

Ferreira de Souza, Pedro Herculano Guimarães

Working Paper

A distribuição de renda nas pesquisas domiciliares brasileiras: Harmonização e comparação entre censos, PNADS e POFS

Texto para Discussão, No. 1832

Provided in Cooperation with:

Institute of Applied Economic Research (ipea), Brasília

Suggested Citation: Ferreira de Souza, Pedro Herculano Guimarães (2013) : A distribuição de renda nas pesquisas domiciliares brasileiras: Harmonização e comparação entre censos, PNADS e POFS, Texto para Discussão, No. 1832, Instituto de Pesquisa Econômica Aplicada (IPEA), Brasília

This Version is available at:

<https://hdl.handle.net/10419/91288>

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.

1832

TEXTO PARA DISCUSSÃO

A DISTRIBUIÇÃO DE RENDA NAS PESQUISAS DOMICILIARES BRASILEIRAS: HARMONIZAÇÃO E COMPARAÇÃO ENTRE CENSOS, PNADS E POFS

Pedro Herculano Guimarães Ferreira de Souza

A DISTRIBUIÇÃO DE RENDA NAS PESQUISAS DOMICILIARES BRASILEIRAS: HARMONIZAÇÃO E COMPARAÇÃO ENTRE CENSOS, PNADS E POFS*

Pedro Herculano Guimarães Ferreira de Souza**

* O autor agradece os comentários de Rodolfo Hoffmann, Marcelo Medeiros e Fabio Veras Soares.

** Técnico de Planejamento e Pesquisa da Diretoria de Estudos e Políticas Sociais (Disoc) do Ipea.

Governo Federal

**Secretaria de Assuntos Estratégicos da
Presidência da República**
Ministro interino Marcelo Côrtes Neri

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
Marcelo Côrtes Neri

Diretor de Desenvolvimento Institucional
Luiz Cezar Loureiro de Azeredo

**Diretor de Estudos e Relações Econômicas e
Políticas Internacionais**
Renato Coelho Baumann das Neves

**Diretor de Estudos e Políticas do Estado, das
Instituições e da Democracia**
Daniel Ricardo de Castro Cerqueira

**Diretor de Estudos e Políticas
Macroeconômicas**
Cláudio Hamilton Matos dos Santos

**Diretor de Estudos e Políticas Regionais,
Urbanas e Ambientais**
Rogério Boueri Miranda

**Diretora de Estudos e Políticas Setoriais
de Inovação, Regulação e Infraestrutura**
Fernanda De Negri

Diretor de Estudos e Políticas Sociais
Rafael Guerreiro Osorio

Chefe de Gabinete
Sergei Suarez Dillon Soares

**Assessor-chefe de Imprensa e
Comunicação**
João Cláudio Garcia Rodrigues Lima

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

Texto para Discussão

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

© Instituto de Pesquisa Econômica Aplicada – **ipea** 2013

Texto para discussão / Instituto de Pesquisa Econômica Aplicada.- Brasília : Rio de Janeiro : Ipea , 1990-

ISSN 1415-4765

1. Brasil. 2. Aspectos Econômicos. 3. Aspectos Sociais.
I. Instituto de Pesquisa Econômica Aplicada.

CDD 330.908

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.

JEL: I31, I32.X

SUMÁRIO

SINOPSE

ABSTRACT

1 INTRODUÇÃO	7
2 CARACTERIZAÇÃO GERAL DAS PESQUISAS.....	9
3 TRATAMENTO DOS DADOS	13
4 A COMPARAÇÃO ENTRE OS CENSOS, AS PNADS E AS POFS	40
5 CONCLUSÃO	54
REFERÊNCIAS	57
APÊNDICE	62

SINOPSE

O objetivo deste trabalho é documentar e explicar as diferenças nas distribuições de renda dos censos demográficos, das Pesquisas Nacionais por Amostra de Domicílios (PNADs) e das Pesquisas de Orçamentos Familiares (POFs). A principal hipótese é que um número relativamente pequeno de procedimentos de harmonização, que aproxima o desenho das três pesquisas, é capaz de promover grande convergência dos resultados. Os resultados confirmam, em larga medida, esta hipótese: o retrato que emerge do Brasil nos censos, nas PNADs e nas POFs é coerente e robusto após a harmonização. Ainda que haja pequenas variações quanto aos níveis de renda, desigualdade e pobreza, pode-se afirmar que, pelo menos de acordo com os dados disponíveis até o momento, o crescimento da renda e a queda da desigualdade e da pobreza podem ser considerados fenômenos bem estabelecidos. Grande parte das discrepâncias entre censos, PNADs e POFs decorre de questões amostrais, conceituais e de tratamento que podem ser minimizadas com os procedimentos sugeridos. O desenho de cada pesquisa influencia fortemente seus resultados e, portanto, não se deve estranhar que os números produzidos não sejam diretamente comparáveis.

Palavras-chave: pesquisas domiciliares; distribuição de renda; desigualdade; pobreza.

ABSTRACTⁱ

This paper documents and tries to explain the discrepancies between the income distributions reported by the three major household surveys in Brazil: the Census, the Pesquisa Nacional por Amostra de Domicílios (PNAD) and the Pesquisa de Orçamentos Familiares (POF). The main hypothesis is that key differences in survey design are responsible for these results, which can be remedied by a relatively small number of harmonization procedures. Our analysis largely confirms this hypothesis: the harmonized data yields very consistent results in the Census, the PNAD and the POF. Although the levels of income, inequality and poverty are not exactly the same, the trends are very similar: income growth, the fall in inequality and poverty reduction in the past decade can be considered well-established facts. Most of the differences between the income data in the three surveys can be explained by disparities in sample design, conceptual framework and the statistical treatment of the data. Thus, it is not surprising that the three surveys report divergent figures, as they are not directly comparable.

Keywords: household surveys; income distribution; income inequality; income poverty.

i. As versões em língua inglesa das sinopses desta coleção não são objeto de revisão pelo Editorial do Ipea.
The versions in English of the abstracts of this series have not been edited by Ipea's publishing department.

1 INTRODUÇÃO

A primeira década de 2000 trouxe pelo menos duas boas notícias para os pesquisadores interessados na distribuição de renda no Brasil. Uma delas, o crescimento econômico pró-pobre, isto é, com diminuição da desigualdade, tem sido objeto de inúmeras investigações (Hoffmann, 2005; Soares, 2006; Barros *et al.*, 2006; entre outros), que só foram possíveis em função de outra novidade muito pouco comentada: a profusão de bancos de dados de alta qualidade, com microdados publicamente disponíveis. A década do início da queda da desigualdade foi, sem dúvida, a mais bem documentada desde o começo dos estudos empíricos sobre a distribuição de renda no país, na década de 1970.¹

Essa abundância de dados resulta de um esforço de longo prazo, especialmente do Instituto Brasileiro de Geografia e Estatística (IBGE), e é extremamente positiva, pois permite que os pesquisadores avaliem a robustez de suas descobertas. No entanto, ela também desperta algumas dúvidas, já que nem sempre os números obtidos são compatíveis entre si.

Este trabalho tem como objetivo principal explicar as discrepâncias nas distribuições de renda nas três pesquisas domiciliares regularmente conduzidas pelo IBGE que têm cobertura nacional e coletam informações sobre rendimentos: os censos demográficos, as Pesquisas Nacionais por Amostra de Domicílios (PNADs) e as Pesquisas de Orçamentos Familiares (POFs).²

Para isso, o texto procede em duas etapas: primeiro, as principais diferenças entre as três fontes de dados são agregadas em três grandes grupos – diferenças amostrais, conceituais e na coleta e tratamento das informações –, documentadas e, dentro do possível, harmonizadas. Em seguida, as distribuições e alguns indicadores mais importantes são comparados antes e depois da harmonização. Na análise, as PNADs são tomadas como parâmetro de referência, isto é, os censos e as POFs são comparados

1. Ao longo de todo o texto, a expressão “queda da desigualdade” diz respeito à desigualdade da renda domiciliar *per capita*. Com efeito, a desigualdade de rendimentos para algumas fontes específicas – como a renda das pessoas economicamente ativas – começou a cair já em meados da década de 1990 (Hoffmann, 2002).

2. O Instituto Brasileiro de Geografia e Estatística (IBGE) possui uma quarta pesquisa domiciliar com informações sobre rendimentos e periodicidade regular, a Pesquisa Mensal do Emprego (PME), mas sua abrangência geográfica restringe-se a seis regiões metropolitanas e o questionário contempla apenas os rendimentos do trabalho.

com ela, mas não entre si, o que se justifica tanto por questões técnicas quanto pelo maior uso das PNADs, decorrente, em larga medida, de sua periodicidade anual.

A principal hipótese do trabalho é que um conjunto relativamente pequeno de ajustes é capaz de promover uma grande convergência nas distribuições de renda nos censos, nas PNADs e nas POFs, o que permite que se chegue a conclusões mais robustas sobre o processo observado nos últimos anos. Como a maior parte destes ajustes tem como objetivo tornar os censos e as POFs mais parecidos com as PNADs – e não o inverso –, a hipótese de trabalho pode também ser colocada em outros termos: no que diz respeito aos rendimentos, as PNADs medem bem aquilo que elas são desenhadas para medir e permitem inferências adequadas tanto sobre os níveis quanto sobre as tendências dos rendimentos ao longo do tempo.

Esse objetivo é diferente da abordagem usual da literatura, que normalmente tenta validar os dados de rendimentos das pesquisas domiciliares comparando-os com as referências externas, as quais, presumivelmente, representariam os valores “corretos”. As estratégias mais comuns são a comparação de agregados com registros administrativos, para testar a participação e os rendimentos vinculados a programas governamentais, ou com as contas nacionais, para avaliar a renda total das famílias e sua composição.

Em geral, os resultados mostram que, no Brasil, as pesquisas domiciliares tendem a subestimar o nível dos rendimentos e a participação em alguns programas governamentais, como o Bolsa Família, mas não há consenso quanto ao nível de desigualdade (Lluch, 1982; Hoffmann, 1988; Barros; Cury; Ulysea, 2006; Hoffmann; Ney, 2008; Souza, 2010). A literatura internacional chega a resultados parecidos. Primeiro, a participação em programas governamentais e os rendimentos associados a eles são quase sempre subestimados, inclusive com tendência de piora ao longo do tempo, embora haja grande variação entre programas (Mathiowetz, Brown e Bound, 2001; Weinberg *et al.*, 1999; Wheaton, 2007; Meyer, Moke e Sullivan, 2009). Segundo, algumas fontes de rendimentos – como renda de propriedades, juros e afins, ou rendimentos de empregadores e trabalhadores por conta própria – também são mal captadas nas pesquisas domiciliares (Canberra Group, 2001; Hurst, 2010). Terceiro, os resultados tendem a ser sensíveis ao desenho e à estrutura das pesquisas, e, portanto, variam bastante entre *surveys* (Coder e Scoon-Rogers, 1996; Moore, Stinson e Welniak Junior, 2000; Gouskova e Schoeni, 2007).

Neste trabalho, optou-se por outra estratégia. A ideia não é comparar os resultados com números “verdadeiros”, mas sim explicar as diferenças entre as pesquisas. Com isso, um dos pressupostos da análise é que um fenômeno pode ser considerado bem estabelecido sempre que as três diferentes fontes de dados – censos, PNADs e POFs – concordarem entre si, independentemente de sua validação ou não por fontes exógenas. Naturalmente, é possível que haja erros ou problemas comuns às três fontes de dados, como, por exemplo, a captação de todos os rendimentos dos mais ricos. No entanto, não há como investigar tais problemas sem o auxílio de fontes externas, o que está além do escopo deste texto.

