tais processos complexos e abstratos sejam inseridos na análise de forma adequada.

8) Por fim, é importante que uma metodologia seja capaz de *avaliar os erros de mensuração das variáveis*, para que os resultados produzidos sejam mais confiáveis. A idéia é que a confiabilidade das variáveis possa ser levada em consideração durante os cálculos, de forma que se possam minorar as distorções exercidas por tais erros nos resultados finais das análises.

No que concerne à *capacidade de superar os limites da simples quantificação dos aspectos relativos à pobreza*, pode-se dizer que os índices sintéticos e os sistemas de indicadores sociais não satisfazem a esse atributo, uma vez que estão calcados na agregação de dados. Um problema inicial a ser resgatado é que, por consistirem em valores agregados dos aspectos socioeconômicos pesquisados, os resultados oferecidos por tais métodos são números ou coeficientes derivados de processos de contagem ou de algum outro tipo de quantificação. Essas quantificações são vistas como insuficientes para o estudo relacional da pobreza, devendo ser apenas um estágio inicial das pesquisas. A idéia é que as investigações prossigam rumo a um entendimento mais profundo do fenômeno estudado, fornecendo conhecimentos sobre a identificação dos fatores que afetam a probabilidade de as pessoas viverem em situações de carências e sobre a intensidade da influência de cada um desses fatores. Da mesma forma, a partir da pressuposição de que existem realidades heterogêneas de privações dentro da parcela desfavorecida da população, são valorizados saberes relativos à forma como os graus de pobreza se distribuem nessas camadas. Como esses propósitos analíticos não podem ser atingidos por meio de agregação de dados, os índices sintéticos e os sistemas de indicadores sociais são de pouca contribuição para que se obtenham esses tipos de conhecimento.

A técnica da regressão comporta-se de maneira diferente em relação a esse atributo da quantificação. Ela não se limita a oferecer como resultado um índice que sumarie aspectos da pobreza, mas calcula suas estimativas segundo procedimentos sensíveis à distribuição dos dados. Com isso, geram-se coeficientes que expressam a maneira como as condições de privações variam dentro da camada desfavorecida da população, segundo as características socioeconômicas escolhidas para compor o estudo.

Essa potencialidade da regressão remete à discussão sobre a segunda propriedade desejável em um método relacional, qual seja: a *capacidade de estabelecer relações entre os diferentes aspectos envolvidos no problema*. No método da análise de regressão, o estabelecimento de relações entre variáveis se dá apenas parcialmente, pois cada uma das co-variáveis associa-se à variável dependente de forma isolada. Conforme discutido anteriormente, os modelos de regressão possuem uma estrutura única (ver gráfico 3), comum a todos eles, independentemente de quais sejam suas variáveis constituintes ou da estrutura do fenômeno que queiram representar. Essa disposição de elementos evidencia justamente uma limitação inerente à regressão, qual seja: não contemplar as interligações de todos os fatores. É a partir da análise de tais interligações que emerge a possibilidade de se conhecerem os efeitos que cada variável incluída na análise exerce sobre as condições de vida, tanto direta como indiretamente. Como a análise de regressão se constrói a partir da imposição de uma estrutura rígida sobre os elementos que compõem o estudo, não há espaço para que a configuração das relações entre os diversos fatores associados à pobreza venha a ser considerada na análise dos dados. Assim, não existe a possibilidade de que a complexidade advinda da articulação daqueles fatores seja incorporada num modelo dessa natureza. Essas constatações permitem afirmar que essa técnica não espelha, na etapa de investigação dos dados empíricos, o sistema de proposições teóricas que serve de guia a uma investigação sobre a pobreza.

No caso dos indicadores sintéticos, o seu procedimento de construção tampouco parte do princípio de que as relações estruturais entre fatores envolvidos no fenômeno, tais como postuladas pela teoria, devem ser refletidas na própria organização do indicador. Por isso, as facetas do fenômeno são dispostas de forma indistinta dentro de sua estrutura; um indicador sintético aglutina todos os elementos como se eles fossem de natureza semelhante. Ou seja, desconsidera-se o fato de que podem estar sendo tratados aspectos relativos à mensuração da pobreza em si (como a renda e o acesso a serviços públicos) ou à mensuração de fatores associados (como educação, saúde, emprego).³² Não há espaço, portanto, para que essa metodologia venha a refletir um sistema de proposições teóricas que descreva a estrutura complexa produzida pela articulação dos diversos fatores envolvidos no fenômeno estudado.

Por fim, a metodologia dos sistemas de indicadores sociais também recebe críticas nessa direção. Como visto, esses sistemas oferecem dificuldade de interpretação de objetos sociais complexos, uma vez que apresentam seus resultados de forma bastante fragmentada. Além de complicar a análise, a desconexão entre os vários indicadores expostos nas tabelas faz com que tal metodologia tenha um baixo potencial de abordagem relacional do objeto social estudado, uma vez que é capaz de oferecer apenas uma visão parcial das articulações existentes na realidade social sob estudo. E outro ponto reforça a noção de que essa metodologia possui pouca capacidade de apreensão relacional do fenômeno. Ainda que se realizem cruzamentos entre variáveis, os sistemas de indicadores sociais nada afirmam acerca das significâncias estatísticas das relações entre as mesmas. Por isso, não se podem fazer inferências sobre a existência daquelas ligações na população estudada.

Diante do exposto, nota-se que nenhum dos três métodos tradicionais tem a potencialidade de *representar graficamente o objeto social estudado*. Quanto a esse terceiro atributo, os indicadores sintéticos caracterizam-se pela opacidade dos seus resultados. Os estudos que se baseiam apenas em índices dessa natureza não propiciam uma perspectiva geral das ligações existentes entre “privações” (representadas por indicadores de condições de vida) e outros indicadores socioeconômicos, assim como não fornecem uma visão sobre as relações mútuas entre estes indicadores.

No caso dos sistemas de indicadores sociais e da análise de regressão, essas representações pictóricas são também inviabilizadas. Em se tratando dos sistemas de indicadores, a excessiva fragmentação com que os resultados são apresentados acaba por impedir a construção de uma representação gráfica da forma como se conectam os fatores associados à pobreza. A análise de regressão, por sua vez, ao assumir uma estrutura única de relações entre fatores, tampouco abre espaço para que se possa emergir um diagrama sobre a articulação dos fatores.

Entre os atributos desejáveis em um método de análise relacional, o quarto e o quinto acima enumerados referem-se, respectivamente, à *capacidade de determinar os fatores mais influentes nas condições de vida* e à *de valorar a intensidade de tais*

32. Essa mesma crítica referente ao fato de que não se levam em conta questões substantivas, relativas à forma como os fatores se estruturam dentro do processo estudado, é colocada em termos metodológicos por Jannuzzi (2001, p. 22): a aglutinação de fatores em um índice sintético não discrimina um “fator-insumo”, de um “fator-processo” ou “fator-produto”.

influências de forma direta e indireta. Pode-se dizer que os índices sintéticos e os sistemas de indicadores sociais não correspondem a essas características. Nesses procedimentos, o pesquisador inclui nas análises os fatores que consideram relevantes. Como não se abre a possibilidade de que a determinação dos fatores mais influentes seja buscada nas próprias informações empíricas, não há como valorar, a partir da análise dos dados, quais as intensidades exercidas por fator sobre as condições de vida da população. Deve-se ressaltar a possibilidade de que, no caso dos índices sintéticos, sejam atribuídas ponderações distintas a cada um dos fatores como uma tentativa de reconhecer as diferentes relevâncias assumidas por elemento dentro do processo estudado. Entretanto, como já mencionado, essa não é uma solução satisfatória, pois envolve uma dose de arbitrariedade, constituindo-se ainda em alvo de controvérsias. A análise de regressão, a seu turno, pode ser considerada em conformidade com o quarto e o quinto atributos desejáveis, uma vez que é possível, a partir do tratamento dos dados, determinar as associações entre variáveis e suas intensidades (ainda que essa metodologia só valore as influências diretas que cada fator exerce sobre as condições de vida).

O método da regressão também se apresenta em conformidade com o sexto atributo: *a habilidade de se explicar uma porcentagem da variabilidade do fenômeno*, o qual costuma ser denotado, no âmbito das investigações empíricas, por meio da renda ou de outros indicadores das condições de vida. Cabe notar que os procedimentos metodológicos capazes de fornecer tal explicação são aqueles que estabelecem associações entre as medidas de pobreza e os demais indicadores socioeconômicos dos indivíduos, como é o caso desse método.³³ No entanto, não sendo esse mesmo atributo oferecido pelos índices sintéticos nem pelos sistemas de indicadores sociais, trabalhos calcados nessas duas outras metodologias não colaboram para a explicação da variabilidade do fenômeno.

A intenção de que se realizem estudos relacionais sobre o fenômeno da pobreza evoca ainda a necessidade de que uma dada metodologia *inclua conceitos complexos e abstratos de forma estável*. No caso de um indicador sintético, esse é o seu principal propósito: representar um conceito abstrato e multidimensional sob a forma de um só índice. Entretanto, muitas vezes, a elaboração de um indicador sintético não se sustenta em procedimentos que contribuam para a sua estabilidade, tais como a realização de testes empíricos que endossem as dimensões escolhidas para compô-lo. Em geral, esses índices são elaborados a partir da incorporação de dimensões que são consideradas relevantes pelo pesquisador, processo apoiado em lastros teóricos, mas que muitas vezes não conta com uma etapa de reafirmação por parte dos dados empíricos. Dessa forma, abre-se a possibilidade de que sejam construídos índices

33. Em linhas gerais, a habilidade de explicar uma proporção da variabilidade da variável-resposta advém do estabelecimento de correlações entre essa variável e as variáveis independentes. No caso da regressão múltipla, é calculado o "coeficiente de correlação múltipla" (R), que denota a predição da variável-resposta pelo conjunto de todas as covariáveis, simultaneamente. De fato, R é definido como a correlação entre o valor *observado* da variável-resposta (o dado empírico) e a melhor combinação linear de covariáveis. Isso equivale a dizer que esse coeficiente indica a correlação entre o valor *observado* da variável-resposta e o valor predito pela equação da regressão, a partir daquele conjunto de variáveis independentes. Sendo assim, ao se elevar o valor de R ao quadrado, tem-se R^2 , o coeficiente de determinação, que indica a porcentagem da variabilidade do fenômeno que é explicada por aquele conjunto de covariáveis (HOWELL, 2002, p. 546).

instáveis, já que se constituem segundo um arcabouço conceitual que pode não encontrar correspondência nas tendências delineadas pelos dados empíricos referentes a certos contextos. Por outro lado, tentativas de contornar essa fragilidade podem fazer com que o índice sintético venha a ser formado por facetas altamente correlacionadas do fenômeno – o que as torna redundantes, de modo que pouco da sua multidimensionalidade seja incorporado à análise.

Os sistemas de indicadores sociais, por sua vez, não propiciam a maneira fácil de comunicar e interpretar conceitos complexos na análise. Para representar tais conceitos abstratos e multifacetados, esses sistemas apresentam os resultados referentes aos vários aspectos envolvidos no fenômeno de forma decomposta. Assim, devido à sua própria natureza fragmentária, essa metodologia dificulta a visão de conjunto da complexidade que marca o fenômeno estudado.

Nos modelos de regressão, os conceitos complexos e abstratos costumam ser admitidos sob a forma de compósitos. A inclusão de compósitos pode trazer alguns fatores complicadores à análise. Os problemas incluem questões como a alteração das informações iniciais através do cálculo de médias, as mudanças causadas na forma de distribuição dos dados e os impactos que essas modificações provocam nas estimativas finais do modelo. Há também os problemas ligados à estabilidade do índice criado para representar um conceito complexo; nesse sentido, os comentários feitos sobre o equilíbrio dos indicadores sintéticos também se aplicam à criação de compósitos.

Por último, cabe comentar como os métodos tradicionais posicionam-se em relação à capacidade de *avaliar erros de mensuração*. Pode-se dizer que os resultados das metodologias tradicionais são permeados pelas complicações geradas por erros de mensuração: as estimativas produzidas estão sujeitas a distorções, uma vez que nenhum desses métodos oferece a possibilidade de avaliar a magnitude desses componentes aleatórios, nem de ajustar os seus cálculos em função da confiabilidade dos dados.

No contexto específico das análises de regressão, os erros de mensuração provocam um tipo de problema nas estimativas dos coeficientes. Trata-se da “atenuação dos coeficientes”. O problema caracteriza-se pelo fato de que o valor da correlação entre duas variáveis tende a diminuir, na medida em que as variáveis que lhe deram origem são menos confiáveis, ou seja, estão mais imbuídas de erro (KLINE, 1998, p. 26). A causa de tal redução no valor dos coeficientes é que, conforme já colocado, os procedimentos estatísticos incluem nos cálculos tanto a parcela “verdadeira” da informação social, como os erros de mensuração. A incorporação da parte aleatória da medida faz com que o comportamento da informação “verdadeira” apresente uma tendência menos definida do que ela teria, se fossem expurgados esses erros. Com isso, as correlações calculadas tendem a ser mais fracas e, por conseguinte, os coeficientes de regressão tendem a ser menores do que realmente são.

Assim, observa-se que as metodologias tradicionais de análise da pobreza têm capacidades variadas de apreender a complexidade do fenômeno, sendo que nenhuma delas apresenta-se plenamente satisfatória para dar conta dessa tarefa. Cada uma delas reflete, de forma parcial, o caráter multidimensional e complexo do fenômeno.

Para retratar essa discussão de forma resumida, elaborou-se o quadro comparativo apresentado na tabela 1. Resumindo os principais tópicos que foram desenvolvidos até aqui, ele aponta a forma como cada uma das metodologias ora abordadas se posiciona em relação às propriedades desejáveis em um método de análise da pobreza. A partir disso, pode-se ter uma idéia dos limites e potencialidades de que cada metodologia dispõe para refletir a construção relacional dessa questão social, na fase de análise dos dados empíricos em uma pesquisa.

TABELA 1
Comparação entre os métodos tradicionais de análise da pobreza e características desejáveis em método de análise relacional

Métodos tradicionais Atributos desejáveis	Indicadores sintéticos	Sistemas de indicadores sociais	Análise de regressão
Superar os limites da simples quantificação	Trabalha no estágio da quantificação, gerando resultados agregados	Trabalha no estágio da quantificação, gerando resultados agregados	Não quantifica; trabalha de forma sensível à distribuição dos dados
Estabelecer as relações entre os diferentes aspectos envolvidos no problema	Não estabelece	Permite a realização de poucos cruzamentos; ainda assim, não estabelece relações	Estabelece parcialmente (apenas entre as covariáveis e as variáveis dependentes)
Oferecer uma representação gráfica do fenômeno	Não oferece (não leva em consideração a estrutura do fenômeno)	Não oferece (trabalha com visões parciais do fenômeno)	Não oferece (trabalha com uma única forma estrutural de relações)
Determinar que fatores se mostram mais influentes nas condições de vida das pessoas	Não determina (inclui fatores considerados relevantes pelo pesquisador)	Não determina (abrange fatores considerados relevantes pelo pesquisador; não estabelece relações)	Determina (associações entre fatores emergem dos dados)
Valorar a intensidade de tais influências, tanto direta como indiretamente	Não valora (imposição externa de ponderações)	Não valora	Valora (intensidade de cada associação emerge dos dados)
Explicar uma proporção da variabilidade do fenômeno	Não explica	Não explica	Explica (indica a proporção explicada pelo conjunto de fatores considerados)
Incluir conceitos sociais complexos e abstratos de forma estável	Inclui (mas pode haver fragilidades, devido ao grau de arbitrariedade envolvido na construção do indicador)	Não inclui (informação fragmentada)	Pode incluir (através de compósitos, sujeitos aos mesmos problemas dos indicadores sintéticos)
Avaliar os erros de mensuração das variáveis	Não avalia (não ajusta resultados em função de uma estimativa de confiabilidade dos dados)	Não avalia (não ajusta resultados em função de uma estimativa de confiabilidade dos dados)	Não avalia (não ajusta resultados em função de uma estimativa de confiabilidade dos dados)

4 A MEE COMO MÉTODO PARA A ANÁLISE RELACIONAL DA POBREZA

Nesse ponto da discussão, há que se apresentar um método de análise de dados adequado para refletir o caráter multidimensional e complexo da pobreza. De acordo com a argumentação que vem sendo desenvolvida até aqui, pode-se dizer que uma metodologia possui o potencial de apreender a natureza relacional dessa questão social se ela dispõe das propriedades já enumeradas, tidas como desejáveis em um método de investigação empírica do fenômeno.

Sob essa ótica, a MEE coloca-se como uma metodologia propícia ao estudo relacional da pobreza, como se verá ao longo desta discussão. Essa técnica é relativamente nova, tendo surgido na década de 1960 como resultado da evolução da modelagem de multiequações, desenvolvida principalmente nos campos da psicologia, da econometria e da sociologia. Dentro do panorama evolutivo dos métodos quantitativos de pesquisa, pode-se dizer que as raízes da MEE remontam ao ano de 1918, quando o geneticista Sewall Wright apresentou um trabalho seminal sobre a técnica da análise de trajetórias, a qual só veio a ser exposta formalmente em uma série de artigos que datam do começo dos anos 1920 (DUNCAN, 1971, p. 116).

A análise de trajetórias caracteriza-se por abordar os problemas de dependência entre variáveis, típicos da regressão. Contudo, ela dá um passo adiante, pois envolve duas ou mais equações de regressão na modelagem do fenômeno estudado. Por isso, esse método delinea problemas de regressão através de um “diagrama de trajetórias”, que representa as associações entre variáveis observáveis, estimando-lhes os coeficientes (TURNER; STEVENS, 1971, p. 75). Com o passar do tempo, essas propriedades foram sendo aprofundadas e refinadas, o que veio a dar origem à MEE.

Em linhas gerais, a MEE caracteriza-se pela sua capacidade de especificar, estimar e testar relações hipotéticas entre um grupo de variáveis. Comporta também a inclusão de conceitos sociais abstratos sob a forma de construtos ou variáveis latentes (elaborados com base em variáveis observáveis que lhes servem de indicadores), que podem ser correlacionados entre si (KLINE, 1998; BENTLER; WU, 2002). Outra feição dessa metodologia é que tais modelos são lineares, uma vez que as relações entre todas as variáveis, latentes ou observáveis, podem ser representadas por equações estruturais lineares ou podem ser assim transformadas (BOLLEN, 1989, p. 3). Além disso, propicia que as correlações entre variáveis e construtos sejam traduzidas sob a forma de um diagrama, ou seja, as relações que conformam o modelo são dispostas de modo gráfico ou pictórico, o que possibilita uma representação mais clara da teoria estudada (KAPLAN, 2000, p. 4).³⁴ Kline (1998, p. 47-48) entende que essa representação pictórica assume centralidade no âmbito da MEE. O autor argumenta que, como essa técnica é usualmente aplicada à análise de fenômenos complexos e intrincados, esses diagramas mostram-se como dispositivos eficientes para ilustrar as diversas relações simultâneas estabelecidas entre as variáveis, além de sumarizarem os resultados das análises.³⁵

Da mesma maneira como procedem outras técnicas de análise quantitativa multivariada, a referida metodologia consiste em testar a plausibilidade de um modelo construído com base em uma teoria que sustenta o fenômeno estudado. Esse procedimento é usual a diversas técnicas tradicionais, tais como análise fatorial e análise de regressão.³⁶ A diferença entre esses procedimentos clássicos e a MEE é que

34. Em um diagrama de MEE, existe a convenção de se denotar os construtos ou variáveis latentes por meio de “balões”. As variáveis observáveis são indicadas graficamente por quadrados.

35. Na MEE, há duas formas equivalentes de se denotarem as relações entre as variáveis: elas podem ser representadas através do diagrama e do sistema de equações estruturais.

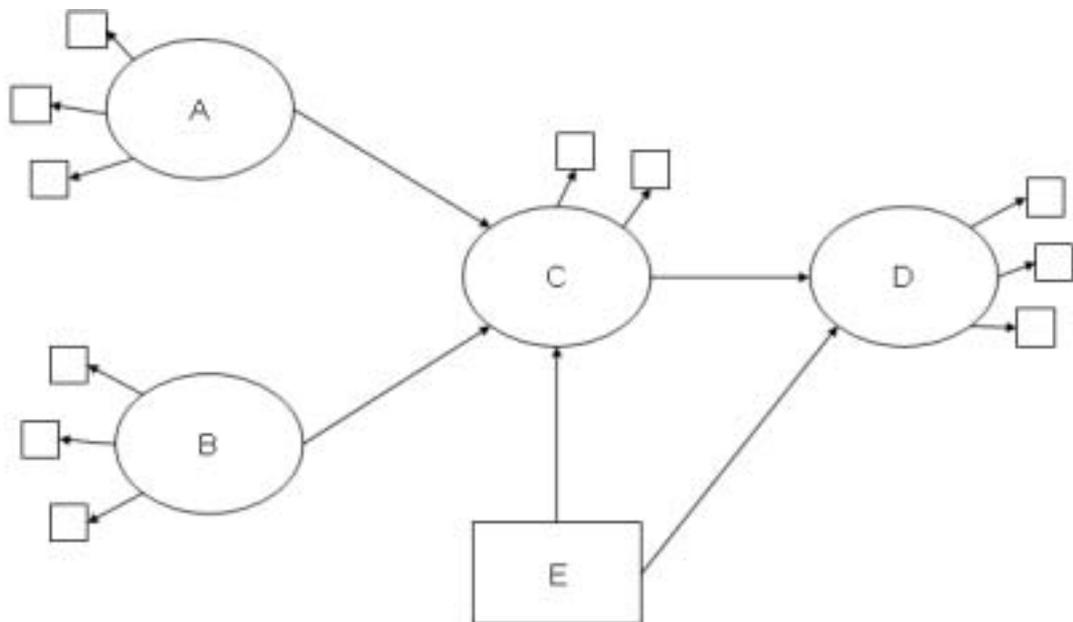
36. Muitos procedimentos estatísticos podem ser vistos como casos particulares da MEE. Dentre eles, pode-se citar a análise de regressão, a análise de trajetórias, a correlação canônica, a análise fatorial e a ANOVA (KLINE, 1998, p. 8).

aqueles se caracterizam por contemplar múltiplas variáveis independentes, mas só admitir uma única relação entre as variáveis dependentes e independentes. A MEE, por sua vez, ultrapassa essa restrição, permitindo que se apreciem diversas relações possíveis entre variáveis e construtos, sejam eles dependentes ou independentes (KLINE, 1998).

Em outras palavras, no âmbito da MEE, é possível que se analisem várias relações de dependência entre variáveis, incluindo a possibilidade de que uma variável dependente em uma equação seja independente em outra. Essa situação é ilustrada no gráfico 4, que mostra o construto C exercendo o papel de variável independente na determinação de D, ao mesmo tempo em que aparece como variável dependente em outra equação, em que é função de A, B e E.³⁷ Tendo em vista o posicionamento relativo que cada variável assume no modelo, cabe esclarecer que A, B e E são chamadas de variáveis “exógenas”, uma vez que não existem outras que exerçam efeito sobre elas; analogamente, C e D são tidas como “endógenas”, pois ambas recebem a influência de outras variáveis presentes no modelo.

A possibilidade de que as variáveis sejam dispostas de modo intrincado dentro do modelo abre espaço para que se possam distinguir e estimar os efeitos diretos, indiretos e totais que uma variável pode exercer sobre a outra (BOLLEN, 1989, p. 5; KAPLAN, 2000, p. 4). Como se pode notar no gráfico 4, a variável E exerce efeito sobre D diretamente e também indiretamente, através de C. Por isso, sua influência total sobre D é dada pela soma dessas duas parcelas.

GRÁFICO 4
Esquema ilustrativo sobre relações estabelecidas em um modelo de equações estruturais



37. Observe-se que a notação usada no referido gráfico designa que as variáveis A, B, C e D são latentes, sendo E uma observável. Deve-se esclarecer também que os quadrados pequenos representam as variáveis observáveis a partir das quais são construídas as latentes, que se caracterizam por serem não diretamente mensuráveis. As relações de dependência entre as variáveis podem ser expressas ainda pelas funções matemáticas $D = f(C, E)$ e $C = f(A, B, E)$.

A partir desse exemplo, pode-se observar que a MEE representa um avanço em relação ao método da regressão: ela expande as possibilidades de análise de uma questão social, uma vez que oferece maior flexibilidade de representação das estruturas existentes entre os fatores envolvidos em um fenômeno. Por permitir que se calculem as correlações entre os diversos fatores que participam do modelo, a MEE contorna o problema da multicolinearidade. Além disso, com a possibilidade de discernir efeitos diretos e indiretos, supera-se aquela limitação inerente à análise de regressão, referente aos problemas gerados pela inclusão de variáveis mediadoras ou intervenientes na modelagem.