Comparações parciais entre os censos, as PNADs e as POFs já foram feitas por alguns autores, junto com o contraste com as contas nacionais, mas sempre em apenas um momento do tempo, nunca entre as três pesquisas simultaneamente.³ Este trabalho tenta preencher esta lacuna, cobrindo todas as edições realizadas desde o Plano Real, o que inclui dois censos (2000 e 2010), quinze PNADs (1995-2011) e duas POFs com cobertura nacional (2002-2003 e 2008-2009, ou, para simplificar, 2002 e 2008).

2 CARACTERIZAÇÃO GERAL DAS PESQUISAS

O censo demográfico foi a primeira pesquisa domiciliar brasileira, realizado pela primeira vez em 1872, ainda durante o Império. Desde então, a periodicidade decenal manteve-se constante, com poucas exceções,⁴ mas o escopo dos dados coletados cresceu continuamente. Com efeito, o período após a Segunda Guerra Mundial, caracterizado pela maior demanda por dados e por acordos de cooperação internacional, marcou uma nova fase para os censos, tanto no que diz respeito à quantidade quanto ao que se refere à qualidade das informações, em especial a partir de 1970 (Médici, 1986; Oliveira e Simões, 2005).

A coleta de informações sobre renda é um exemplo dessa evolução. Sua primeira aparição foi no Censo de 1890, cujas tabulações só foram divulgadas, de forma incompleta, cerca de dez anos depois. Os resultados sobre rendimentos nunca foram

3. Para mais informações acerca desse tema, ver, entre outros, Lluch (1982); Barros, Cury e Ulyseia (2006); e Hoffmann e Ney (2008).

4. Em função de turbulências políticas, os censos de 1910 e 1930 não foram realizados (Médici, 1986). Mais recentemente, por causa da conjuntura de crise política e econômica, o Censo de 1990 foi a campo somente em 1991 (Oliveira e Simões, 2005).

publicados, aparentemente em função da má qualidade dos dados e das altas taxas de não resposta. Depois disso, os rendimentos só reapareceram no Censo de 1960 – o primeiro a incluir a distinção entre questionários da amostra e do universo –, na forma de uma única questão, com respostas fechadas em múltiplos do salário mínimo (Médici, 1986). A partir daí, a renda continuou a ser coletada em todos os censos subsequentes, com nível de detalhamento crescente até 2000, quando foram pesquisadas oito diferentes fontes de rendimentos – trabalho principal; demais trabalhos; Previdência pública; alugueis; pensão alimentícia, mesada e doações; programas sociais; outros rendimentos.

O Censo de 2010 reverteu essa tendência, discriminando os rendimentos em apenas três fontes – trabalho principal; demais trabalhos; outros rendimentos. Os resultados empíricos da literatura internacional sugerem que a maior agregação pode introduzir um duplo viés, pois tende a levar à subestimação do nível da renda e, pelo menos em algumas pesquisas, a superestimar a desigualdade, uma vez que os mais pobres são mais afetados (Davern *et al.*, 2005). Contudo, como será visto mais adiante, isso não parece ter ocorrido neste censo, pelo menos quanto ao nível da renda, talvez em função de perguntas indicadoras sobre o recebimento ou não das “outras rendas” e/ou do treinamento dos recenseadores.

Cabe notar que os censos brasileiros são dos poucos que coletam regularmente informações sobre rendimentos. O próprio documento oficial da Organização das Nações Unidas (ONU) sobre censos populacionais recomenda que o tema é mais apropriado para pesquisas domiciliares por amostragem, dados os custos e a dificuldade de obter informações de boa qualidade (ONU, 2008). Apenas nove dos 69 países (13%) com microdados de censos populacionais disponíveis no Integrated Public Use Microdata Series – International (IPUMS-International) coletam, ou coletaram em algum momento, informações sobre a renda total dos indivíduos. Somente dezessete (14%) dos 123 censos coletados a partir de 1990 contêm esta variável, incluindo dois censos brasileiros (1991 e 2000).⁵

5. O Integrated Public Use Microdata Series – International (IPUMS-International) é o maior repositório público de microdados de censos populacionais no mundo. Mantido pelo Minnesota Population Center, da Universidade de Minnesota, o IPUMS-International oferece, até o momento, microdados harmonizados de 210 censos de 69 países, cobrindo o período entre 1960 e 2010. Até dezembro de 2012, o censo brasileiro de 2010 ainda não estava disponível no *site*. Para mais informações, ver <<http://international.ipums.org/international/>>.

As PNADs surgem como parte do esforço de aperfeiçoamento das estatísticas nacionais no Pós-Guerra, também sob a influência da intensificação da cooperação internacional, mas com objetivos distintos dos censos. Sua criação deriva do chamado “Plano Atlântida”, concebido pelo US Bureau of the Census e promovido, na América Latina, pela United States Agency for International Development (USAID) e pelo Inter-American Statistical Institute (IASI). O plano era direcionado a países com informações básicas precárias e detalhava como levar a campo pesquisas que permitissem o monitoramento contínuo das condições socioeconômicas do país, com atenção especial para o mercado de trabalho (Martine *et al.*, 1988; Médici, 1988; Martine, 2005).

A PNAD foi conduzida pela primeira vez pelo IBGE em 1967, com caráter trimestral e abrangência geográfica restrita. Tanto o formato geral quanto a periodicidade da pesquisa ficaram basicamente constantes até 1973, com mudanças apenas na abrangência geográfica, que se tornou mais inclusiva, e em alguns pontos do questionário. Depois da interrupção, em 1970, para a realização do censo, a PNAD novamente deixou de ir a campo em 1974 e 1975, em função do Estudo Nacional de Despesas Familiares (ENDEF). Seu retorno, em 1976, marcou uma nova etapa, com mais mudanças no questionário e a periodicidade anual, exceto em anos censitários.⁶

Entre 1976 e 1979, o questionário continuou a sofrer alterações; mas, em 1981, seu desenho básico se consolidou, não havendo grandes mudanças estruturais nos últimos trinta anos. Neste período, houve apenas a continuação da expansão da cobertura geográfica para as áreas residuais até então não incluídas – como as áreas rurais da região Norte, as últimas a entrar nas PNADs, em 2004 – e mudanças pontuais no questionário e em algumas definições, sem prejudicar a comparabilidade da série histórica.

No caso da captação dos rendimentos, desde 1981, houve apenas pequenas mudanças nas perguntas sobre os rendimentos do trabalho e uma maior desagregação dos “outros rendimentos”, principalmente com a pergunta, a partir de 1992, para registrar as pensões alimentícias e as aposentadorias e pensões da Previdência privada, que elevou para onze os tipos de rendimentos pesquisados. Apesar de positiva, tamanha

6. De acordo com o então presidente do IBGE, Simon Schwartzman, os cortes orçamentários de 1992 e 1993 e a não aprovação do orçamento de 1994 no prazo apropriado forçaram a alteração do cronograma das pesquisas. Com isso, e como o processamento dos dados das pesquisas nacionais por amostra de domicílios (PNADs) 1992 e 1993 ainda não havia terminado, a PNAD de 1994 acabou não sendo realizada (Schwartzman, 1994).

consistência histórica também produziu efeitos colaterais indesejados: por exemplo, até 2011, as PNADs continuaram captando separadamente os rendimentos do abono de permanência – extinto em 1993 –, mas nunca incluíram no questionário básico perguntas específicas sobre programas mais recentes e de importância crescente, como o Bolsa Família e o Benefício de Prestação Continuada.⁷ Isto só deverá ocorrer nos próximos anos, com a reformulação do sistema de pesquisas domiciliares do IBGE.

Finalmente, as origens das POFs remontam ao ENDEF, que foi não apenas a primeira pesquisa minuciosa e de grande abrangência a trazer dados de rendimentos e despesas das famílias como também foi a primeira a coletar informações antropométricas e relativas ao consumo alimentar.⁸ A complexidade e o alto custo desta pesquisa, no entanto, fizeram que uma nova edição viesse a ser realizada somente mais de uma década depois, no biênio 1987-1988, já com o novo nome – Pesquisa de Orçamentos Familiares – e escopo geográfico e temático mais restritos. A POF 1987-1988 pesquisou apenas as nove regiões metropolitanas criadas entre 1973 e 1974, além do Distrito Federal e de Goiânia, e não coletou dados sobre consumo não monetário, consumo alimentar ou antropometria. As edições subsequentes da POF, no entanto, reverteram por completo este movimento: tanto o consumo alimentar quanto a antropometria voltaram em 1995-1996, e as duas últimas edições – 2002-2003 e 2008-2009 – não apenas recuperaram os aluguéis estimados para as famílias que moram em residência própria como também voltaram a ter abrangência nacional (Diniz *et al.*, 2007; Vaz, 2012).

A coleta dos rendimentos nas POFs é mais exaustiva que nas PNADs e nos censos. Os microdados da primeira edição com cobertura nacional, em 2002-2003, discriminam mais de oitenta tipos de rendimentos monetários; em 2008-2009, este número subiu para cerca de 110. Além disso, ao contrário dos censos e das PNADs, as

7. Para mais detalhes acerca desse assunto, ver Médici (1988) e Rocha (2003).

8. O marco inicial dos inquéritos alimentares no Brasil foram as pesquisas de Josué de Castro, Horace Davis e Samuel Lowrie, ainda na década de 1930, baseadas em amostras não probabilísticas e de caráter estritamente local. Posteriormente, houve pesquisas maiores, como o Northeast Brazil survey, da Universidade Federal de Pernambuco (UFPE) e do Interdepartmental Committee on Nutrition for National Development (ICNND), em 1962, além das pesquisas de orçamentos familiares da Fundação Getúlio Vargas, da Fundação Instituto de Pesquisas Econômicas (Fipe), entre outras (Vasconcelos, 2001; Mondini e Monteiro, 1994; Cavalcante, Priore e Franceschini, 2004; Menezes e Osório, 2009; Yuba, 2011; IBRE/FGV, 2012). Nenhuma delas, no entanto, chegou perto do escopo, detalhamento, abrangência e custos do Estudo Nacional de Despesas Familiares (ENDEF).

POFs também recolhem informações sobre a renda não monetária (ou seja, a soma do consumo não monetário com o aluguel imputado líquido) e a tributação direta (necessária para o cálculo da renda líquida).

Apesar de todas essas transformações, as três pesquisas continuam fortemente marcadas por seus objetivos originais. Em última instância, o papel principal dos censos no sistema nacional de estatística é fornecer dados sociodemográficos para recortes que vão do Brasil como um todo até níveis geográficos bem detalhados, permitindo, portanto, a atualização das projeções demográficas e os ajustes no desenho das pesquisas por amostragem. As PNADs, por sua vez, servem para o monitoramento mais frequente e detalhado destas características gerais da população, com ênfase no mercado de trabalho. Estes dados são usados, por exemplo, para o cálculo da quantidade de ocupações e da massa salarial de atividades específicas nas contas nacionais, como empregados sem carteira assinada, trabalhadores por conta própria e empregadores informais (Hallak Neto *et al.*, 2008). As POFs, por sua vez, continuam com uma missão principal dupla: fornecer parâmetros para atualizar a estrutura de ponderação dos dois principais índices oficiais de preços – o Índice Nacional de Preços ao Consumidor (INPC) e o Índice Nacional de Preços ao Consumidor Amplo (IPCA) – e para estabelecer a parcela do consumo familiar nas contas nacionais.