Devido a essas potencialidades, a MEE permite que a articulação entre os diversos fatores associados à pobreza, tal como discutida pela teoria, seja traduzida na forma de um modelo empírico, que pode ser estatisticamente estimado e testado. Assim, essa técnica de análise multivariada é capaz de identificar quais são os fatores mais imediatamente ligados a situações de carências e privações, de oferecer uma visão da forma como eles se inter-relacionam e de calcular o peso de cada um na determinação das condições de pobreza. Nesse sentido, o método também apresenta a potencialidade de evitar que, durante a análise, as diferentes dimensões associadas ao fenômeno se confundam com os indicadores da pobreza propriamente dita, assim como discute Mestrum (2002, p. 37), em suas colocações sobre a complexidade inerente à abordagem multidimensional dessa questão social.

Dentro de uma perspectiva mais técnica, pode-se dizer que a MEE é um método de modelagem estatística mais abrangente, que pode ser visto como a combinação da análise fatorial³⁸ com a análise de regressão. As principais potencialidades oferecidas por tal método decorrem dessa combinação: a possibilidade de inclusão de conceitos abstratos na modelagem através da síntese conceitual da variável latente; e o estabelecimento de relações estruturais entre os construtos teóricos (BOLLEN, 1989, p. 4).

De fato, um dos motivos que faz a MEE ser particularmente adequada aos estudos sociais é que muitas teorias e modelos elaborados nessa área do conhecimento baseiam-se em conceitos teóricos abstratos. Como já se mencionou, esses conceitos não podem ser diretamente observados, não sendo passíveis de mensuração. Suas inserções em modelagens estatísticas se dão através dos chamados “construtos” ou “variáveis latentes”, elaborados a partir da utilização de variáveis observáveis, que funcionam como indicadores daqueles conceitos trabalhados na seara teórica (BYRNE, 1994). No âmbito da MEE, a incorporação desses conceitos se deve à

38. A análise fatorial é um tipo de procedimento estatístico aplicado quando se desejam construir “fatores” – que usualmente representam conceitos abstratos ou não diretamente mensuráveis – a partir de variáveis observáveis disponíveis em uma base de dados. Os principais propósitos para a aplicação de técnicas analíticas de fatores são: *a*) a redução do número de variáveis (isso acontece quando se dispõem de variáveis altamente correlacionadas, que indicam um mesmo conceito; para evitar tal redundância, é criado um único fator para representá-las); *b*) a detecção da estrutura das relações entre variáveis, o que equivale a realizar uma “classificação de variáveis” em grupos ou “fatores”. Nesse caso, a análise fatorial revela a existência de padrões entre as variáveis, padrões esses que conduzem o processo de agrupamento daquelas variáveis em distintos fatores; a “classificação” se dá, portanto, por meio da criação de grupos constituídos por aquelas que forem altamente correlacionadas entre si. No âmbito da MEE, quando se fala em análise fatorial, está-se referindo à sua modalidade “confirmatória”, aquela que permite testar hipóteses que tenham sido previamente elaboradas acerca da estrutura fatorial de um grupo de variáveis.

presença da análise fatorial, metodologia que estabelece quais variáveis observáveis designam, conjuntamente, uma noção abstrata. O fator criado dessa maneira constitui-se então em uma síntese conceitual, que é inserida no modelo sob a forma de uma variável latente. Assumindo tal formato, o conceito abstrato está apto a se correlacionar com os demais fatores e variáveis incluídos na análise (HOX; BECHGER, 1998, p. 373).

A possibilidade de que se estabeleçam associações entre esses construtos liga-se à presença da análise de regressão no bojo da MEE. É possível que se calculem relações não apenas entre variáveis latentes, mas também aquelas estabelecidas com as observáveis. Todas essas inter-relações são parâmetros do modelo, representados por coeficientes de regressão. Dessa forma, os já mencionados benefícios oferecidos pela análise de regressão também fazem parte dos resultados produzidos pela metodologia ora abordada. Mais especificamente, é possível que se identifiquem quais os elementos mais relevantes dentro da análise, assim como podem ser determinadas as intensidades das correlações entre fatores. Deve-se salientar ainda que, da mesma maneira como acontece com os modelos de regressão, é possível calcular as proporções das variâncias das variáveis dependentes que são explicadas pelo conjunto das variáveis independentes. Nesse ponto deve-se lembrar que a MEE permite que, em um só modelo, existam diversas variáveis dependentes – são dependentes todas aquelas que assumem o *status* de “endógenas”. Para todas as variáveis endógenas presentes em um modelo de equações estruturais, sejam elas observáveis ou latentes, são estimadas as porcentagens explicadas de suas variâncias.

Os aspectos recém-abordados – isto é, a inclusão de variáveis latentes e o estabelecimento de relações estruturais entre os fatores – são tratados por duas partes distintas ou “submodelos”, que compõem um modelo de equações estruturais completo: o de “mensuração” dos construtos e o das “relações estruturais” entre eles. O submodelo de mensuração especifica como variáveis latentes dependem das ou são indicadas pelas variáveis observáveis. Como mencionado, isso se dá por meio da realização de análise fatorial confirmatória, que determina a forma como os construtos são mensurados a partir das variáveis observadas. Esse submodelo oferece ainda uma descrição das propriedades de mensuração (validade e confiabilidade) dessas variáveis.

A parte estrutural, por seu turno, fundamenta-se no cálculo de sistemas de equações simultâneas. É ela que encerra os procedimentos de especificação e estimação das associações das variáveis latentes entre si ou com outras observáveis, descrevendo seus efeitos e intensidades.³⁹ Além disso, estão incluídas nessa parte as informações sobre a variância explicada e a não-explicada de cada termo endógeno presente no modelo (KAPLAN, 2000, p. 5).

39. A estimação dos parâmetros do modelo de equações estruturais utiliza o método da máxima verossimilhança, que se caracteriza por fazer uma estimação “simultânea”, ou seja, por calcular todos os parâmetros do modelo de uma só vez. A denominação desse método descreve o princípio estatístico que subjaz ao processo de derivação dos parâmetros: assume-se que os parâmetros do modelo são valores da população; assim, tal método estima os coeficientes que maximizam a semelhança (ou a probabilidade, “L”) de que os dados observados tenham sido tirados dessa população (KLINE, 1998, p. 125).

A proposta central da MEE é que se construam modelos que possam refletir as associações entre conceitos, tais como descritas pela teoria. Como se viu, essas associações são representadas nos modelos por uma série de relações estruturais, semelhantes às equações de regressão. Para sustentar essa construção, o procedimento metodológico consiste em elaborar e testar um modelo hipotético, que espelhe aquelas disposições teóricas. Esse teste constitui-se na aplicação de um procedimento estatístico que analisa todo o sistema de variáveis simultaneamente, com a intenção de que se determine em que medida aquelas postulações que conformam o modelo hipotético são consistentes com os dados. Se o ajuste do modelo for considerado adequado, admite-se a plausibilidade das relações supostas entre variáveis; se for inadequado, a postulação de tais relações é rejeitada (BYRNE, 1994).

Tecnicamente, a verificação da consistência entre o modelo hipotético e os dados empíricos consiste na análise da estrutura da matriz de covariância⁴⁰ das medidas. Uma vez estimados os parâmetros do modelo, sua matriz de covariância deve ser comparada à matriz de covariância do “modelo nulo” – aquele que não impõe nenhum tipo de relação entre as variáveis, ou seja, elas são consideradas livremente, de modo que a sua matriz de covariância reflete as correlações “naturais” entre os dados, que são as efetivamente encontradas naquela base empírica. Se ambas as matrizes forem consistentes entre si, tem-se que as relações postuladas no modelo que está sendo testado se aproximam do comportamento natural dos dados. Com isso, aquele modelo construído com base na teoria, que intenta traduzir o fenômeno investigado, pode ser considerado uma explicação plausível para as relações entre as medidas⁴¹ (KLINE, 1998).

O fato de que a MEE utiliza-se da estrutura de covariância das variáveis observáveis coloca-se como uma idiosincrasia dessa metodologia. A esse respeito, Bollen (1989, p. 1) nota que a maioria dos pesquisadores que aplicam procedimentos estatísticos conduzem seus raciocínios em termos da modelagem de observações individuais. Na regressão múltipla, por exemplo, os coeficientes de regressão ou as estimativas de outros parâmetros são gerados pela minimização da soma das diferenças dos quadrados das variáveis dependentes preditas e observadas, para cada caso individual. Da mesma forma, a análise dos resíduos, etapa que visa a aferir o ajuste do modelo aos dados empíricos, expõe as discrepâncias entre valores preditos e observados para cada membro da amostra.

40. A covariância é definida como uma medida absoluta do grau em que duas variáveis tendem a mudar juntas. A noção subjacente a esse conceito é similar à do coeficiente de correlação, já que ambos medem a intensidade da associação entre um par de variáveis. Uma matriz de covariância, por sua vez, é um quadro bidimensional que contém todas as variâncias e covariâncias existentes entre as variáveis de uma base de dados. No caso de um modelo que está sendo testado, sua matriz de covariância consiste no cálculo das relações entre cada uma de suas variáveis observadas e as demais. Por ser baseada em matrizes de covariância, a MEE é também conhecida como Análise de Estruturas de Covariância (HOX; BECHGER, 1998, p. 373).

41. Em seu trabalho, Bollen (1989, p. 1) desenvolve uma explicação teórica sobre a metodologia da MEE. O autor afirma que o procedimento de comparação entre a matriz de covariância do modelo testado e a do “modelo nulo” encontra justificativa na assunção de que a matriz de covariância das variáveis observadas é função de um conjunto de parâmetros. Assim, se ambas as matrizes se aproximarem, de modo que o modelo testado possa ser considerado “correto” ou bem ajustado, pode-se dizer que foram revelados os parâmetros que regem as relações entre aquelas variáveis no mundo empírico.

Para Bollen (1989, p. 1), a MEE demanda uma reorientação nesse tipo de pensamento, uma vez que tal procedimento enfatiza as covariâncias em vez dos casos.⁴² Ao se referir a procedimentos metodológicos de verificação do ajuste de um modelo de equações estruturais aos dados, o autor chama a atenção para o fato de que, no caso da MEE, os resíduos são formados pelas diferenças entre as covariâncias observadas e as previstas. Assim, em vez de se minimizarem funções de valores observados e previstos de casos individuais, devem-se minimizar as diferenças entre as covariâncias da amostra (isto é, do modelo nulo) e as covariâncias previstas pelo modelo que está sendo testado. A partir da comparação entre essas matrizes de covariância – e da noção de que é desejável que suas estruturas mostrem-se conciliadas – é que se desenvolvem os coeficientes voltados para a aferição do ajuste do modelo aos dados.

A literatura sobre MEE mostra a existência de diversos testes que avaliam o ajuste de um modelo hipotético aos dados observados.⁴³ Como mencionado, esses testes de ajuste, em geral, consistem em comparar a matriz de covariância dos dados (a observada) com a implicada pelo modelo. Em sua forma mais simples, essa noção é traduzida na realização de teste que calcula a diferença entre ambas as matrizes e analisa os resíduos existentes entre elas.

Conforme exposto, o cenário ideal na avaliação dos resíduos é que não haja diferença entre aquelas matrizes. Em sua aplicação a situações reais, essa lógica indica apenas que os resíduos devem ser o mais próximos de zero quanto possível, para que o ajuste do modelo seja considerado satisfatório. De fato, é usual que as matrizes apresentem diferenças, o que denota que algumas variâncias e covariâncias das variáveis observadas não são exatamente previstas pelo modelo (BOLLEN, 1989, p. 257).

Esse tipo de avaliação do ajuste do modelo aos dados se concretiza através de um teste de qui-quadrado.⁴⁴ Ao comparar ambas as matrizes, o qui-quadrado testa a hipótese de que elas são idênticas. Como se disse, é desejável que elas sejam próximas, similares, de modo que tal teste não aponte uma discrepância significativa entre as covariâncias observadas e as implicadas pelo modelo. Entretanto, há que se notar que esse teste mostra-se falível, uma vez que é altamente dependente do tamanho da

42. Como se disse, os cálculos realizados por programas computacionais de estimação de modelos de equações estruturais baseiam-se em matrizes de covariância. Entretanto, o pesquisador pode alimentar esse programa com uma base de dados composta por casos individuais (microdados), uma vez que o próprio *software* se encarrega de calcular a matriz de covariância correspondente àqueles dados (KLINE, 1998, p. 8).

43. A literatura sobre MEE é bastante farta, tanto com relação aos estudos que discutem conceitualmente esse método e suas propriedades, quanto àqueles que o utilizam como instrumento para a análise de questões sociais substantivas. Entre os trabalhos realizados na área, podem-se citar Bedeian, Day e Kelloway (1997); Boomsma (2000); Edwards e Bagozzi (2000); Fincham *et al.* (1997); Hoyle e Smith (1994); Klem (2000); MacCallum *et al.* (1993); Marshall e Lang (1990); McDonald e Ringo Ho (2002); Pierce, Vinokur e Buck (1998); Raykov, Tomer e Nesselroade (1991); Tomas e Oliver (1999); Vinokur, Price e Caplan (1996); Vinokur e Schul (1997, 2002). Cabe notar aqui que, apesar da existência de uma plethora de índices de ajuste dos modelos de equações estruturais, serão abordados neste trabalho apenas os mais utilizados nos estudos recentes.

44. O qui-quadrado é um teste estatístico que prima detectar a significância de relações entre variáveis categóricas. Mais especificamente, ele avalia se são significantes as diferenças entre os valores *observados* e os *esperados* de certas variáveis. No contexto da MEE, o qui-quadrado é usado para comparar a matriz de covariância observada e a implicada pelo modelo.

amostra (quando se trabalha com grandes amostras, o valor do qui-quadrado tende a ser alto). Nessas circunstâncias, ele tende a acusar a significância da diferença entre aquelas matrizes, mesmo quando esse resultado não é apropriado. Em outras palavras, uma amostra grande pode conduzir a um resultado em que a matriz residual não é zero, mas apresenta magnitudes ínfimas, de modo que a significância substantiva dessas diferenças pode ser negligenciada⁴⁵ (BOLLEN, 1989, p. 268).

Por isso, é necessário dispor de medidas mais robustas e estáveis, que não sejam tão dependentes do tamanho da amostra. Nesse sentido, foram desenvolvidos outros índices de verificação do ajuste do modelo aos dados, que são menos vulneráveis a essas influências amostrais. Tal propriedade mostra-se particularmente necessária no contexto da MEE já que, em geral, a estimação desses modelos tende a requerer o uso de grandes amostras (KLINE, 1998, p. 8).

Para se contornar o problema da influência do tamanho amostral, o ajuste de um modelo deve ser aferido relativamente a outro modelo que tenha sido construído com o mesmo conjunto de dados. Sob esse prisma, três tipos de modelos fazem-se fundamentais: *a)* o modelo hipotético, aquele que foi elaborado a partir de alguma base teórica e que está sendo testado; *b)* o modelo nulo que, como já foi dito, caracteriza-se por não impor nenhum tipo de relação entre as variáveis; e *c)* o modelo saturado ou “cheio”, que estabelece todas as relações possíveis entre as variáveis, contemplando, portanto, a estimação de todos os parâmetros (com isso, esse modelo oferece o ajuste máximo, produzindo covariâncias exatamente iguais às dos dados observados).⁴⁶

Sendo o modelo saturado aquele que apresenta o ajuste mais perfeito aos dados observados, ele é o primeiro a ser utilizado como base de comparação quando se quer verificar a adequação de um modelo hipotético. Assim, a medida de ajuste do modelo saturado é comparada à do hipotético. Como o método de estimação usado na MEE é o da máxima verossimilhança – que calcula os valores dos coeficientes que maximizam a probabilidade (L) de que os dados da amostra tenham saído da população que é regida por certos parâmetros –, o ajuste de um modelo estimado por esse método é aferido por meio do logaritmo da probabilidade L , o $\log(L)$.⁴⁷ Como se disse, o modelo saturado é aquele que, por simplesmente reproduzir os dados observados, apresenta o melhor ajuste. Assim, o valor do $\log(L_s)$ é tomado como referência de ajuste perfeito. A idéia é então checar se o valor do $\log(L_h)$ aproxima-se do de $\log(L_s)$; se eles forem próximos, tem-se que o modelo hipotético apresenta um

45. Conforme ilustra Bollen (1989, p. 268), essa situação é análoga a uma outra em que se rejeite a hipótese nula de que um coeficiente de correlação é zero entre duas variáveis, por se ter encontrado uma correlação amostral no valor de 0,03 com base em uma amostra extremamente grande. Segundo a argumentação do autor, o grande tamanho da amostra pode indicar que é improvável que a correlação populacional seja zero, mas o significado substantivo de uma correlação no valor de 0,03 é nulo, na maioria dos casos.

46. O modelo saturado é uma construção sem lastro teórico; seu objetivo é apenas ajudar o processo de aferição do ajuste de um modelo que está sendo testado. Como ele é pensado para reproduzir exatamente a distribuição dos dados observados, estimando todas as relações possíveis, não acrescenta informação alguma que possa iluminar a análise da estrutura de um fenômeno estudado.

47. Para facilitar a fluência deste texto, será utilizada a seguinte notação: $\log(L_h)$ designa o logaritmo da probabilidade L do modelo hipotético; de modo análogo o $\log(L_s)$ e o $\log(L_n)$ referem-se aos modelos saturado e nulo, respectivamente.

bom ajuste. Uma grande diferença entre esses valores indica que o modelo testado não tem uma boa adequação aos dados (XIE; POWERS, 2000, p. 26).

O procedimento ora descrito conforma um indicador usual de verificação do ajuste de um modelo, que é a estatística da razão de verossimilhança (G^2). Ela mede o quanto o modelo hipotético afasta-se do saturado. Em termos técnicos, a razão de verossimilhança é calculada como “menos duas vezes o logaritmo da razão de probabilidades do modelo hipotético com o saturado”⁴⁸ (XIE; POWERS, p. 66-67).

A estatística da razão de verossimilhança apresenta uma distribuição semelhante à do qui-quadrado, a ponto de serem consideradas como equivalentes (DOBSON, 2002, p. 76). Por isso, em vez de se trabalhar com a própria função do logaritmo da probabilidade, os cálculos são feitos com base na função do qui-quadrado, que se mostra uma alternativa mais conveniente do ponto de vista computacional (BOLLEN, 1989, p. 270). Assim, os resultados são apresentados em termos de valores de qui-quadrado, que designam a diferença de um modelo qualquer em relação ao saturado.

O ajuste do modelo estimado, entretanto, não pode se dar apenas com base na razão de verossimilhança. De modo similar ao que acontece com o qui-quadrado, o valor dessa razão é dependente do tamanho da amostra – menores valores de G^2 associam-se com grandes amostras. Assim, a verificação do ajuste deve prosseguir dentro da lógica da comparação entre diferentes modelos construídos com a mesma amostra, passando a levar em consideração o modelo nulo (XIE; POWERS, 2000, p. 26).

A inclusão do modelo nulo como base de comparação para o ajuste do modelo hipotético baseia-se na lógica da “redução proporcional no erro”. Essa lógica avalia a adequação de um modelo hipotético relativamente ao nulo que, como já se definiu, é aquele em que todas as correlações são zero. Ele é o mais simples que pode existir, já que o número de parâmetros a serem estimados é o menor possível.⁴⁹

Outra forma de compreender o modelo nulo liga-se à já mencionada assunção de que a matriz de covariância das variáveis observadas é função de um conjunto de parâmetros que regem as relações entre aquelas variáveis no mundo empírico. Sob tal prisma, um modelo é considerado adequado se ele ajudar a descobrir quais são aqueles parâmetros. O modelo nulo é então aquele que está o mais longe possível do ajuste aos dados, uma vez que ele não estima tais parâmetros, por não postular a existência de relação alguma entre as variáveis (em outras palavras, ele não oferece nenhuma predição melhor do que a média da variável dependente). As estimativas por ele fornecidas são as que contêm maior parcela de erro, uma vez que esse modelo não conta com nenhuma associação entre variável independente e dependente que possa ajudar a explicar a variabilidade dessa última. Assim, a diferença entre o $\log(Lnu)$ e o $\log(Ls)$ – ou o seu equivalente, o valor do qui-quadrado, X^2nu – assume o maior valor possível, já que essa

48. A expressão matemática para esse cálculo é $G^2 = -2\log(Lh/Ls)$, o que é equivalente a $G^2 = -2(\log Lh - \log Ls)$.

49. Cabe esclarecer que a MEE estima apenas os coeficientes de correlação entre variáveis, pois não trabalha com o parâmetro da “constante” que, usualmente, se apresenta em uma equação linear. Nesse contexto, o modelo nulo não estima parâmetro algum. Entretanto, em outras técnicas de modelagem que calculem o valor da “constante”, o modelo nulo estima apenas esse parâmetro, que indica o valor médio da variável dependente estudada.

discrepância denota a distância entre os modelos que apresentam o pior e o melhor ajuste, respectivamente.

A partir dessas considerações, a lógica da “redução proporcional do erro” consiste em comparar a medida de ajuste do modelo hipotético à do nulo. Dessa forma, mede-se quão melhor é o ajuste de um modelo testado relativamente ao modelo nulo. Essa melhoria é calculada a partir da diferença entre a medida de ajuste do modelo nulo ($X^2 nu$) – que é considerado o pior possível em termos de poder de predição da variável dependente – e a medida de ajuste de um modelo hipotético ($X^2 h$), que foi construído com base em alguma teoria. Como se trata de calcular a proporção da melhoria oferecida pelo modelo hipotético em relação ao modelo nulo, essa diferença deve ser dividida pela medida de ajuste deste último.

Em termos formais, essa operação pode ser expressa de acordo com a seguinte fórmula:

$$\Delta = (X^2 nu - X^2 h) / X^2 nu$$

onde:

$X^2 nu$ = medida de ajuste do modelo nulo; e

$X^2 h$ = medida de ajuste do modelo hipotético.

Assim, pode-se calcular a porcentagem de melhoria no ajuste oferecida pelo modelo hipotético em relação ao nulo – ou, dito de outra forma, a proporção da diminuição do erro na predição, fornecida pelo modelo que está sendo testado.

Para um índice dessa natureza, o melhor cenário é indicado quando $X^2 h$ é zero, o que leva Δ ao valor máximo de 1. Por outro lado, um ajuste ruim será designado por um Δ mais próximo de zero, já que o $X^2 h$ vai assumir um valor mais alto e, portanto, mais próximo do valor de $X^2 nu$. Neste caso, o modelo testado não estará acrescentando melhorias no ajuste ou na predição, relativamente ao modelo nulo (BOLLEN, 1989, p. 270).

A partir dessa lógica de comparação entre os ajustes dos modelos nulo e hipotético, são criados coeficientes que verificam o ajuste global de um modelo de equações estruturais. Esses coeficientes são também chamados de “índices de ajuste incremental”, por indicarem qual é a melhora incremental do modelo hipotético em relação ao nulo, em termos do seu poder de predição e de ajuste. Vale frisar que, como se disse, a verificação do ajuste de um modelo com base na comparação com outros que tenham sido formulados mediante a mesma amostra apresenta a vantagem de contornar o problema da influência do tamanho amostral, complicação à qual está sujeito o teste do qui-quadrado inicialmente comentado, que coteja diretamente as matrizes de covariância observada e predita.

Após esta discussão, vale a pena mencionar exemplos de alguns índices de ajuste incremental que são encontrados na literatura sobre MEE.⁵⁰ Um dos mais utilizados e

50. Para uma discussão mais aprofundada sobre os índices de ajuste mencionados a seguir, ver Bollen (1989) e Kaplan (2000).

que mais se aproxima da lógica recém-exposta é o *Bentler-Bonnet Normed Fit Index*⁵¹ (NFI). A sua estrutura corresponde exatamente ao que foi colocado, uma vez que ele é calculado pela mesma fórmula já explicada:

$$NFI = (X^2_{nu} - X^2_h) / X^2_{nu}$$

O valor do *NFI* varia de 0 a 1, indicando a proporção na melhoria do ajuste geral do modelo proposto pelo pesquisador em relação ao modelo nulo. Se o *NFI* assumir o valor de 0,80, por exemplo, então o ajuste geral do modelo hipotético é 80% melhor do que o do modelo nulo, que foi estimado com os mesmos dados amostrais.