3 TRATAMENTO DOS DADOS

Diante dessa heterogeneidade de origens, trajetórias e objetivos, não é surpreendente que os censos, as PNADs e as POFs apresentem diferenças significativas em todas as suas etapas, o que traz problemas para a comparação direta entre as pesquisas. O objetivo desta seção, portanto, é enumerar algumas das principais divergências e descrever os procedimentos adotados para harmonizar, na medida do possível, as três fontes de dados.

Para isso, essas divergências foram agrupadas em três categorias – plano amostral, definições conceituais e coleta e tratamento dos dados –, que são analisadas em sequência. Os resultados apresentados mais adiante comparam entre si não apenas os “dados originais” dos censos, das PNADs e das POFs mas também os dados harmonizados. Com isso, pode-se testar em que medida a harmonização das definições mais básicas provoca ou não convergência nos resultados.

Ressalte-se que a seleção das unidades de análise constituiu uma etapa preliminar à harmonização. Em consonância com a tradição brasileira, todos os resultados apresentados neste trabalho dizem respeito à distribuição da população brasileira de acordo com a renda domiciliar *per capita*. A definição de população abrange todos os moradores de domicílios particulares permanentes, exceto pensionistas, empregados domésticos e afins.

Por fim, todas as informações sobre rendimentos – tanto dos microdados harmonizados quanto dos originais – foram deflacionadas pelo INPC para 1º de outubro de 2011, seguindo a sugestão de Corseuil e Foguel (2002). A linha de pobreza extrema utilizada em todas as análises é de R\$ 70 *per capita* para os rendimentos deflacionados.

3.1 Diferenças no desenho amostral

As questões relativas ao plano amostral envolvem três aspectos principais: representatividade geográfica, tamanho e forma de seleção da amostra das pesquisas. Embora não sejam idênticos quanto a estes aspectos, as diferenças entre os censos, as PNADs e as POFs podem ser minimizadas com alguns procedimentos simples.

3.1.1 Representatividade geográfica

A representatividade geográfica é o aspecto que impõe menos restrições, em especial porque as análises abaixo se restringem ao país como um todo. A única alteração necessária é ditada pelas PNADs: até 2004, sua cobertura não incluía as áreas rurais da região Norte (exceto Tocantins), ao contrário dos censos, das POFs e das próprias PNADs de 2004 em diante. Assim, o primeiro ajuste para harmonização amostral implica descartar estas áreas. Tendo em vista que elas representam apenas cerca de 2% da amostra expandida destas pesquisas, e, portanto, sua eliminação altera muito pouco os resultados, não foi adotado nenhum procedimento para reponderar os pesos amostrais.

3.1.2 Tamanho das amostras

As PNADs têm mais ou menos o dobro do tamanho das POFs, mas ambas têm amostras bem robustas: entre 1995 e 2011, as PNADs cresceram de 85 mil para 112 mil domicílios particulares permanentes entrevistados por ano,⁹ enquanto as POFs visitaram com sucesso entre 49 mil e 56 mil domicílios. As amostras dos censos são outra história: por incluírem uma porcentagem relativamente constante do total de domicílios – 11,7% em 2000 e

9. As maiores variações nas amostras das PNADs ocorrem, em geral, logo após os censos demográficos. Por exemplo, de 1999 para 2001, a amostra aumentou de 94 para 103 mil domicílios; de 2009 para 2011, ela caiu de 121 mil – o maior valor registrado no período – para 112 mil.

10,7% em 2010 (IBGE, 2003; 2012a) –, os censos são entre cinquenta e cem vezes maiores. Há dados válidos para 5,2 milhões de domicílios em 2000 e 6,1 milhões em 2010.

Para a maior parte das variáveis, isso não representa um problema, mas o caso da renda é especial. Os rendimentos não têm nenhum limite superior necessário e sabidamente a distribuição de renda é log-normal apenas até certo ponto, aproximando-se de uma distribuição de Pareto no topo.¹⁰ Isto significa que amostras descomunais como as dos censos, ao captarem, de forma potencialmente mais eficiente, os muito ricos, podem apresentar indicadores muito distintos de amostras menores. Em outras palavras, uma característica da concentração de renda no topo é que a inclusão de pouquíssimos indivíduos pode provocar efeitos desproporcionais sobre algumas estatísticas convencionais.

A maior capacidade dos censos em localizar e entrevistar os muito ricos é uma vantagem dessa pesquisa; *ceteris paribus*, os níveis de renda e desigualdade registrados estarão mais próximos dos “verdadeiros” parâmetros populacionais. Isto se torna um problema, contudo, se o objetivo é comparar os censos com outras pesquisas. Afinal, qualquer comparação razoável envolve a harmonização dos universos pesquisados; dadas as características amostrais das três fontes de dados, a capacidade de as PNADs ou as POFs atingirem grupos tão diminutos quanto os muito ricos é tão irrisória que, para fins práticos, pode-se considerar que eles não fazem parte do universo representado por ambas.¹¹

O gráfico 1 testa isso empiricamente ao mostrar as razões entre as rendas do centésimo mais rico nos censos e nas PNADs¹² (gráfico 1A), e nas POFs e nas PNADs (gráfico 1B). No primeiro caso, há uma descontinuidade marcante entre o 0,1% mais rico; no segundo, não. Assim, nos censos, a maior parte do 1% mais rico ganha entre 25% e 50% a mais que suas contrapartes nas PNADs, mas quando se chega ao 0,1% mais rico, esta porcentagem dispara para algo entre 100% e 150% a mais. Tendo em vista que isso não ocorre nas POFs, é razoável concluir que a influência do tamanho da amostra dos censos se manifesta, no topo da distribuição, sobretudo entre este 0,1% mais rico.

10. Para mais informações acerca desse assunto, ver Inhaber e Carroll (1992).

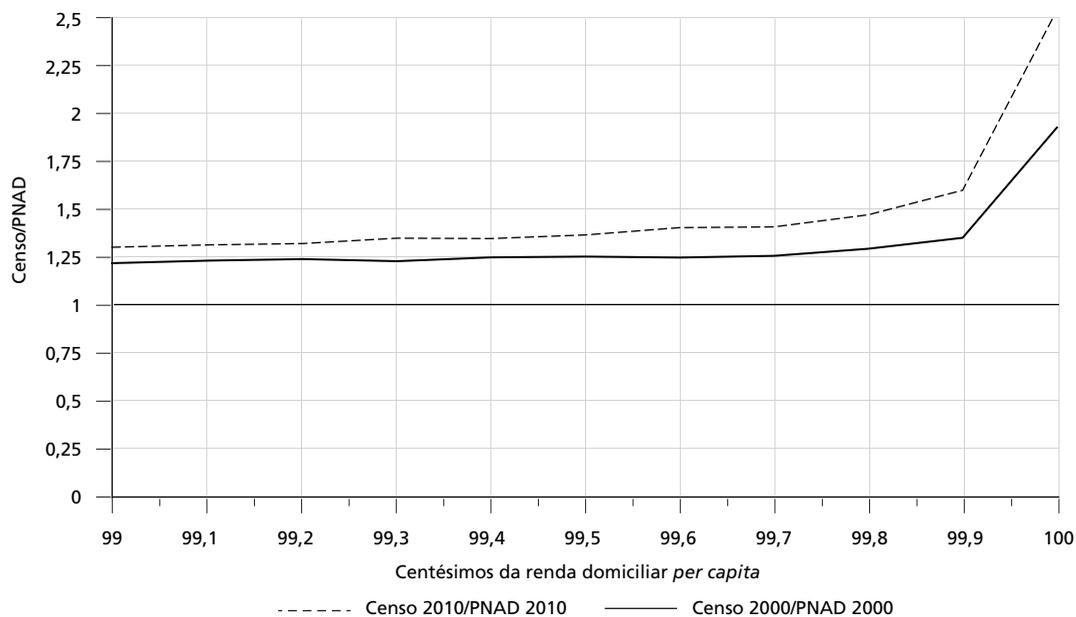
11. É razoável argumentar que pelo menos alguns desses indivíduos extraordinariamente ricos encontrados nos censos são erros de registro, resposta e afins, especialmente em decorrência da escala maciça do recenseamento, que demanda a contratação de um batalhão de entrevistadores temporários e inexperientes. Este é o ponto levantado, por exemplo, por Hoffman e Ney (2008). Neste caso, esses *outliers* deveriam ser vistos como um problema de coleta e tratamento, e não de amostra. Todavia, como o IBGE já faz uma extensa crítica para identificar, entre outras coisas, rendimentos incompatíveis com o perfil dos domicílios, o autor optou por interpretar os dados pelo seu valor de face, e, portanto, considerar como válida a informação registrada para todos os “muito ricos”.

12. A seção 3.4 explica como foram construídas as PNADs “sintéticas” para 2000 e 2010.

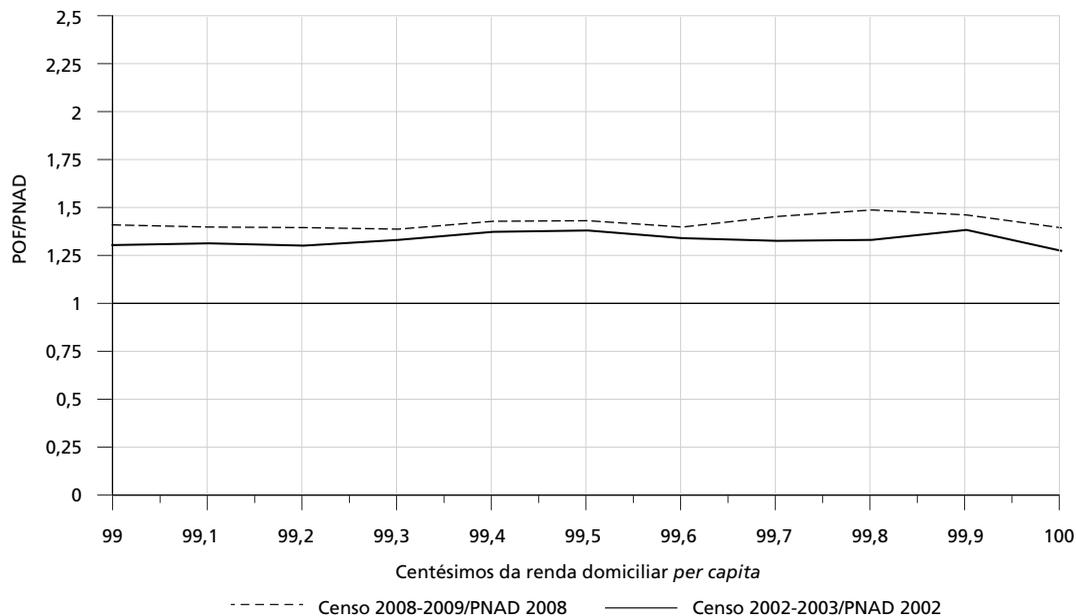
GRÁFICO 1

Razão entre as rendas dos milésimos dentro do 1% mais rico – Censo/PNAD e POF/PNAD

1A – Razão entre as rendas dos milésimos nos censos e nas PNADs



1B – Razão entre as rendas dos milésimos nas POFs e nas PNADs



Fonte: microdados das pesquisas nacionais por amostra de domicílios (PNADs) 1999, 2001, 2002, 2008, 2009 e 2011; dos censos 2000 e 2010; das POFs 2002-2003 e 2008-2009.