Vale notar que esse índice de ajuste incremental não controla para o número de graus de liberdade⁵² do modelo avaliado. O número de graus de liberdade de um modelo é um indicador de sua complexidade: quanto mais complexo for um modelo (isto é, quanto mais relações forem estabelecidas entre as variáveis), menor será o seu número de graus de liberdade. Assim, a inclusão dos graus de liberdade na fórmula de cálculo de um índice dessa natureza fornece um resultado ajustado para a complexidade do modelo.

Tendo em vista essa noção, foi criada uma versão modificada do *NFI*, o *Comparative Fit Index* (CFI),⁵³ que é calculado de acordo com a seguinte fórmula:

$$CFI = (GL_{nu} - GL_h) / GL_{nu}$$

onde:

GL_{nu} = graus de liberdade do modelo nulo; e

GL_h = graus de liberdade do modelo hipotético.

Como se vê, o *CFI* apresenta uma estrutura equivalente à do *NFI*. Dessa forma, ambos os índices podem ser interpretados da mesma maneira, ou seja de acordo com a perspectiva de ajuste incremental do modelo testado em relação ao nulo. Por não se basear no valor do qui-quadrado, mas no número de graus de liberdade de cada modelo, o *CFI* mostra-se menos afetado pelo tamanho da amostra.

Outro índice que se apresenta como um refinamento do *NFI* é o *Bentler-Bonnet Non-Normed Fit Index* (NNFI).⁵⁴ Ao embutir o número de graus de liberdade do modelo em seus cálculos, o *NNFI* fornece um valor de coeficiente corrigido para a

51. Índice de Ajuste Normalizado de Bentler-Bonnet.

52. Os graus de liberdade podem ser definidos como o número de “valores” ou “pedaços de informação” que estão livres para variar, após o processo de estimação de uma estatística. De modo mais amplo, essa noção pode ser também aplicada a uma técnica de modelagem que estime um ou mais parâmetros. No caso da MEE, o número de graus de liberdade é calculado de acordo com a seguinte fórmula: $n(n + 1)/2 - k$, onde n é o número de variáveis observáveis e k é o número de parâmetros a serem estimados pelo modelo. Como os índices de ajuste incremental baseiam-se no qui-quadrado – que varia em função do nível de confiança e do número de graus de liberdade –, eles podem ser refinados se seus cálculos inserirem o número de graus de liberdade do modelo sob análise.

53. Índice de Ajuste Comparativo.

54. Índice de Ajuste Não Normalizado de Bentler-Bonnet.

complexidade da modelagem que se quer avaliar. O cálculo de tal índice é expresso da seguinte maneira:

$$NNFI = (X^2 nu / GLnu - X^2 h / GLh) / (X^2 nu / GLnu - 1)$$

Como observado, ao dividir o valor do qui-quadrado pelo número de graus de liberdade, o NNFI realiza a comparação do *ajuste por grau de liberdade* entre ambos os modelos. Com isso, obtém-se uma diminuição da sensibilidade daquela estatística ao tamanho da amostra. Note-se também que, por ser oriundo da fórmula acima, os valores do NNFI podem cair fora da faixa de 0 a 1.

Há outra forma bastante utilizada de se aferir a adequação de um modelo aos dados, que não segue a lógica do ajuste incremental. Trata-se do *Root Mean Square Error of Approximation* (RMSEA),⁵⁵ que se fundamenta no “parâmetro da não-centralidade”.⁵⁶ Sem se ater ao arsenal estatístico que subjaz à formulação desse índice, é suficiente dizer que o RMSEA pode ser entendido como uma medida do “desajuste” do modelo, que focaliza o quão distante ele se situa em relação aos dados. Além disso, o RMSEA é um índice que leva em conta a complexidade de um modelo, pois fornece uma avaliação do *desajuste por grau de liberdade*, o que pode ser notado na sua expressão matemática:

$$RMSEA = [(X^2 h / GLh - 1) / (N - 1)]^{1/2}$$

onde N = número de variáveis observáveis no modelo.

Os exemplos de índices recém-expostos são as medidas de ajuste mais usualmente reportadas em estudos baseados na MEE. Vale notar que essas são medidas de ajuste global, que se referem à adequação do modelo como um todo. Caso uma parte da modelagem mostre uma melhor adequação aos dados do que outra, esses diferenciais de ajuste não são captados nem reportados pelos índices ora mencionados.

Há que se notar também que os referidos coeficientes refletem facetas diferentes do ajuste de um modelo. Por isso, é recomendável que a adequação de um modelo aos dados seja aferida por esse conjunto de índices. Em decorrência de tal recomendação, é de praxe que os pesquisadores reportem os valores de todos eles, de modo que o leitor tenha uma noção mais completa do ajustamento daquilo que lhe é apresentado (KLINE, 1998, p. 129-130).

Outro ponto a ser comentado diz respeito à interpretação da magnitude desses coeficientes. Nesse sentido, há uma convergência na literatura da área, que considera como adequados os modelos que obtiverem índices de ajuste incremental (NFI, CFI, NNFI) maiores que 0,90. Com relação ao RMSEA, são tidos como aceitáveis aqueles modelos que apresentem valores inferiores a 0,10; valores superiores a tal patamar são

55. Índice Raiz Quadrada do Quadrado Médio do Erro de Aproximação.

56. São chamadas de “distribuições não-centrais” aquelas que são derivadas da distribuição normal, cujas médias são diferentes de zero, como é o caso do qui-quadrado. O “parâmetro da não-centralidade” dessa distribuição coincide com a expressão apresentada no numerador da fórmula do RMSEA. Uma explicação aprofundada desse assunto pode ser encontrada em Huber (1977).

considerados insatisfatórios, sugerindo que o modelo não explica bem as correlações observadas. Lembre-se que esse é um coeficiente de “desajuste”, de modo que uma adequação aceitável aos dados é indicada através de baixos valores (KLINE, 1998, p. 131; MCDONALD; RINGO HO, 2002, p. 72).

A questão do ajuste do modelo aos dados remete a uma outra discussão, que se relaciona à existência de erros de mensuração que permeiam a modelagem e sua análise. Sobre esse assunto, deve-se ter em mente que uma das principais características dos modelos de equações estruturais é que eles incluem o estudo de variáveis não-observáveis ou não-mensuráveis. Essas variáveis não-mensuráveis referem-se não apenas às variáveis latentes ou construtos – que, conforme já discutido, designam conceitos sociais abstratos a serem examinados –, mas podem também representar os resíduos que são estimados por essa modelagem. Como se sabe, o resíduo é um termo que “perturba” a variável, seja ela observável ou latente; ele denota o fato de que existem aspectos desconhecidos ou não levados em consideração pelo modelo, que exercem influências sobre as variáveis estudadas.

Dentro de um modelo de equações estruturais, os resíduos podem se apresentar sob duas formas diferentes. Primeiro, há os resíduos associados à mensuração de cada variável observável, que são indicativos de que aquelas medidas estão imbuídas de erros.⁵⁷ Há também os resíduos ligados às previsões dos fatores endógenos do modelo, designando o fato de que as previsões de tais termos não-observáveis são “menos que perfeitas”. Apesar de ambos representarem a existência de erros, convencionou-se tratar o primeiro tipo de resíduo sob a denominação de “erro” e o segundo tipo sob a de *disturbance*⁵⁸ (BYRNE, 1994, p. 8).

Ainda com relação aos tipos de resíduos que são levados em conta pela MEE, deve-se comentar que os erros de mensuração das variáveis observáveis são particularmente relevantes para essa metodologia. Isso se deve ao fato de que eles estão amalgamados com as variáveis observáveis, encontrando-se, portanto, na base de todos os cálculos que serão desenvolvidos no modelo. Assim, a consideração desses elementos é fundamental para que possam ser estimados e ajustados os demais parâmetros, inclusive o outro tipo de termo residual, as *disturbances*.

Os termos residuais são incluídos na MEE na condição de variáveis não-observáveis, de modo que suas magnitudes sejam estimadas e avaliadas. Isso denota que tal metodologia apresenta o diferencial de admitir em seus procedimentos a noção de que as medidas das variáveis são “menos que perfeitas”, ou seja, estão imbuídas do elemento aleatório. É a partir da incorporação dessa noção que se torna possível estimar as relações “verdadeiras” – que são aquelas consideradas como “puras” ou “descontaminadas dos erros” – entre as variáveis dependentes e independentes. Por isso, a MEE é tida como um método que permite o estudo da influência que um construto “livre de erros” exerce sobre outro, eliminando-se, dessa forma, potenciais vieses devidos à atenuação. Dito de outra forma, a consideração dos erros no processo de cálculo dos parâmetros do modelo

57. A natureza desses erros de mensuração e seus impactos na estimação de coeficientes de correlações entre variáveis (particularmente o problema da “atenuação dos coeficientes”) já foram abordados na discussão relativa à análise de regressão.

58. “Perturbação”.

faz com que os resultados produzidos por essa metodologia sejam mais confiáveis (BEDEIAN; DAY; KELLOWAY, 1997, p. 787).

O problema da atenuação dos coeficientes já foi mencionado na discussão sobre análise de regressão. A esse respeito, os ajustes dos parâmetros em função dos erros de mensuração, que são realizados no âmbito da MEE, acarretam uma “correção para a atenuação”. Esse processo de correção baseia-se na assunção de que todos os erros aleatórios nos indicadores observáveis de uma variável latente são essencialmente variâncias residuais; assim sendo, a correlação entre dois construtos é estimada sem essas variâncias (BEDEIAN; DAY; KELLOWAY, 1997, p. 789).

Em termos técnicos, a “correção para a atenuação” consiste em se estimar a correlação “teórica” entre duas variáveis, x e y . Tal correlação “teórica” é entendida como aquela que seria encontrada se ambas fossem perfeitamente confiáveis, ou seja, se fossem medidas sem erro. O cálculo para isso consiste na divisão da correlação observada entre as variáveis pela raiz quadrada do produto de suas confiabilidades.⁵⁹ Dessa maneira, obtém-se a correlação “desatenuada”⁶⁰ ou “verdadeira”.

Como o próprio nome indica, a magnitude do coeficiente de correlação “desatenuado” é maior que a do observado. Sendo o valor desse coeficiente inversamente proporcional à confiabilidade da mensuração das variáveis, tem-se que quanto menos confiáveis forem as medidas de x ou y , maior será a diferença entre suas correlações observadas e “desatenuadas” (KLINE, 1998, p. 197). Note-se que é através do processo de “desatenuação” das correlações observadas que as confiabilidades das variáveis são levadas em conta pela MEE.⁶¹ É por meio desse procedimento que se expurga dos cálculos de correlação a parcela de erro de mensuração que se encontra misturada com a informação social “verdadeira”. Dessa forma, é possível trabalhar com a parcela “líquida” da informação social, o que faz com que os coeficientes calculados sejam mais precisos.

Como indicam as características da MEE que foram discutidas até aqui, essa metodologia apresenta alguns traços distintivos em relação aos procedimentos tradicionais de análise multivariada. Em princípio, note-se que esse tipo de

59. Quando se trata de estimar a confiabilidade de um fator, que é composto por um grupo de variáveis observáveis, o cálculo é feito com base nas correlações entre elas. A medida mais comum de confiabilidade é o Coeficiente Alfa de Cronbach, (α) que afere a consistência interna entre aqueles itens que compõem o fator. De modo simplificado, tal coeficiente fornece a proporção da variação da variável latente que é comum entre os itens que a compõem. Em outras palavras, ele indica a proporção da informação social “verdadeira” ou “confiável” existente naquele fator. Em geral, variáveis latentes que tenham valores de α iguais ou superiores a 0,70 são consideradas satisfatórias. A partir desse patamar, as variáveis latentes são vistas como suficientemente confiáveis para participarem das modelagens.

60. O uso desse neologismo justifica-se pelo fato de que se quer passar a idéia de que o problema da atenuação dos coeficientes, causado pelos erros de mensuração, foi revertido. De maneira análoga, essa questão é designada na literatura de língua inglesa pela expressão *disattenuated correlation coefficient*.

61. Variáveis com distintos níveis de confiabilidade podem participar de um mesmo modelo. Para que as diferenças entre os níveis de confiabilidade das variáveis possam ser percebidas de modo intuitivo, basta pensar que a variável “idade” está menos sujeita a erros de mensuração do que a variável “grau de inserção no mercado de trabalho”. Entretanto, frente às diferenças de suas precisões, não se pode confiar nelas igualmente, quando se deseja calcular suas influências sobre as condições de pobreza dos indivíduos. Nessas circunstâncias, reforça-se a necessidade de que as estimativas geradas por um tal modelo sejam ajustadas para essas diferenças de confiabilidade das variáveis; caso contrário, a utilização de métodos que desconsiderem os diferentes níveis de mensuração e operem como se as medidas fossem igualmente (e perfeitamente) confiáveis pode gerar resultados enganosos.

modelagem desenvolve uma análise confirmatória em vez de exploratória – apesar de que, em algumas etapas do processo de desenvolvimento de um modelo de equações estruturais, existe a possibilidade de se realizarem procedimentos de exploração dos dados (KLINE, 1998, p. 8). Em outras palavras, pesquisadores tendem a utilizar tal metodologia para determinar se um certo modelo é válido, em vez de usar essa técnica para “achar” um modelo que combina com os dados. Nesse sentido, Bollen (1989, p. 4) ressalta que o conhecimento substantivo sobre o objeto estudado desempenha um papel fundamental como guia para a realização da maioria dos estágios do processo de modelagem.

Outro ponto a ser notado é que, por requerer que as relações entre variáveis sejam formuladas *a priori* (apesar de serem passíveis de modificações posteriores), essa metodologia presta-se bem à análise de dados para propósitos inferenciais. Por contraste, a maioria dos outros procedimentos multivariados possui natureza essencialmente descritiva (a exemplo da análise fatorial exploratória e da criação de indicadores sociais), dificultando ou mesmo impossibilitando o teste de hipóteses.

Note-se também que os procedimentos tradicionais multivariados são incapazes de avaliar ou corrigir erros de mensuração. A MEE, por seu turno, é capaz de prover estimativas explícitas desses parâmetros. Conforme discutido, é justamente por incorporar e modelar os erros de mensuração que essa metodologia gera estimativas não-viesadas para as relações entre os construtos latentes. Ligada a esse aspecto está ainda outra diferença que deve ser frisada: enquanto os procedimentos clássicos de análise de dados modelam apenas as mensurações observáveis, análises feitas com base na MEE incorporam tanto variáveis não-observáveis (latentes) quanto observáveis. Isso faz com que o uso de tal metodologia seja recomendado em estudos que trabalham com conceitos sociais abstratos.

Para que um modelo de equações estruturais seja construído e testado, Kline (1998, p. 48) descreve, de modo sumarizado, todas as etapas que devem ser percorridas pelo pesquisador. Em princípio, é necessário especificar as relações entre variáveis que compõem o modelo. Nesse estágio, as hipóteses de pesquisa são expressas sob a forma de um sistema de equações. Contudo, é possível que o processo de especificação seja iniciado com o desenho do diagrama, que deve ser posteriormente traduzido sob a forma da série de equações. Em tais equações, são definidos os parâmetros a serem estimados pelo modelo; como já exposto, eles correspondem às correlações entre variáveis latentes e observáveis, a serem testadas e calculadas por programas de computador, com base nos dados da amostra. A esse respeito, cabe mencionar a existência de diversos *softwares* disponíveis para a estimação de modelos de equações estruturais. Entre os mais populares, podem-se citar: AMOS, EQS, SAS e LISREL.

Em seguida, vem uma etapa de natureza “matemática”: deve-se determinar se o modelo está identificado. Um modelo está identificado se for teoricamente possível para o computador derivar uma estimativa única para cada um dos parâmetros que devem ser calculados. Como acontece com a resolução de qualquer sistema de equações lineares, é necessário que haja uma compatibilidade entre o número de parâmetros desconhecidos (aqueles que devem ser estimados pelo modelo) e o número de parâmetros conhecidos (que se relaciona à quantidade de variáveis

observáveis que o compõem).⁶² Entretanto, diversas complicações se interpõem ao longo desse caminho: diferentes tipos de modelos de equações estruturais devem atender a requerimentos específicos para serem identificados. Se um modelo falhar em atendê-los, ele não estará identificado e as tentativas de estimá-lo podem não ser bem-sucedidas⁶³ (KLINE, 1998, p. 49).

Após a etapa da identificação, deve-se partir para a coleta e preparação das variáveis que fazem parte do modelo.⁶⁴ Uma vez que se esteja de posse dessas variáveis, pode-se chegar ao procedimento de análise dos dados. Esse passo envolve o uso de programas computacionais capazes de calcular os parâmetros desejados, com base nos dados empíricos. Como já se disse, tal estimação se dá, em geral, por meio de métodos sofisticados, como o da máxima verossimilhança.

Após uma primeira estimação do modelo hipotético, é chegada a hora de avaliar o seu ajuste, de acordo com os coeficientes apropriados para esse fim. Conforme já discutido, isso significa determinar quão adequadamente o modelo explica os dados. Segundo Kline (1998, p. 50), é freqüente – e até esperado – que os modelos iniciais dos pesquisadores não se ajustem bem aos dados. Quando isso acontecer, deve-se então partir para a reespecificação do modelo que, da mesma forma como ocorreu com sua especificação inicial, deve incluir modificações que sejam guiadas pelas hipóteses do pesquisador. Tal reespecificação deve reiniciar toda a trajetória ora descrita, de modo que se chegue novamente à etapa de avaliação do ajuste do modelo que foi revisado e construído com base nos mesmos dados, e assim sucessivamente.

Uma vez que se chegue a um resultado estável e bem ajustado, é hora de partir para a discussão substantiva dos achados. Além das correspondências encontradas entre teoria e prática, é possível que as proposições teóricas iniciais não encontrem ressonância perfeita no modelo empírico. Isso pode abrir espaço para a realização de diversos tipos de análises, havendo inclusive a possibilidade de que a teoria venha a ser atualizada ou reformulada com base nos achados que não tenham sido por ela antecipados.

62. Para que um modelo esteja identificado, o número de parâmetros conhecidos deve ser maior do que os incógnitos. No âmbito da MEE, a quantidade de parâmetros conhecidos em um modelo é calculada pela fórmula $k = N(N + 1)/2$; onde N é o número de variáveis observáveis. O número de desconhecidos é obtido pela contagem da quantidade de relações a serem estimadas pelo modelo.

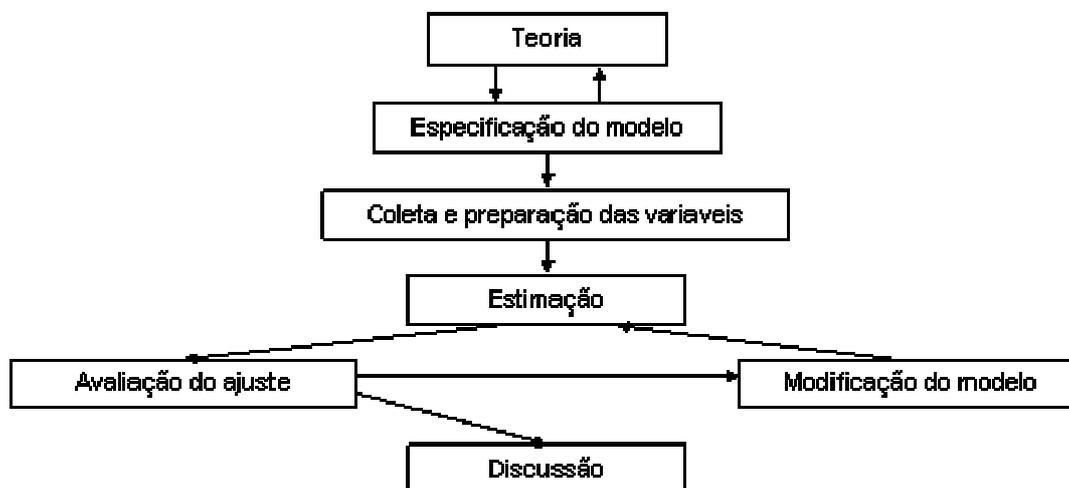
63. A etapa de identificação de um modelo a ser testado é considerada como uma das mais difíceis em todo o processo de modelagem. Diversos detalhes e exigências se impõem ao pesquisador. Para uma abordagem mais aprofundada desse tópico, ver Kenny, Kashy e Bolger (1998).

64. A MEE baseia-se nas suposições de que suas variáveis devem ser contínuas e normalmente distribuídas. Para que sigam a distribuição normal, pode ser necessário que se realizem algumas transformações nas variáveis antes que elas sejam introduzidas na modelagem. Com relação à suposição sobre a natureza contínua das variáveis, vale esclarecer que é possível incluir variáveis categóricas na modelagem. Tal inclusão respalda-se na abordagem econométrica sobre a natureza da variável categórica, que assume a existência de uma variável contínua não-observável ou "latente", subjacente a toda categórica observável. É essa variável latente que carrega a verdadeira informação social que se quer aferir. No entanto, sua natureza não permite a aferição daqueles valores subjacentes de forma precisa: esse tipo de variável só é "parcialmente observável", deixando-se apreender apenas sob o formato discreto. Nessas condições, quando o valor da latente cruza certo patamar, é possível observar uma mudança no valor na variável categórica, que é a parte que se manifesta no mundo empírico. Isso significa que os valores observados das categóricas referem-se apenas aos intervalos dentro dos quais as latentes se inserem (XIE; POWERS, 2000, p. 10). É com base nessa concepção que as variáveis categóricas são inseridas nos modelos de equações estruturais. Um maior aprofundamento sobre esse tema é encontrado em Bollen (1989).

Esse processo de construção e testagem de um modelo de equações estruturais constitui-se em uma abordagem convencional de proceder à modelagem. Essa abordagem é também descrita por Kaplan (2000, p. 8), que sumariza e dispõe as etapas recém-mencionadas (ver gráfico 5).

GRÁFICO 5

Etapas a serem percorridas para a construção de um modelo de equações estruturais



Os traços gerais da MEE que foram delineados até aqui fornecem subsídios suficientes para que se possa analisar sua capacidade de apreender relacionamente a pobreza. Dentro da argumentação que vem sendo desenvolvida neste trabalho, a habilidade de um método de pesquisa refletir a complexidade daquele fenômeno pode ser indicada por meio da presença de algumas características vistas como desejáveis, uma vez que propiciam a realização de tal abordagem relacional. Desse modo, cabe assinalar em que medida a MEE apresenta afinidade com tais características.⁶⁵

Por não trabalhar com a agregação de dados, mas conferir-lhes um tratamento sensível a suas distribuições, a MEE corresponde à primeira característica, que é *superar os limites da simples quantificação dos aspectos relativos à pobreza*. Uma vez que não busca condensar as informações em uma só medida, esse método propicia que se examine a heterogeneidade das condições de vida existente entre aqueles que vivenciam a pobreza. De fato, pode-se notar que a MEE vai além dos propósitos da quantificação, uma vez que desenvolve uma análise que busca capturar a complexidade do fenômeno, analisando a articulação das múltiplas dimensões que compõem essa questão social. Nesse sentido, essa técnica atende também ao segundo atributo desejável, um método de estudo da pobreza, que é *estabelecer relações entre os diferentes aspectos envolvidos no problema*, o que se desdobra na possibilidade de que se possam analisar os efeitos diretos, indiretos e totais que cada fator exerce sobre as condições de vida dos indivíduos.

65. A essa altura da discussão, os referidos atributos já foram suficientemente discutidos, assim como já foram analisados os motivos que fazem com que um método de pesquisa corresponda àquelas características ou não. Por isso, o exame da aproximação entre a MEE e aqueles atributos será feito de maneira sucinta, para que se evitem repetições de argumentos que já foram expostos no presente trabalho.

É também evidente a aproximação da MEE à terceira propriedade desejável em um método de análise daquele fenômeno complexo, que é *ter uma representação gráfica do fenômeno*. Conforme explicitado, está no cerne dessa metodologia o procedimento de construção de um diagrama que exponha tanto as relações existentes entre fatores socioeconômicos e medidas de pobreza, como as associações existentes entre os próprios fatores. Esse atributo faz com que a MEE apresente de maneira transparente a forma como se estruturam as relações analisadas. Por oferecer essa explicitação, o método mostra-se adequado à investigação de fenômenos complexos e intrincados, já que comunica a configuração das diversas relações entre as variáveis – ou seja, o caráter relacional do objeto social estudado – de maneira inteligível.