A tabela 1 confirma o fenômeno ao mostrar as participações médias do 0,1% mais rico na renda total e as razões médias entre a renda deste grupo e a do quantil 99,9% nos censos, nas PNADs e nas POFs. Os censos são claramente discrepantes: nestes, os muito ricos não só detêm uma porcentagem muito maior da renda total como também ganham, em média, mais que o dobro da linha de corte estabelecida pelo quantil 99,9%, contra 50% a mais nas PNADs e POFs. Logo, o segundo procedimento de harmonização das amostras consistiu na eliminação do 0,1% mais rico nos censos.

TABELA 1
Discrepâncias entre o 0,1% mais rico nos censos, nas PNADs e nas POFs

Pesquisa	Amostra média (milhares)	Participação média do 0,1% mais rico na renda total (%)	Razão entre renda média do 0,1% mais rico e a do quantil 99,9%
Censos de 2000 e 2010	20.263,9	5,7	2,2
PNADs de 1995-2011	373,0	3,0	1,5
POFs de 2002 e 2008	185,8	3,0	1,5

Fonte: microdados das PNADs 1999, 2001, 2002, 2008, 2009 e 2011; dos censos 2000 e 2010; das POFs 2002-2003 e 2008-2009.

3.1.3 Seleção das amostras

O terceiro e último aspecto amostral diz respeito à seleção das amostras propriamente ditas. As amostras dos censos incluem todos os municípios do país e “compreende[m] a seleção sistemática e com equiprobabilidade, dentro de cada setor censitário, de uma amostra dos domicílios particulares e (...) coletivos, com fração amostral constante para setores de um mesmo município” (IBGE, 2003). Embora o tamanho relativo da amostra tenha permanecido relativamente constante para o Brasil, houve mudanças nas frações amostrais por tipos de municípios. Em 2000, havia apenas duas faixas de frações amostrais: 10%, para os municípios com população estimada em mais de 15 mil habitantes, e 20%, para os demais municípios. Em 2010, o IBGE utilizou quatro faixas, que variaram entre 5%, para os municípios com mais de 500 mil habitantes, e 50%, para aqueles com população de até 2,5 mil pessoas.

As PNADs seguem uma amostragem probabilística em três estágios, na qual os municípios são as unidades primárias, os setores censitários são as unidades secundárias e os domicílios são as terciárias. De início, cada município do país é classificado em três estratos: o primeiro engloba os municípios de nove regiões metropolitanas (Belém, Belo Horizonte, Curitiba, Fortaleza, Porto Alegre, Recife, Rio de Janeiro,

São Paulo e Salvador) e o Distrito Federal (doravante região metropolitana – RM); o segundo estrato é composto pelos municípios autorrepresentativos (AR), que inclui os municípios não metropolitanos de grande porte; e o terceiro abarca os municípios não autorrepresentativos (NAR), reunindo todos os demais municípios do país.

No primeiro estágio, os municípios RM e AR são necessariamente selecionados, isto é, são sorteados com probabilidade igual a 1; os municípios NAR, por seu turno, são agrupados principalmente por critérios geográficos – para garantir a representatividade das áreas urbanas e rurais de todas as Unidades da Federação (UFs) – e sorteados, dentro dos seus grupos, com probabilidades proporcionais ao seu tamanho. No segundo estágio, os setores censitários são selecionados, em cada município, também com probabilidade proporcional ao seu tamanho e com reposição; e no terceiro e último estágio, os domicílios são sorteados com equiprobabilidade dentro de cada setor (Silva, Pessoa e Lila, 2002; IBGE, 2012b).

Finalmente, o plano amostral das POFs é conglomerado em dois estágios, com os setores censitários como unidade primária e os domicílios particulares permanentes como unidade secundária. O sorteio do primeiro estágio é feito a partir da estratificação geográfica e estatística – vale dizer, socioeconômica – das unidades primárias, que são selecionadas, dentro de cada estrato, com probabilidade proporcional ao seu tamanho; o segundo estágio, por sua vez, sorteia os domicílios por amostra aleatória simples, sem reposição (IBGE, 2004). Uma mudança da POF de 2002 para a de 2008 é que esta foi feita com base na Amostra Mestra, a qual deverá municiar as pesquisas domiciliares do IBGE nos próximos anos¹³ (Freitas *et al.*, 2007; IBGE, 2010).

A harmonização perfeita dos desenhos amostrais das três fontes é ao mesmo tempo impossível e possivelmente desnecessária, dado que todas são planejadas para representar o Brasil. No caso dos rendimentos, no entanto, é possível levantar um problema que, se verdadeiro, pode ser parcialmente sanado: é possível que parte das diferenças entre as três pesquisas decorra do fato de que, por um lado, existe no Brasil uma grande correlação

13. Além de mudanças nos questionários e na periodicidade das pesquisas propriamente ditas, uma das novidades do sistema de pesquisas domiciliares que o IBGE está desenvolvendo diz respeito à criação de uma única Amostra Mestra para subsidiar todas as pesquisas domiciliares do instituto. Trata-se de um cadastro de unidades de área escolhidas de forma probabilística, a partir do qual será possível sortear subamostras para atender às demandas das diversas pesquisas. A este respeito, ver Freitas *et al.* (2007).

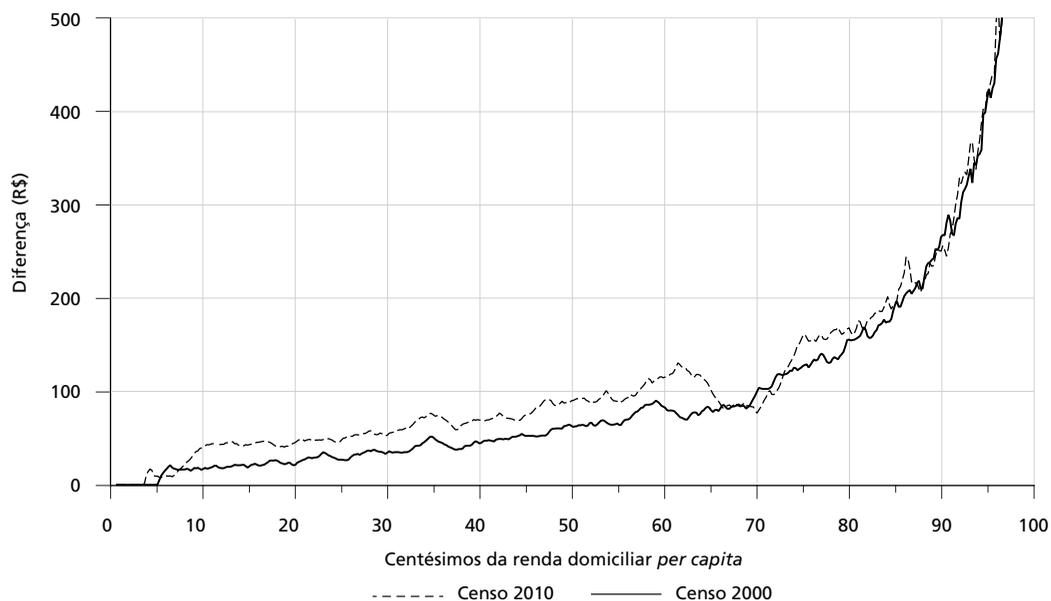
entre o tamanho dos municípios e o nível de renda *per capita*, e, por outro, o desenho amostral das PNADs tende a privilegiar os municípios não autorrepresentativos maiores e, portanto, mais ricos. Esta possibilidade ganha força quando se observa que as PNADs não só visitam menos municípios que as POFs e, obviamente, os censos, como também os municípios escolhidos – pelo menos para o período 2001-2009 – são muito mais concentrados geograficamente (IBGE, 2007).¹⁴ Com efeito, tanto no Censo 2000 quanto no Censo 2010, os municípios não autorrepresentativos selecionados para as PNADs de 2001 a 2009 (NAR-PNAD) têm, em média, mais que o dobro do tamanho dos demais municípios não autorrepresentativos (NAR-Outros).

A implicação distributiva desse viés é apresentada nos gráficos 2 e 3, que trazem, para os censos 2000 e 2010, as diferenças entre as curvas de quantis (gráfico 2) e as diferenças entre as curvas de Lorenz entre as populações dos municípios NAR-PNAD e NAR-Outros (gráfico 3). No primeiro caso, como as diferenças são sempre positivas, pode-se concluir que há dominância de primeira ordem da população dos municípios NAR-PNAD sobre a dos NAR-Outros tanto em 2000 quanto em 2010; no segundo, desconsiderando os pequenos valores negativos em torno do 90^o percentil em 2010, pode-se afirmar que a distribuição de renda nos municípios NAR-PNAD tem dominância de Lorenz sobre a dos NAR-Outros.

Em bom português, isso significa que: *i*) não só a renda nos municípios incluídos nas PNADs 2001-2009 é maior na média como também ao longo de toda a distribuição, e, assim, o bem-estar nestes municípios é inequivocamente maior que nos outros municípios tanto em 2000 quanto em 2010; *ii*) se o pequeno intervalo negativo perto do topo em 2010 for deixado de lado, a desigualdade de renda entre a população destes municípios será menor que a registrada entre a população dos municípios não autorrepresentativos, excluídos das PNADs para qualquer uma das medidas de desigualdade que obedeça ao princípio das transferências, o que abarca todas as medidas convencionais de desigualdade; e *iii*) conseqüentemente, para qualquer linha absoluta de pobreza, tanto a incidência quanto a intensidade da pobreza serão menores na população dos municípios NAR sorteados para as PNADs 2001-2009 que para a população dos demais municípios não autorrepresentativos.

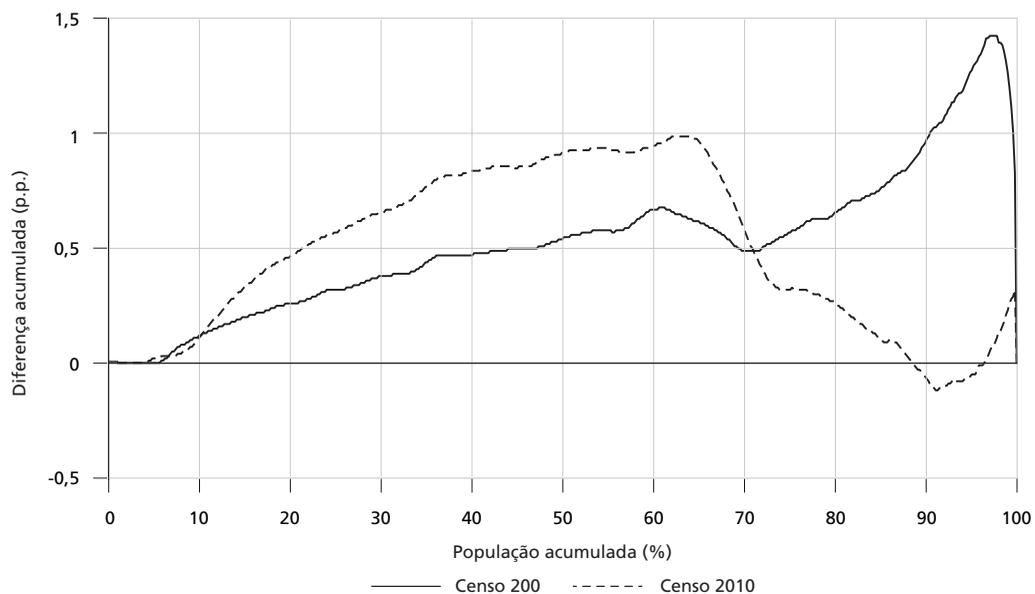
14. As PNADs mantêm fixa a lista de municípios pesquisados no período intercensitário. Entre 2001 e 2009, foram 851 municípios; em 2011, o número cresceu para 1.100. Em comparação, a Pesquisa de Orçamento Familiar (POF) 2002-2003 visitou 1.632 municípios; e a edição de 2008-2009, 1.757.

GRÁFICO 2
Diferença entre as curvas de quantis (NAR-PNAD – NAR-Outros)



Fonte: microdados dos censos 2000 e 2010.

GRÁFICO 3
Diferença entre as curvas de Lorenz (NAR-PNAD e NAR-Outros)



Fonte: microdados dos censos 2000 e 2010.

Não há o que fazer para harmonizar os planos amostrais das PNADs e os das POFs, mas, dado o tamanho das amostras dos censos, é possível aproximar, substancialmente, estas dos planos amostrais das PNADs. O terceiro e último procedimento de harmonização faz precisamente isto: todos os casos dos municípios não autorrepresentativos não selecionados nas PNADs de 2001 a 2009 foram descartados e, em seguida, o peso amostral dos indivíduos em municípios NAR-PNAD foi recalibrado para que a população de cada Unidade da Federação continuasse idêntica à dos dados originais de cada censo.

3.1.4 Resumo do filtro amostral

Para facilitar, o quadro 1 lista os procedimentos aplicados a cada uma das três fontes de dados. No restante do texto, o conjunto dos três procedimentos será sempre tratado como o “filtro amostral”. As três etapas conseguem harmonizar relativamente bem as diferenças entre censos, PNADs e POFs, com um grau de intervenção mínimo nas últimas duas pesquisas. Apenas os censos são afetados em maior escala, tanto pela exclusão dos indivíduos com rendas extremas quanto pela exclusão dos municípios não autorrepresentativos não incluídos nas PNADs de 2001 a 2009.

QUADRO 1
Etapas do filtro amostral

Etapas	Censos	PNADs	POFs
Exclusive Norte rural (exceto Tocantins)	Sim	2004-2011	Sim
Excl. 0,1% mais rico	Sim	Não	Não
Excl. municípios NAR-Outros, reponderação dos NAR-PNAD	Sim	Não	Não

Elaboração do autor.

O lado negativo da harmonização é que ela é artificial, embora não arbitrária: naturalmente, se o desejo é conhecer a distribuição de renda no Brasil como um todo, não haveria por que excluir o Norte rural, o 0,1% mais rico ou uma parte dos municípios não autorrepresentativos. Logo, o filtro amostral serve para compatibilizar as três pesquisas, mas pode enviesar as estimativas de nível e trajetória das estatísticas de interesse, e, portanto, a interpretação dos resultados deve levar isto em conta.

3.2 Diferenças conceituais de rendimentos

As diferenças nos conceitos de renda definidos pelos censos, pelas PNADs e pelas POFs tendem a ser mais relevantes que as diferenças amostrais e também mais difíceis de harmonizar. Há quatro aspectos básicos a considerar: a definição geral do conceito de renda, o período de referência para captação dos rendimentos, a definição específica dos rendimentos que se deseja captar e o grau de agregação das perguntas sobre o assunto.

3.2.1 Definição geral da renda

A definição geral da renda decorre de duas decisões: se os rendimentos são brutos ou líquidos e se eles são abrangentes ou devem se restringir aos rendimentos monetários. O padrão internacional normalmente recomenda o uso da renda disponível – isto é, os rendimentos monetários e não monetários líquidos de tributos e contribuições – por ser uma aproximação mais fidedigna da capacidade real de consumo das famílias (Canberra Group, 2001). Quase toda a literatura brasileira, no entanto, se restringe à análise da renda monetária bruta, porque nem os censos nem as PNADs trazem informações sobre rendimentos não monetários, tributos e contribuições¹⁵.

A questão passa a ser, portanto, como definir renda nas POFs. A opção foi apresentar somente os dados relativos aos rendimentos brutos – mesmo para as tabulações dos dados originais das POFs – e considerar como primeira etapa do “filtro conceitual” a exclusão dos rendimentos não monetários. Em outras palavras, todos os rendimentos das POFs apresentados abaixo são brutos, isto é, sem subtrair tributos e contribuições; no entanto, os dados das POFs “originais” incluem também os rendimentos não monetários – formados pelo consumo não monetário e, principalmente, pelo aluguel imputado líquido –, enquanto os dados das POFs com o “filtro conceitual” se restringem, entre outras coisas, aos rendimentos monetários.

3.2.2 Períodos de referência

Nos censos e nas PNADs, a referência é limitada ao mês específico de realização da pesquisa (julho e setembro, respectivamente), enquanto as POFs pretendem registrar todos os rendimentos recebidos nos doze meses anteriores à entrevista. Se a renda das

15. Alguns autores, como Costa (1988) e Rocha (2003), suspeitam que, na prática, muitos entrevistados tendem a confundir os dois tipos de rendimentos – brutos e líquidos – na hora de responder aos questionários dos censos e das PNADs, apesar de os manuais de campo afirmarem explicitamente que se trata da renda bruta – por exemplo, o *Manual de entrevista da Pesquisa Nacional por Amostra de Domicílios 2009* (IBGE, 2009, p. 258).

famílias fosse sempre constante, isto não seria um problema, mas há sérios motivos para duvidar deste pressuposto (Soares, 2009).

Apenas esse fato já geraria a possibilidade de discrepâncias entre pesquisas geradas pela volatilidade e/ou sazonalidade dos rendimentos das famílias, mas há ainda outro problema: tanto os censos quanto as PNADs dão espaço para considerável dose de subjetividade ao perguntarem qual o rendimento “habitualmente” recebido em seus meses de referência. Nenhuma das duas pesquisas pergunta de fato quanto o indivíduo recebeu em um dado mês, mas sim qual o valor “normal” que ele deveria receber. Embora a intenção seja justamente evitar que os dados fiquem contaminados por excepcionalidades, isto gera um sem-número de dificuldades adicionais.¹⁶ Não há como fazer qualquer harmonização destas diferenças com os dados existentes, mas uma parte delas pode ser atenuada com os procedimentos adotados para o terceiro aspecto da harmonização conceitual, relativo à definição dos componentes específicos da renda.

3.2.3 Definição específica dos rendimentos captados

Grosso modo, os censos e as PNADs captam um conjunto semelhante de rendimentos, ainda que em níveis de agregação diferentes, enquanto as POFs são muito mais detalhadas. Toda a harmonização consiste, então, em descartar os rendimentos incompatíveis com as definições de renda vigentes nos censos e nas PNADs. O quadro 2 traz os principais exemplos de rendimentos não captados nas outras duas pesquisas e, portanto, excluídos das versões das POFs com definições harmonizadas.

QUADRO 2

Principais rendimentos captados pelas POFs e não captados pelos Censos e pelas PNADs

Categoria	Rendimentos não captados pelos censos e PNADs
Trabalho	13º salário; abono de férias; auxílios e benefícios (alimentação, transporte, entre outros); aviso prévio; ganho com viagem de trabalho; hora extra; participação nos lucros; rendimento e/ou saque do Fundo de Garantia do Tempo de Serviço (FGTS) e do Programa de Integração Social e do Programa de Formação do Patrimônio do Servidor Público (PIS/PASEP).
Previdência pública	13º salário da aposentadoria ou pensão; restituição ou devolução da Previdência pública.
Outros rendimentos	Aluguel imputado líquido; auxílio-funeral; comissão de vendas esporádicas; consumo não monetário; dinheiro achado; empréstimos contraídos; heranças; indenizações judiciais; prêmios; resgate de título de capitalização; restituições ou devoluções de impostos e afins; vendas de bens móveis ou imóveis.

16. Para mais informações, ver a discussão em Médici (1988) e Rocha (2003).

A tabela 2 mostra o impacto da harmonização sobre a renda *per capita* total e sobre cada tipo de rendimento nas POFs. Na edição de 2002, a renda média cai 18% com a exclusão das fontes de rendimentos não captadas nos outros bancos de dados; em 2008, a queda é de 24%. Duas fontes são as principais responsáveis por isto: os rendimentos dos empregados formais e os outros rendimentos. Os motivos são simples: por um lado, os trabalhadores com carteira têm direito ao 13º salário, às férias e aos benefícios que são excluídos dos dados harmonizados; por outro, os “outros rendimentos” são afetados principalmente pela exclusão da renda não monetária.

Ressalte-se que não foi possível desagregar o 13º salário dos aposentados e pensionistas na POF em 2002. De qualquer modo, se os dados de 2008 forem tomados como parâmetro, o impacto desta exclusão seria modesto.

TABELA 2
Rendimentos *per capita* e composição da renda nas POFs antes e depois da harmonização conceitual

Fonte	POF 2002			POF 2008		
	Original (R\$ 2011)	Harmonizado (R\$ 2011)	Δ (%)	Original (R\$ 2011)	Harmonizado (R\$ 2011)	Δ (%)
Renda <i>per capita</i>	709	581	-18	941	712	-24
Trabalho	514	453	-12	591	528	-11
Com carteira	275	230	-17	333	285	-14
Sem carteira	73	66	-9	80	73	-9
Conta própria	106	99	-6	125	120	-4
Empregador	60	58	-4	53	50	-7
Demais rendas	195	128	-34	349	184	-47
Previdência pública	89	89	0	145	136	-6
Aluguéis	15	15	0	16	16	0
Doações e afins	19	19	0	21	21	-3
Programas sociais	2	2	0	7	7	0
Outros rendimentos	71	4	-94	159	3	-98

Fonte: microdados das POFs 2002-2003 e 2008-2009.

3.2.4 Grau de agregação dos rendimentos

Compatibilizar o grau de agregação dos rendimentos em cada banco de dados não chega a constituir uma etapa do processo de harmonização propriamente dito, uma vez que as mudanças são apenas nas classificações, e não afetam nem o nível nem a distribuição dos rendimentos. Apesar disso, este ponto merece ser mencionado, porque a variação entre as três fontes de dados é muito grande e há algumas evidências de

que o grau de agregação na coleta dos dados pode afetar o nível e a distribuição dos rendimentos declarados (Davern *et al.*, 2005). Por sinal, mesmo quando a coleta é bastante desagregada, permanece o fato de que pesquisas domiciliares são, em geral, particularmente deficientes na captação de rendimentos de trabalhadores por conta própria e empregadores e rendas de investimentos e propriedades.

O Censo 2000 e as PNADs têm um grau de agregação relativamente próximo, situando-se entre a extrema desagregação das POFs e a coleta sintética do Censo 2010, embora obviamente sejam mais semelhantes a este último. O quadro 3 apresenta os rendimentos explicitamente desagregados, ou facilmente desagregáveis, nos dados originais de cada pesquisa.