No que diz respeito à forma como se estruturam as relações, há que se comentar que a MEE tem a potencialidade de refletir a configuração dos fatores associados ao fenômeno de maneira flexível. Isso significa que esse método não impõe a aplicação de uma estrutura única, inerente ao método, aos diversos tipos de fenômenos que forem estudados, tal como acontece com a análise de regressão. De modo contrário, há aqui a possibilidade de que sejam testados modelos com configurações diversas, delineadas a partir de teorias específicas que iluminem cada estudo.

Essa flexibilidade acompanha todo o processo de construção e testagem de um modelo de equações estruturais. Observe-se, por exemplo, a situação em que a configuração postulada pela hipótese inicial não corresponda aos achados empíricos. Nessas circunstâncias, essa configuração deve sofrer sucessivas modificações, até que venha a emergir um modelo que se ajuste bem aos dados.⁶⁶

A estimação das relações entre fatores consiste na detecção de quais são as associações significativas e na quantificação de suas magnitudes, através do cálculo de coeficientes de correlação. Assim, a MEE é capaz de *determinar que fatores se mostram mais influentes nas condições de vida das pessoas*, mostrando-se também apta a *valorar a intensidade de tais influências, tanto direta como indiretamente*. Por meio desses atributos, torna-se possível analisar as circunstâncias que afetam a probabilidade de as pessoas vivenciarem situações de carências e privações, além de se abrir espaço para a compreensão da maneira como a pobreza se distribui entre os desfavorecidos.

Ligado a essas potencialidades encontra-se ainda o fato de que, com a utilização de tal técnica, é possível *explicar a variabilidade do fenômeno*. Conforme já explicado, a capacidade de se explicar uma proporção da variabilidade da variável-resposta de um estudo deriva do estabelecimento de correlações entre essa variável e as independentes. Por dispor dessa prerrogativa, a MEE ajuda a determinar qual é a porcentagem da variação das condições de vida das pessoas pobres que se encontra associada ao conjunto de fatores que está sendo analisado.⁶⁷

66. Deve-se chamar a atenção para o fato de que é nesse processo de busca do ajuste satisfatório, em que se incluem as sucessivas reespecificações e testagens empíricas do modelo, que a análise confirmatória abre espaço para a realização do procedimento de natureza exploratória.

67. Como se pode observar, por ter a análise de regressão como uma de suas partes constitutivas, a MEE dispõe de todas as vantagens oferecidas por essa técnica. Os pontos fracos da análise de regressão (no que tange à abordagem relacional de um objeto social complexo) são contornados por outros atributos presentes na MEE.

Com relação à capacidade de *incluir conceitos sociais complexos e abstratos de forma estável*, pode-se dizer que essa é uma das principais propriedades da MEE. Essa inclusão se dá através da síntese conceitual proporcionada pela presença de variáveis latentes na modelagem, que podem ser correlacionadas entre si e com outras observáveis que venham a participar do modelo. Elas são construídas a partir da utilização de variáveis observáveis, segundo um procedimento que se baseia não apenas em indicações teóricas, mas que encontra respaldo também na fase de análise dos dados empíricos, por meio da técnica de análise fatorial. A vantagem desse procedimento está no fato de que essa técnica confere certa estabilidade à variável latente, pois agrupa itens cujas variações tendem a confluir em uma mesma direção; isso indica que os itens que compõem um dado construto, de fato, refletem um mesmo conceito subjacente. Outro benefício advindo da análise fatorial é que ela possibilita a estimação do coeficiente de confiabilidade de um construto, coeficiente esse que, posteriormente, será utilizado para refinar os resultados fornecidos pelo modelo. Por abranger esses aspectos, a MEE propicia que conceitos complexos e abstratos sejam inseridos na análise de forma confiável e estável.

Há ainda que se comentar sobre a potencialidade de se *avaliar os erros de mensuração das variáveis*. De acordo com a discussão que foi desenvolvida, pode-se dizer que esse é um dos maiores diferenciais oferecidos pela MEE. Essa técnica reconhece o fato de que medidas de variáveis sociais estão imbuídas de erros e assimila essa noção em seus procedimentos. Nessas circunstâncias, a MEE não apenas fornece estimativas explícitas desses erros, mas ajusta os cálculos dos demais coeficientes em função da presença desses elementos aleatórios na análise. Isso se baseia em procedimentos que levam em consideração os diferentes níveis de confiabilidade das variáveis, de modo que as estimativas finais das análises sejam mais precisas, conforme já discutido.

Como se vê, a MEE oferece todos os atributos que são apontados na literatura como desejáveis em um método de pesquisa sobre a pobreza. De acordo com a argumentação que vem sendo desenvolvida neste trabalho, a presença de tais características em uma metodologia confere-lhe a capacidade de apreender a articulação dos fatores associados ao fenômeno. Dessa forma, a MEE coloca-se como uma técnica adequada para refletir a complexidade e a multidimensionalidade que marcam as discussões teóricas sobre essas situações sociais.

5 CONSIDERAÇÕES FINAIS

A intenção central deste trabalho foi discutir o potencial de abordagem relacional dos métodos de pesquisa tradicionais aplicados aos estudos sobre pobreza e MEE. Tal discussão foi estimulada pela necessidade de convergência entre teoria e método em uma pesquisa social. No caso do tema da pobreza, é importante que os métodos traduzam, na etapa de análise dos dados empíricos, a crescente complexidade que as discussões teóricas têm ensejado.

Para guiar a análise, realizou-se uma revisão da literatura sobre métodos de abordagem da pobreza. Foram então discernidas algumas características tidas como desejáveis em uma técnica, para que ela permita investigar a forma como se articulam os fatores associados às situações de carências e privações, assim como aqueles que

exercem as maiores influências sobre as condições de vida dos indivíduos. Essa revisão da literatura indicou que a construção empírica desse objeto social deve ser capaz de *superar os limites da simples quantificação dos aspectos relativos à pobreza; estabelecer as relações entre os diferentes aspectos envolvidos no problema; ter uma representação gráfica do fenômeno; determinar que fatores se mostram mais influentes nas condições de vida dessas pessoas; valorar a intensidade de tais influências, tanto direta como indiretamente; explicar a variabilidade do fenômeno; incluir conceitos sociais complexos e abstratos de forma estável; e avaliar os erros de mensuração das variáveis.*

Os atributos desejáveis passaram então a ser utilizados como parâmetros de avaliação de uma metodologia, em termos de sua capacidade de apreender relacionalmente a pobreza. Sob essa ótica, procedeu-se à análise das técnicas mais usualmente aplicadas aos estudos sobre o tema: os índices sintéticos, os sistemas de indicadores sociais e as análises de regressão. Essa discussão foi motivada pela observação de que, até os dias de hoje, as pesquisas que buscam captar a multidimensionalidade do fenômeno freqüentemente se baseiam em tais procedimentos metodológicos; dessa forma, vem se delineando um descompasso entre a evolução teórico-conceitual – que aponta para a necessidade de se pensar esse objeto social de forma articulada – e as metodologias tradicionais. Assim, partiu-se para a discussão de algumas limitações desses métodos enquanto instrumentos relacionais de análise.

Ao se confrontar cada uma dessas técnicas com os atributos supracitados, notou-se que nenhuma delas dispõe de todas aquelas propriedades tidas como desejáveis para a análise relacional da pobreza. Com isso, tem-se um indicativo de que essas formas metodológicas de abordar o tema são capazes de apreender o caráter multidimensional e complexo do fenômeno apenas de forma parcial.

Este trabalho veio então avaliar o potencial da MEE enquanto instrumento relacional para a construção empírica de tal objeto social. A descrição analítica da MEE, por sua vez, mostrou que esse método possui todos aqueles atributos. Ele apresenta-se como um instrumento apto a refletir, nas análises empíricas, o caráter relacional do fenômeno da pobreza.

Frente a essas reflexões conceituais acerca dos métodos de pesquisa sobre a pobreza, evidenciou-se que a MEE coloca-se como um procedimento mais adequado à abordagem relacional desse objeto social do que os métodos usualmente empregados. Ela enriquece e refina as possibilidades de compreensão do fenômeno, apresentando-se, assim, como uma técnica promissora para o desenvolvimento de análises relacionais no campo de estudos sobre a pobreza. Resta, portanto, construir modelos de equações estruturais empíricos sobre os fatores articulados à pobreza, para que as potencialidades analíticas oferecidas pelo método venham, de fato, beneficiar esse campo de estudos.

Entretanto, ainda que tenham sido ressaltados todos esses méritos da MEE, cabe notar que existem limites para a sua aplicação em análises de fenômenos sociais. Em princípio, deve-se colocar que essa técnica é fortemente dependente da variação dos elementos considerados. Com isso, seus resultados podem ser ofuscados quando se referem a populações marcadas por certa homogeneidade de características. Dito de modo mais específico, tal limitação refere-se à possibilidade de que relações

teoricamente esperadas entre variáveis não se mostrem significativas no modelo empírico, caso os espectros de variação desses elementos sejam insuficientes.

Além disso, é necessário reforçar que a aplicação da MEE mostra-se adequada apenas a abordagens que primem por captar a estrutura de fenômenos complexos. Nos casos de estudos que tenham propósitos mais simples, sem a preocupação de uma análise aprofundada dessa estrutura, métodos tradicionais podem ser mais propícios do que a MEE. Para a rápida comparação entre várias situações socioeconômicas distintas, por exemplo, indicadores sintéticos permitem construir tabelas para consulta instantânea; seria difícil comparar centenas de modelos de equações estruturais com a mesma agilidade. Da mesma forma, para uma simples averiguação da associação entre poucas variáveis observáveis, a regressão pode dar conta da tarefa sem que seja necessária uma sofisticada modelagem sobre as inter-relações dos fatores. Não se tem aqui, portanto, a pretensão de indicar a MEE como um método quantitativo genericamente substitutivo aos tradicionais, mas como uma alternativa analítica para o avanço do estudo dos fenômenos sociais como objetos relacionais.

REFERÊNCIAS

ANAND, S.; SEN, A. Concepts of human development and poverty: a multidimensional perspective. In: FUKUDA-PARR, S.; KUMAR, A. K. S. (Orgs.). *Readings in human development: concepts, measures and policies for a development paradigm*. Oxford: Oxford University Press, p. 204-219, 2003a.

_____. Human development index: methodology and measurement. In: FUKUDA-PARR, S.; KUMAR, A. K. S. (Orgs.). *Readings in human development: concepts, measures and policies for a development paradigm*. Oxford: Oxford University Press, p. 115-125, 2003b.

BABBIE, E. *Métodos de pesquisas de survey*. Belo Horizonte: Ed. UFMG, 1999.

BECCARIA, L.; FERES, J. C.; SÁINZ, P. *Poverty measurement present status of concepts and methods*. Bangkok: Eclac, 1999.

BEDEIAN, A. G.; DAY, D. V.; KELLOWAY, E. K. Correcting for measurement error attenuation in structural equation models: some important reminders. *Educational and Psychological Measurement*, v. 57, n. 5, p. 785-799, 1997 (Copyright Holder: Sage Publications, Inc.)

BENTLER, P. M.; WU, E. J. C. *EQS 6 for windows user's guide*. Encino: Multivariate Software Inc., 2002.

BETTI, G.; D'AGOSTINO, A.; NERI, L. *Panel regression models for measuring multidimensional poverty dynamics*. Colchester: Institute for Social and Economic Research of the University of Essex, 2000.

BLUM, Z.; ROSSI, P. H. Social class research and images of the poor: a bibliographic review. In: MOYNIHAN, D. P. (Org.). *On understanding poverty: perspectives from the social sciences*. New York: American Academy of Arts and Sciences, p. 343-397, 1969.

BOLLEN, K. A. *Structural equations with latent variables*. New York: John Wiley & Sons, Inc., 1989.

BOOMSMA, A. Reporting analyses of covariance structures. *Structural Equation Modeling*, v. 7, n. 3, p. 461-482, Copyright Holder: Lawrence Erlbaum Associates Inc., 2000.

BOURDIEU, P. *O poder simbólico*. Rio de Janeiro: Bertrand Brasil, 1998.

BYRNE, B. M. *Structural equation modeling with EQS and EQS/windows: basic concepts, applications and programming*. Thousand Oaks: Sage Publications, 1994.

CEPAL. *Anuario Estadístico de América Latina y el Caribe 2003*. Santiago: Naciones Unidas, 2004.

CONDE, M. M. Los indicadores de pobreza utilizados em Colombia: una crítica. In: *Pobreza y desigualdad: reflexiones conceptuales y de medición*. Santafé de Bogotá: Cinep, p. 139-155, 1999.