QUADRO 3

Desagregação das fontes de rendimentos nos censos, nas PNADs e nas POFs

Desagregação	Censo 2000	Censo 2010	PNAD (1995-2011)	POF (2002 e 2008)
Renda <i>per capita</i>	Sim	Sim	Sim	Sim
Trabalho	Sim	Sim	Sim	Sim
Com carteira	Sim	Sim	Sim	Sim
Sem carteira	Sim	Sim	Sim	Sim
Conta própria	Sim	Sim	Sim	Sim
Empregador	Sim	Sim	Sim	Sim
Demais rendas	Sim	Sim	Sim	Sim
Previdência pública	Sim	Não, só indicação	Sim	Sim
Aluguéis	Sim	Não, residual	Sim	Sim
Doações e afins	Sim	Não, residual	Sim	Sim
Programas sociais	Sim	Não, só indicação	Sim, estimado	Sim
Outras	Sim	Não, residual	Sim, estimado	Sim

Fonte: microdados das PNADs 1999, 2001, 2002, 2008, 2009 e 2011; dos censos 2000 e 2010; das POFs 2002-2003 e 2008-2009.

Dois comentários merecem ser feitos. Primeiro, o Censo 2010 não representou propriamente um total retrocesso à forma de captação do Censo 1970 e afins porque, apesar de as “outras rendas” serem coletadas em um único item, este foi precedido por três perguntas indicadoras da participação, ou não, nos programas sociais mais importantes. Como a Previdência Social, em particular, representa fração expressiva da renda das famílias, é possível que estas perguntas tenham atenuado o viés de subestimação da renda ao lembrar os entrevistados de incluir estas rendas em suas respostas. Não há

como testar detalhadamente esta possibilidade, mas ela parece razoável se considerado que os resultados da próxima seção mostram grande congruência entre os dados harmonizados dos dois censos.

O segundo comentário diz respeito à identificação do Programa Bolsa Família e do Benefício de Prestação Continuada (BPC) nas PNADs. Como não é possível identificar os beneficiários e as transferências recebidas diretamente, o método dos “valores típicos” foi utilizado, tal como descrito em Souza, Osorio e Soares (2011).¹⁷

3.2.5 Resumo do filtro conceitual

Analogamente à seção anterior, o quadro 4 resume as etapas do “filtro conceitual”. Na prática, somente as POFs são afetadas por estes procedimentos, pois tanto os censos quanto as PNADs capturam fontes de rendimentos semelhantes.

QUADRO 4
Etapas do filtro conceitual

Etapas	Censos	PNADs	POFs
Definição geral da renda bruta	Sem alteração	Sem alteração	Só renda monetária
Período de referência	Sem alteração	Sem alteração	Sem alteração
Definição específica dos componentes da renda	Sem alteração	Sem alteração	Exclusive rendimentos não captados na PNAD
Agregação da renda	Só reclassificação, sem alterar nível ou distribuição da renda		

Ao contrário das diferenças amostrais, a harmonização das definições conceituais é bastante imperfeita, limitando-se à exclusão, nas POFs, dos rendimentos obviamente não pesquisados pelas PNADs e pelos censos. Contudo, assim como o filtro amostral, o filtro conceitual também representa uma manipulação artificial, mas não arbitrária, das informações. A comparação apropriada entre censos, PNADs e POFs impõe a necessidade deste filtro, mas obviamente não há motivos intrínsecos para descartar determinadas fontes de rendimentos da análise, o que também deve ser levado em conta na interpretação dos resultados.

17. Para mais informações, ver também Soares *et al.* (2007); Barros (2007); Barros, Carvalho e Franco (2007).

3.3 Diferenças na coleta e tratamento dos dados divulgados

Se as diferenças conceituais são mais graves e menos tratáveis que as amostrais, as diferenças na coleta e no tratamento dos dados são ainda piores. O caso da coleta, em particular, é mais sério e irreversível, afetando principalmente os censos: ao contrário das PNADs e das POFs, o próprio escopo dos censos faz que eles só possam ser viabilizados pelo recrutamento de um exército de recenseadores temporários e inexperientes, cujo treinamento, por sua vez, é inevitavelmente menos minucioso que em outras pesquisas.

Para além desse tipo de problema, há muitas outras questões mais sutis relacionadas às idiossincrasias de cada pesquisa. Variações na duração da entrevista, ordem das perguntas, conhecimento dos entrevistados sobre as pesquisas, forma de coleta, entre outras, têm o potencial para afetar os resultados. Naturalmente, boa parte do planejamento e da preparação, anteriores à coleta dos dados, busca minimizar estas influências. Uma vez realizado o trabalho de campo, no entanto, cabe à crítica dos dados detectar e corrigir, na medida do possível, eventuais problemas.

No caso dos rendimentos, a questão é o tratamento das informações, o qual não é idêntico nas três pesquisas. *Grosso modo*, nos censos e nas POFs, o processo primeiro detecta indivíduos que declararam rendimentos positivos improváveis para o seu perfil e, em seguida, substitui a renda declarada destes indivíduos por uma renda imputada. Os indivíduos que não souberam ou se recusaram a fornecer resposta para alguma das perguntas sobre rendimentos – ou seja, os indivíduos sem declaração ou com renda ignorada – também passam por esta segunda etapa.

Nos dois casos – indivíduos com rendimentos improváveis ou com rendas ignoradas –, a imputação é feita via *hot deck*: todos os indivíduos da amostra são primeiro estratificados de acordo com um conjunto pré-selecionado de variáveis socioeconômicas e, depois, cada caso improvável ou ignorado recebe, na variável relevante, o valor declarado por um “doador” escolhido aleatoriamente dentro do seu estrato (IBGE, 2010; 2012a).

Uma peculiaridade dos censos é que esse procedimento não é realizado para os indivíduos que declararam ter renda igual a zero. Em decorrência disto, alguns problemas já haviam sido notados quanto aos dados de 2000. Hoffman e Ney (2008), por exemplo, optaram por descartar todos os domicílios com renda zero de sua análise:

O Censo Demográfico de 2000 apresenta dados estranhos na cauda inferior da distribuição da renda. Há uma proporção de domicílios com renda total declarada nula [isto é, zero] substancialmente maior do que na PNAD, o que só não representaria grave problema se essa tendência estivesse restrita às famílias com todos os membros desocupados ou com algum deles ocupado em atividade mal remunerada. Em boa parte dos casos, porém, vive no domicílio pelo menos um membro empregado como militar do exército, delegado, juiz, entre outras profissões cuja renda é relativamente alta (Hoffmann e Ney, 2008).

Em 2010, o problema se agravou em função de questões técnicas:

No Censo 2000, os dados foram coletados em papel. No Censo 2010, a coleta foi realizada com um Computador de Mão (PDA) que continha regras mais restritivas de preenchimento, e dificultava a possibilidade de o Recenseador registrar que a pessoa possuía rendimento e não declarar o valor. (...) Quando se realiza a soma dos domicílios sem rendimento com os domicílios “sem declaração” na PNAD, chega-se a uma estimativa (...) muito próxim[a] do total de domicílios com rendimento zero encontrados no Censo Demográfico 2010. Acredita-se que este fato é um forte indício a contribuir para o raciocínio de que a opção “não tem rendimento” serviu no Censo, em um número significativo de situações, como alternativa para a categoria “ignorado” (IBGE, 2012c).

O IBGE chegou a desenvolver um modelo para imputar rendimentos para parte das famílias com renda zero, mas acabou optando por abandoná-lo, pois sua adoção exigiria, em muitos casos, a imputação de grande parte das variáveis de mercado de trabalho para garantir a consistência dos microdados (IBGE, 2012c). Com isso, tanto o Censo de 2000 quanto o Censo de 2010 contam com uma porcentagem muito elevada de indivíduos com renda domiciliar *per capita* igual a zero. Nas POFs isto não ocorre porque a coleta de dados é feita de modo muito mais detalhado e, principalmente, o período de referência de doze meses faz que a incidência de domicílios com renda zero seja irrisória.

Nas PNADs, o tratamento dos rendimentos é completamente diferente. Embora o IBGE também faça a crítica dos dados antes de sua divulgação, não há nas PNADs qualquer tipo de imputação para os indivíduos que não quiseram ou não souberam responder a alguma das perguntas sobre seus rendimentos pessoais. Nestes casos, os microdados trazem o código para rendimentos ignorados.

O procedimento padrão da maioria dos pesquisadores é simplesmente excluir da análise todos os moradores de domicílios em que pelo menos um membro possui pelo menos um rendimento ignorado. Todos os resultados relativos aos “dados originais” das

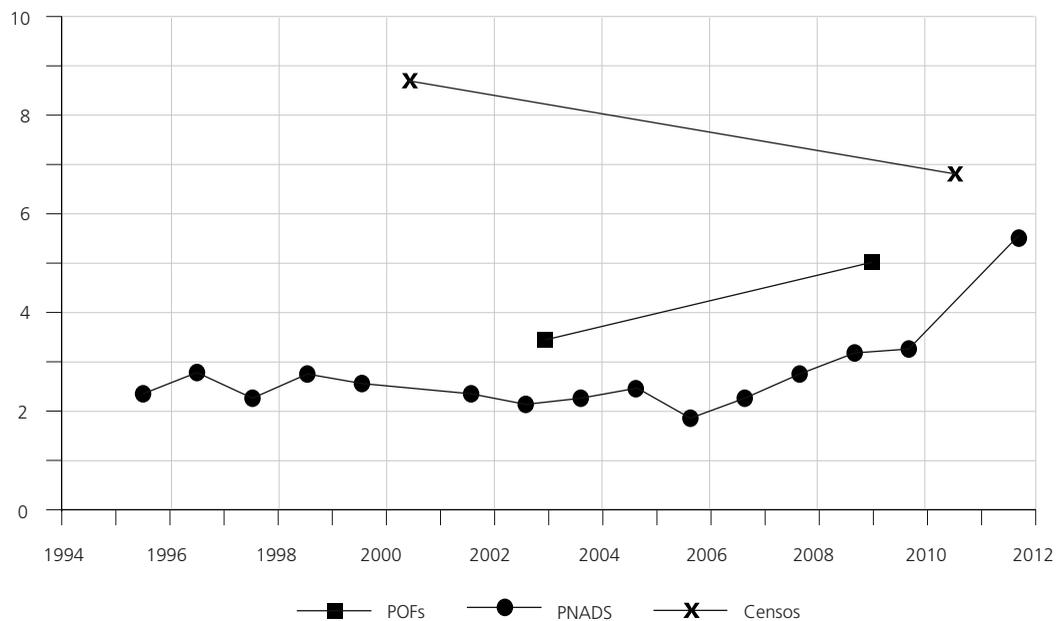
PNADs apresentados mais abaixo fazem precisamente isto. No entanto, para a comparação mais perfeita entre as três fontes de dados, isto é duplamente problemático: primeiro, porque destoa do tratamento aplicado aos censos e às POFs; segundo, porque há, desde 2005, uma nítida tendência de aumento dos casos ignorados, que deixaram de ser residuais e chegaram em 2011 a mais de 5% da população.