CPS/FGV. *Mapa do fim da fome: metas sociais contra a miséria nos municípios baianos*. São Paulo: Centro de Pesquisas Sociais da Fundação Getulio Vargas, 2001.

—————. *Mapa do fim da fome II*. São Paulo: Centro de Pesquisas Sociais da Fundação Getulio Vargas, 2004

DATT, G. *et al. Pobreza e bem-estar em Moçambique: primeira avaliação nacional (1996-97)*. Ministério do Plano e Finanças, Universidade Eduardo Mondlane, Instituto Internacional de Pesquisa em Políticas Alimentares, dez. 1998.

DIAS, T. de L.; OLIVEIRA, M. da P.; CÂMARA, G.; CARVALHO, M. S. Problemas de escala e a relação área-indivíduo em análise espacial de dados censitários. *Revista IP - Informática Pública*, Belo Horizonte: Centro de Desenvolvimento e Estudos da Empresa de Informática e Informação do Município de Belo Horizonte, Ano 4, n. 1, p. 89-104, jun. 2002.

DOBSON, A. J. *An introduction to generalized linear models*. Boca Raton: Chapman & Hall, 2002.

DRAPER, N. R.; SMITH, H. *Applied regression analysis*. New York: John Wiley & Sons Inc., 1998.

DUNCAN, O.; DUDLEY, P. Analysis: sociological examples. In: BLALOCK JR., H. M. (Org.). *Causal models in the social sciences*. Chicago: Aldine Publishing Company, p. 115-138, 1971.

EDWARDS, J. R.; BAGOZZI, R. P. On the nature and direction of relationships between constructs and measures. *Psychological Methods*, v. 5, n. 2, p. 155-174. Copyright Holder: American Psychological Association, 2000.

FINCHAM, F. D.; BEACH, S. R. H.; HAROLD, G. T.; OSBORNE, L. N. Marital satisfaction and depression: different causal relationships for men and women? *Psychological Science*, v. 8, n. 5, p. 351-357, 1997 (Copyright Holder: American Psychological Association).

FITOUSSI, J.-P.; ROSANVALLON, P. *La nueva era de las desigualdades*. Buenos Aires: Ediciones Manantial, 1998.

FUNDAÇÃO GETULIO VARGAS. *Mapa do fim da fome*. São Paulo: Centro de Políticas Sociais, FGV, 2001.

HAQ, Mahbub ul. The birth of the human development index. In: FUKUDA-PARR, S.; KUMAR, A. K. S. (Orgs.). *Readings in human development: concepts, measures and policies for a development paradigm*. Oxford: Oxford University Press, p. 103-113, 2003.

HOWELL, D. C. *Statistical methods for psychology*. Pacific Grove: Duxbury, 2002.

HOX, J. J.; BECHGER, T. M. An introduction to structural equation modeling. *Family Science Review*, Minneapolis, v. 11, p. 354-373, 1998.

HOYLE, R. H.; SMITH, G. T. Formulating clinical research hypotheses as structural equation models: a conceptual overview. *Journal of Consulting and Clinical Psychology*, v. 62, n. 3, p. 429-440, 1994 (Copyright Holder: American Psychological Association).

HUBER, P. J. *Robust statistical procedures*. Philadelphia: SIAM, 1977.

- IBGE. *Síntese de indicadores sociais 2000*. Rio de Janeiro: IBGE, 2001.
- JANNUZZI, P. de M. *Indicadores sociais no Brasil*. Campinas: Alínea, 2001.
- KAPLAN, D. *Structural equation modeling: foundations and extensions*. Thousand Oaks: Sage Publications, 2000.
- KENNY, D. A.; KASHY, D. A.; BOLGER, N. Data analysis in social psychology. In: GILBERT, D.; FISKE, E.; LINDZEY, G. (Orgs.). *Handbook of Social Psychology*, v. 1, p. 252-259, Boston: McGraw-Hill, 1998.
- KISH, L. *Survey sampling*. New York: John Wiley & Sons Inc., 1995.
- KLEM, L. Structural equation modeling. In: GRIMM, L. G.; YARNOLD, P. R. (Eds.). *Reading and understanding more multivariate statistics*. Washington, D. C.: American Psychological Association. Copyright Holder: American Psychological Association, 2000.
- KLINE, R. B. *Principles and practice of structural equation modeling*. New York: The Guilford Press, 1998.
- MACCALLUM, R. C.; WEGENER, D. T.; UCHINO, B. N.; FABRIGAR, L. R. The problem of equivalent models in applications of covariance structure analysis. *Psychological Bulletin*, v. 114, n. 1, p. 184-199, 1993 (Copyright Holder: American Psychological Association).
- MCCLAVE, J.; BENSON, T.; GEORGE, P.; SINCICH, T. *Statistics for business and economics*. New York: Prentice Hall, 2004.
- MCDONALD, R. P.; RINGO HO, M.-H. Principles and practice in reporting structural equation analyses. *Psychological Methods*, v. 7, n.1, p. 64-82, Washington, D. C.: American Psychological Association, 2002.
- MCGEE, R.; BROCK, K. *From poverty assessment to policy change: processes, actors and data*. Brighton: Institute of Development Studies, Jul. 2001 (IDS Working Paper, n. 133). Disponível em: <<http://www.ids.ac.uk/ids/bookshop/wp/wp133.pdf>>.
- MARSHALL, G. N.; LANG, E. L. Optimism, self-mastery, and symptoms of depression in women professionals. *Journal of Personality and Social Psychology*, v. 59, n. 1, p. 132-139, 1990 (Copyright Holder: American Psychological Association).
- MEDEIROS, M. O que faz os ricos: um estudo sobre fatores que determinam a riqueza. Tesde (Doutorado) – Universidade de Brasília, Departamento de Sociologia, Brasília, 2003.
- MEJÍA, J. A.; VOS, R. *Poverty in Latin America and the Caribbean – an inventory*. 1980-95. Washington, D. C.: The Inter-American Development Bank, 1999.
- MENEZES-FILHO, N.; VASCONCELLOS, L. Has economic growth been pro-poor in Brazil? Why? A country case study on Brazil. In: *Operationalising propoor growth: a joint initiative of AFD, BMZ (GTZ, KfW Development Bank), DFID, and the World Bank*. São Paulo: University of São Paulo, Oct. 2004.
- MESTRUM, F. *Mondialisation et pauvreté*. de l'utilité de la pauvreté dans le nouvel ordre mondial. Paris: L'Harmattan, 2002.
- MOORE, D. *The basic practice of statistics*. New York: Freeman, 2004.
- MORETTIN, L. G. *Estatística básica*. Volume 2 – INFERÊNCIA. São Paulo: Makron Books, 2000.
- PAULA, G. A. *Modelos de regressão com apoio computacional*. São Paulo: Instituto de Matemática e Estatística da Universidade de São Paulo, 2004.
- PIERCE, P. F.; VINOKUR, A. D.; BUCK, C. L. Effects of war-induced maternal separation on children's adjustment during the Gulf War and two years later. *Journal of Applied Social Psychology*, v. 28, n. 14, p. 1.286-1.311. Copyright Holder: V. H. Winston e Sons, Inc., 1998.

- PNUD. *Human Development Report*. New York: Oxford University Press, 1997.
- . *Human Development Report*. New York: Oxford University Press, 1999.
- POCHMANN, M. *et al. Atlas da inclusão social no Brasil*. 2003.
- RAVALLION, M. *Poverty comparisons*. Chur: Harwood Academic Publishers, 1994.
- RAWORTH, K.; STEWART, D. Critiques of the human development index: a review. In: FUKUDA-PARR, S.; KUMAR, A. K. S. (Orgs.). *Readings in human development: concepts, measures and policies for a development paradigm*. Oxford: Oxford University Press, p. 141-151, 2003.
- RAYKOV, T.; TOMER, A.; NESSELROADE, J. R. Reporting structural equation modeling results in psychology and aging: some proposed guidelines. *Psychology and Aging*, v. 6, n. 4, p. 499-503, 1991 (Copyright Holder: Copyright Holder: American Psychological Association).
- ROCHA, S. *Pobreza no Brasil: afinal, de que se trata?* Rio de Janeiro: Editora FGV, 2003.
- SALAMA, P.; DESTREMAU, B. *O tamanho da pobreza: economia política da distribuição de renda*. Petrópolis: Garamond, 2001.
- SALAMA, P.; VALIER, J. *Pobrezas e desigualdades no terceiro mundo*. São Paulo: Nobel, 1997.
- SCHWARTZMAN, S. Brazil: expert group on poverty statistics. In: *Seminar on poverty statistics*. Santiago, 1997.
- . *As causas da pobreza*. Rio de Janeiro, FGV, 2004.
- SLIWIANY, R. M. *Sociometria: como avaliar a qualidade de vida e projetos sociais*. Petrópolis: Vozes, 1997.
- SPOSATI, A. *et al. Mapa da exclusão/inclusão social da cidade de São Paulo: dinâmica social nos anos 90*. 2000.
- SZÉKELY, M. *Explaining poverty*. Oxford: St. Anthony's College, Queen Elizabeth House Development Studies Oct. 1996 (Working Papers, n. 98).
- TOMAS, J. M.; OLIVER, A. Rosenberg's self-esteem scale: two factors or method effects. *Structural Equation Modeling*, v. 6, n. 1, p. 84-98, 1999 (Copyright Holder: Lawrence Erlbaum Associates Inc.).
- TURNER, M. E.; STEVENS, C. D. The regression analysis of causal paths. In: BLALOCK, JR., H. M. (Org.). *Causal models in the social sciences*, p. 75-100, Chicago: Aldine Publishing Company, 1971.
- VERHINE, R. E. Educação e mercado de trabalho. In: CASTRO, N.; FLAVO, L. (Orgs.). *População, educação e emprego*. Salvador: UFBA/CRH, 1982.
- VINOKUR, A. D.; SCHUL, Y. Mastery and inoculation against setbacks as active ingredients in the jobs intervention for the unemployed. *Journal of Consulting and Clinical Psychology*, v. 65, n. 5, p. 867-877, 1997 (Copyright Holder: American Psychological Association).
- . The web of coping resources and pathways to reemployment following a job loss. *Journal of Occupational Health Psychology*, v. 7, n. 1, p. 68-83, 2002 (Copyright Holder: American Psychological Association).
- VINOKUR, A. D.; PIERCE, P. F.; BUCK, C. L. Work-family conflicts of women in the air force: their influence on mental health and functioning. *Journal of Organizational Behavior*, v. 20, p. 865-878, 1999 (Copyright Holder: John Wiley e Sons, Ltd.).
- VINOKUR, A. D.; PRICE, R. H.; CAPLAN, R. D. Hard times and hurtful partners: how financial strain affects depression and relationship satisfaction of unemployed persons and

their spouses. *Journal of Personality and Social Psychology*, v. 71, n. 1, p. 166-179, 1996 (Copyright Holder: American Psychological Association).

WHITE, H.; KILLICK, T. *African poverty at the millennium: causes, complexities and challenges*. Washington, D. C.: The World Bank, 2001.

WILBER, G. *Poverty: a new perspective*. Lexington: The University Press of Kentucky, 1975.

WISNIK, J. M. *O som e o sentido: uma outra história das músicas*. São Paulo: Companhia das Letras, 1989.

WORLD BANK. *World Development Indicators 2004 (WDI)*. Washington, D. C.: World Bank, 2004.

XIE, Y.; POWERS, D. A. *Statistical methods for categorical data analysis*. San Diego: Academic Press, 2000.

EDITORIAL

Coordenação

Iranilde Rego

Revisão

Lucia Duarte Moreira

Alejandro Sainz de Vicuña

Eliezer Moreira

Elisabete de Carvalho Soares

Miriam Nunes da Fonseca

Editoração

Roberto das Chagas Campos

Aeromilson Mesquita

Camila Guimarães Simas

Camila Oliveira de Souza

Carlos Henrique Santos Vianna

Brasília

SBS – Quadra 1 – Bloco J – Ed. BNDES,

9º andar – 70076-900 – Brasília – DF

Fone: (61) 3315-5090

Fax: (61) 3315-5314

Correio eletrônico: editbsb@ipea.gov.br

Rio de Janeiro

Av. Nilo Peçanha, 50/609

20044-900 – Rio de Janeiro – RJ

Fone: (21) 3515-8522 – 3515-8426

Fax (21) 3515-8585

Correio eletrônico: editrj@ipea.gov.br

Tiragem: 135 exemplares