Além disso, há também a questão dos indivíduos com renda zero. Apesar de serem menos numerosos que nos censos, sua estabilidade é, no mínimo, suspeita. Entre 1995 e 2011, houve crescimento econômico, queda da desigualdade e redução da extrema pobreza, mas a porcentagem de indivíduos com renda zero permaneceu basicamente a mesma. Isto se torna ainda mais estranho quando se leva em conta que as PNADs perguntam explicitamente qual a renda “habitualmente” recebida em setembro, pois é plausível pensar que apenas um grupo restrito de indivíduos – por exemplo, pequenos agricultores – tenham rendimentos tão sazonais a ponto de ser “normal” ter renda zero naquele mês. Para os demais, o mais provável é que se trate de um erro de declaração ou coleta. Com efeito, Osório, Soares e Souza (2011) já argumentaram neste sentido, mostrando que boa parte dos indivíduos com renda zero nas PNADs tem um perfil tão discrepante que apenas é possível explicar sua situação em termos de choques adversos temporários ou de algum tipo de erro.

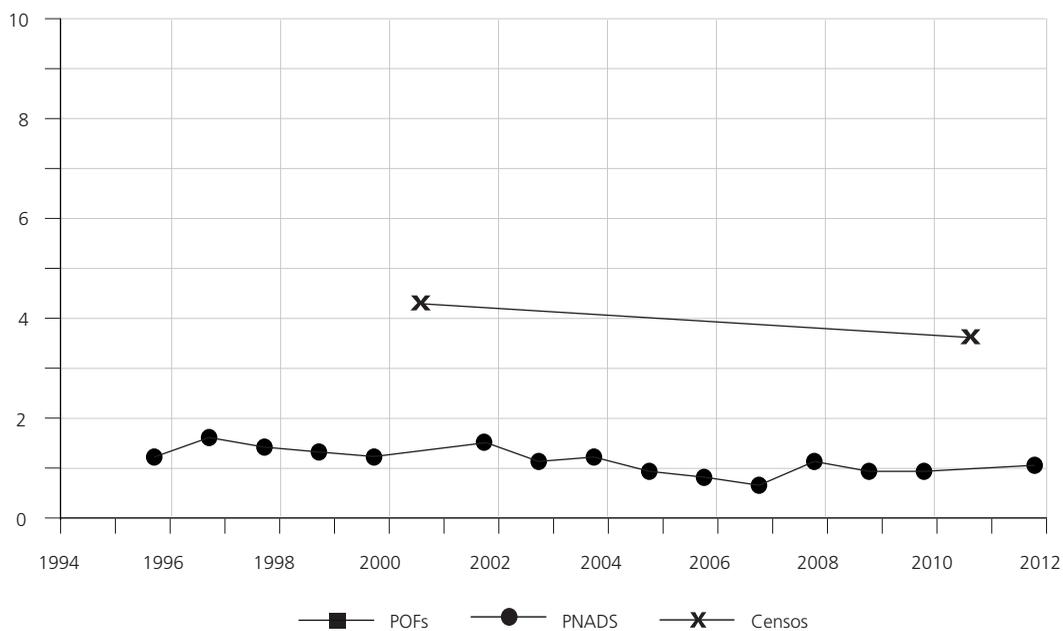
Para harmonizar o tratamento das informações sobre renda, é preciso lidar com estas duas questões. Trata-se, primeiro, de tentar encontrar, entre a massa de indivíduos com renda zero nos censos e nas PNADs, quais não têm perfil de pobres e imputar-lhes rendimentos. Além disso, as PNADs também demandam a imputação de rendimentos para os indivíduos que não quiseram ou não souberam declarar quanto recebiam de alguma fonte específica.

O gráfico 4 ilustra o problema. O gráfico 4A mostra que as taxas de indivíduos com rendimentos ignorados são menores nas PNADs que nas outras duas pesquisas, mas vêm crescendo continuamente desde meados da década de 2000. Como não parece muito razoável descartar mais de 5% dos casos, a imputação de tais rendimentos – como ocorre nos censos e nas POFs – se apresenta como a solução mais recomendável. A partir do gráfico 4B, é possível perceber como o problema dos indivíduos em domicílios com renda zero afeta sobretudo os censos, mas nem mesmo as questões operacionais já mencionadas impedem que haja uma leve queda nas porcentagens entre 2000 e 2010. Nas PNADs, há uma grande estabilidade.

GRÁFICO 4
População em domicílios com rendimentos ignorados ou imputados e em domicílios com renda igual a zero – censos, PNADs e POFs
 4A – População em domicílios com rendimentos ignorados ou imputados
 (Em %)



4B – População em domicílios com renda igual a zero
 (Em %)



Fonte: microdados das PNADs 1999, 2001, 2002, 2008, 2009 e 2011; dos censos 2000 e 2010; das POFs 2002-2003 e 2008-2009.

3.3.1 Imputação de rendimentos para indivíduos com rendimentos ignorados nas PNADs

A imputação de valores válidos de renda para os indivíduos com rendimentos ignorados seguiu um procedimento do tipo *hot deck* semelhante ao aplicado aos censos e às POFs, ainda que com outras variáveis e outro *software*. Para os rendimentos do trabalho – responsáveis sempre por cerca de 80% dos casos de não declaração – e abono de permanência, o *hot deck* foi estratificado por Unidade da Federação, área censitária – região metropolitana, município autorrepresentativo e município não autorrepresentativo; educação – sem escolaridade, ensino fundamental completo, médio completo e superior completo; e posição na ocupação – empregados formais, empregados informais, trabalhadores por conta própria e empregadores. Para as demais rendas, a estratificação foi feita por Unidade da Federação, área censitária, gênero e educação.¹⁸

A imputação foi precedida por uma análise do perfil dos domicílios com renda ignorada para verificar se houve mudanças bruscas ao longo do tempo. Por motivos de espaço, os resultados não serão exibidos ou discutidos em detalhe no texto, mas, de modo geral, pode-se afirmar que o aumento recente da não declaração de renda foi concomitante às transformações significativas: antes, indivíduos com alta escolaridade e/ou empregadores e trabalhadores por conta própria tinham probabilidade muito maior de não saber ou não querer declarar seus rendimentos, mas nos últimos anos houve um estreitamento nítido dos diferenciais. Ou seja, a não declaração aumentou, mas se tornou menos enviesada que antes.

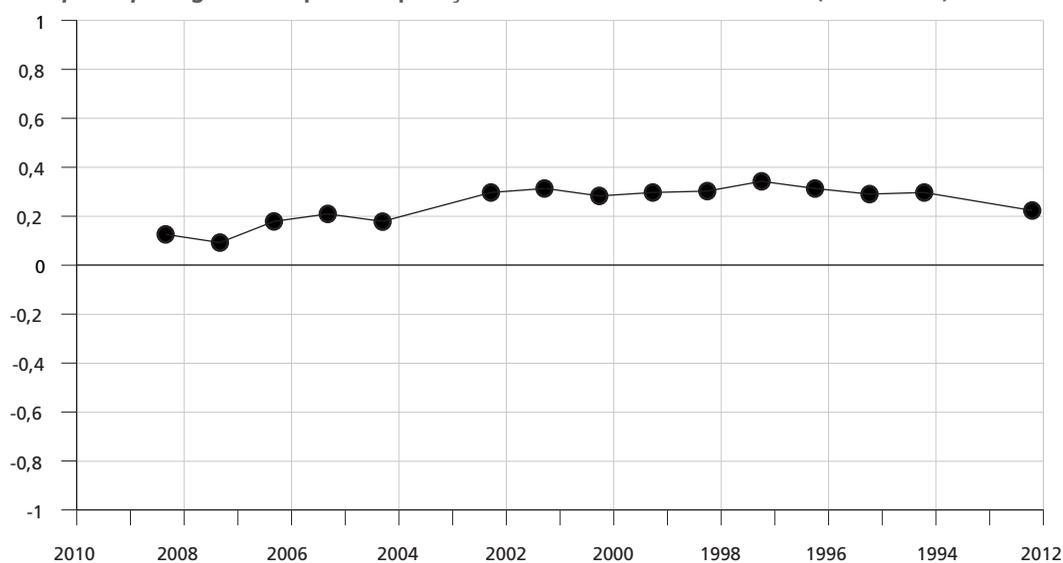
O gráfico 5 traz os coeficientes de concentração ajustados dos indivíduos em domicílios com rendimentos ignorados – isto é, em que pelo menos um membro não respondeu pelo menos a uma questão sobre renda – após os procedimentos de imputação. Estes coeficientes podem variar entre -1 (se todos estes indivíduos fossem os mais pobres da distribuição após a imputação) e +1 (se todos eles fossem os mais ricos); coeficientes iguais a zero implicam que os indivíduos se dispõem de forma aleatória ao longo da distribuição de renda após a imputação.¹⁹

18. Para evitar a ocorrência de células vazias ou com um número muito pequeno de casos, somente as variáveis com maior capacidade de discriminação foram incluídas nos modelos. Dada a preponderância da renda do trabalho e a disponibilidade de informações exclusivas desta fonte – por exemplo, posição na ocupação –, foi elaborado um modelo específico para este tipo de rendimento. Para testar a sensibilidade, algumas especificações alternativas foram experimentadas, sem mudanças substantivas nos resultados.

19. Para mais informações, ver Osorio (2009).

Em todas as PNADs, os coeficientes de concentração ajustados são positivos, o que indica que os indivíduos com renda domiciliar *per capita* ignorada tendem a estar entre os mais ricos, mas não muito elevados, e com tendência consistente de queda nos últimos anos. Logo, é possível que o aumento da porcentagem de ignorados não tenha maiores consequências para a comparabilidade da série histórica das PNADs, mesmo quando simplesmente se opta por descartar estes casos. Independentemente desse fato, o procedimento de imputação garante maior comparabilidade tanto com os censos quanto com as POFs.

GRÁFICO 5
Coeficientes de concentração ajustados dos indivíduos com renda domiciliar *per capita* ignorada após a imputação dos rendimentos nas PNADs (1995-2011)



Fonte: microdados das PNADs 1995-2011.

3.3.2 Imputação de rendimentos para indivíduos com renda *per capita* igual a zero e sem perfil de pobreza ou extrema pobreza

O tratamento dos indivíduos em domicílios com renda zero é feito em duas etapas. Primeiro, aplica-se um modelo de análise de conglomerados para identificar os domicílios com renda igual a zero que não têm perfil socioeconômico compatível com

uma situação de pobreza ou extrema pobreza.²⁰ Em seguida, é feita a imputação de rendimentos (positivos) para os moradores destes domicílios.²¹

Quatro variáveis foram utilizadas na análise de conglomerados: *dummies* para domicílios em áreas urbanas; *dummies* para domicílios cujas pessoas de referência eram homens; *dummies* para domicílios com algum membro com pelo menos o ensino médio completo; e uma variável contínua relativa ao índice socioeconômico do domicílio. Este índice foi construído a partir da análise de componentes principais de uma série de variáveis relacionadas às características do domicílio, como a presença de eletrodomésticos, saneamento básico, entre outras, e padronizado para que sua média geral fosse igual a zero em todos os anos.²²

O *hot deck* para imputar rendimentos para esses domicílios é estratificado por Unidades da Federação, área censitária – regiões metropolitanas, municípios autorrepresentativos e municípios não autorrepresentativos; maior nível educacional entre os moradores – sem escolaridade, ensino fundamental completo, médio completo e superior completo; e número de moradores por cômodo – truncado para uma casa decimal. Para garantir a consistência dos microdados, os rendimentos imputados para estes domicílios com renda zero foram considerados como “outros rendimentos”.

Os gráficos 6 e 7 trazem os resultados da análise de conglomerados para os censos e para as PNADs, confirmando que estas duas fontes de dados tendem a superestimar a quantidade de pessoas com renda domiciliar *per capita* igual a zero. Desconsiderando os motivos operacionais que tornam o problema mais agudo no censo, isto se explica provavelmente pela combinação de períodos de referência curtos e erros de coleta ou declaração: tanto nas PNADs quanto nos censos há muitas famílias que estão, no máximo, apenas temporariamente com renda zero.

20. Para mais informações, ver Osorio, Soares e Souza (2011).

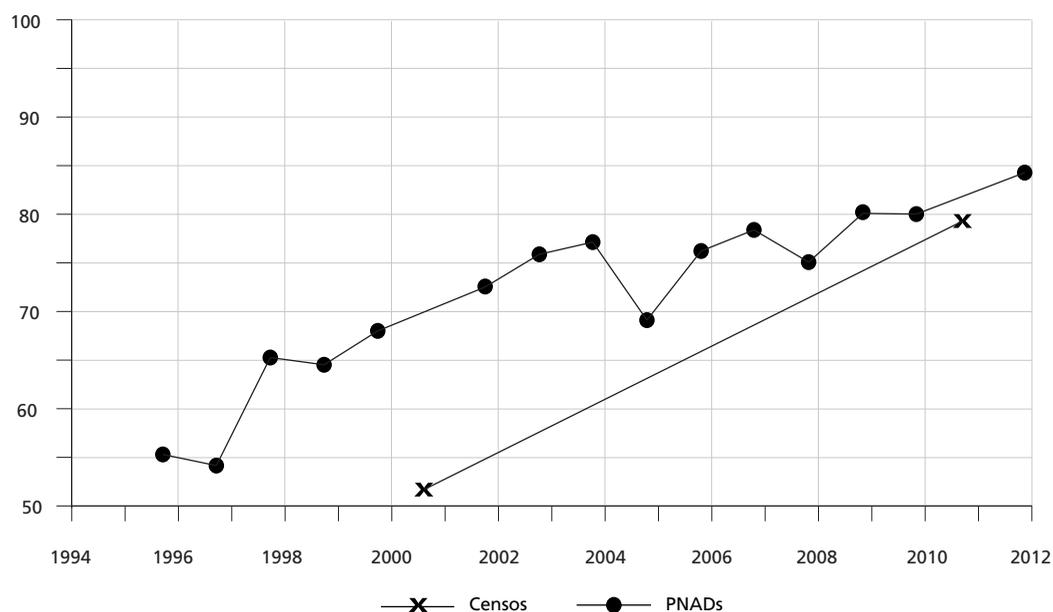
21. A rigor, a imputação da renda também para os pobres com renda zero poderia ser defendida com base na ideia de que não há famílias cuja renda “normalmente” – como perguntam os censos e as PNADs – recebida em determinado mês seja igual a zero. Este procedimento não foi adotado neste trabalho em nome da parcimônia: como tal imputação teria muito pouco efeito prático sobre os números finais, em razão do pequeno número de famílias nesta situação e do baixo nível dos rendimentos esperados, o autor optou por manipular o mínimo possível os dados originais.

22. Para mais detalhes, ver apêndice A desta obra.

O gráfico 6 mostra que, nos dois casos, o percentual de “não pobres” entre as pessoas com renda domiciliar *per capita* igual a zero aumentou significativamente ao longo do tempo. Em 1995, na PNAD, cerca de metade dos indivíduos desse grupo não tinham perfil de pobres; em 2011, já eram 85%. Nos censos, o crescimento foi muito semelhante: de pouco mais de 50% em 2000 para 80% em 2010.

O gráfico 7 confirma que os dois grupos – extremamente pobres e não pobres – identificados pela análise de conglomerados entre os indivíduos com renda domiciliar *per capita* igual a zero são de fato muito distintos. Como o índice socioeconômico é construído para ter média zero em cada ano para cada pesquisa, pode-se perceber que os não pobres com renda zero estão um pouco acima da média no Brasil, enquanto os extremamente pobres têm números muito piores.

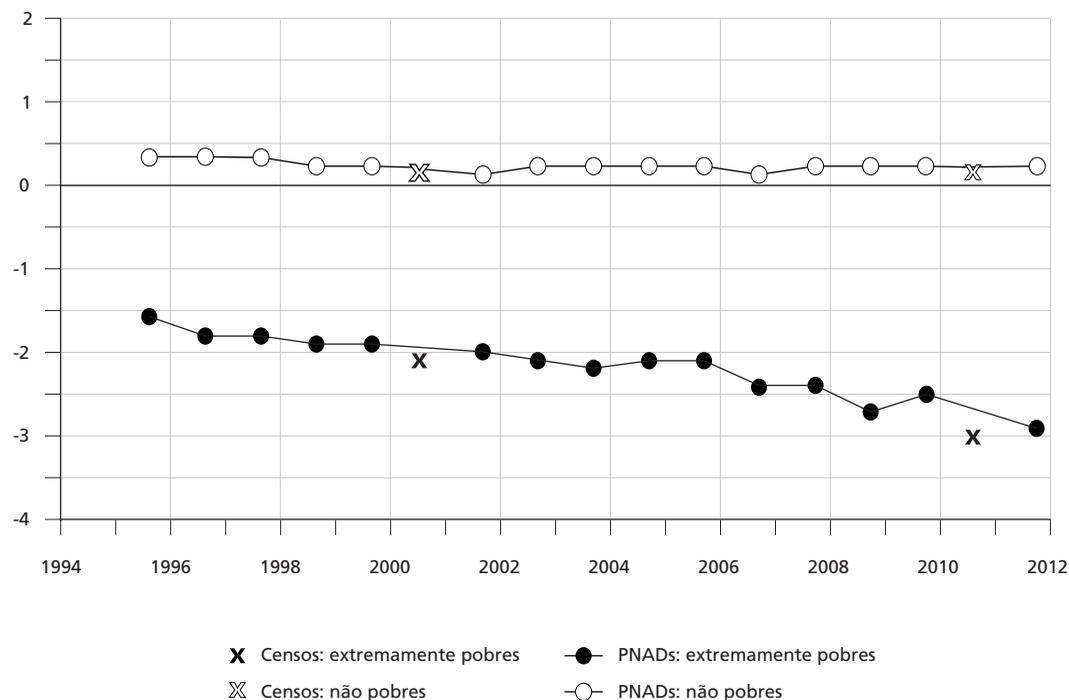
GRÁFICO 6
Resultados da análise de conglomerados para indivíduos com renda *per capita* igual a zero: porcentagem de indivíduos não pobres
(Em %)



Fonte: microdados das PNADs 1995-2011 e dos censos 2000 e 2010.

GRÁFICO 7

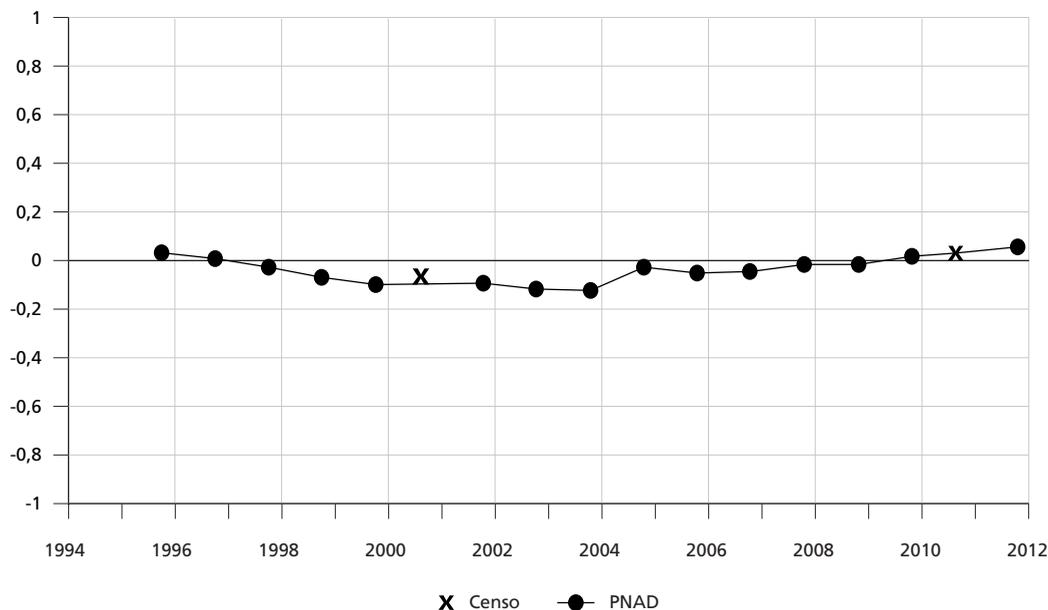
Resultados da análise de conglomerados para indivíduos com renda *per capita* igual a zero: índice socioeconômico de pobres extremos e não pobres com renda zero (Em %)



Fonte: microdados das PNADs 1995-2011 e dos censos 2000 e 2010.

O gráfico 8 apresenta os coeficientes de concentração ajustados após a imputação da renda para os indivíduos não pobres com renda domiciliar *per capita* originalmente igual a zero nos censos e nas PNADs. Mais uma vez, as duas fontes de dados geram resultados semelhantes, com coeficientes de concentração bem próximos de zero, o que sinaliza uma distribuição mais ou menos aleatória.

GRÁFICO 8
Coefficientes de concentração ajustados dos indivíduos não pobres com renda zero após a imputação dos rendimentos nos censos e nas PNADs



Fonte: microdados das PNADs 1995-2011 e dos censos 2000 e 2010.

3.3.3 Resumo do filtro de tratamento

O quadro 5 resume os procedimentos de harmonização do tratamento dos dados – filtro de tratamento. A partir destes dados, depreende-se que as POFs não foram afetadas pelos ajustes mencionados acima, uma vez que o próprio IBGE faz a crítica e a imputação dos rendimentos ignorados e há apenas um número ínfimo de domicílios com renda igual a zero em ambas as edições com cobertura nacional.

QUADRO 5
Etapas do filtro de tratamento dos rendimentos

Etapas	Censos	PNADs	POFs
Imputação dos rendimentos ignorados	Não	Sim	Não
Identificação dos indivíduos “não pobres” com renda igual a zero	Sim	Sim	Não
Imputação de rendimentos para “não pobres” com renda zero	Sim	Sim	Não

Ao contrário dos filtros anteriores, o filtro de tratamento não representa uma manipulação artificial dos dados. Com efeito, pode-se argumentar que ele aproxima tanto os censos quanto as PNADs dos parâmetros populacionais “reais”.

3.4 Seleção dos anos para comparação

Considerando-se que não faz sentido comparar pesquisas realizadas em anos muito diferentes, é preciso escolher quais edições da PNAD serão pareadas com os dois censos e com as duas POFs.

Levando-se em conta também o fato de que a PNAD não vai a campo em anos censitários, deve-se, portanto, escolher ou a PNAD imediatamente anterior ou a imediatamente posterior a cada censo. Esta decisão não faz muita diferença para os resultados, mas, para evitar uma escolha puramente arbitrária entre 1999 ou 2001 e 2009 ou 2011, optou-se por criar PNADs “sintéticas” para os anos de 2000 e 2010. O nível e a distribuição dos rendimentos em cada caso representam a média entre a PNAD do ano anterior e a do ano posterior.

Há diversas formas mais ou menos sofisticadas de fazer isso; neste trabalho, o autor optou por agregar as PNADs 1999, 2001, 2009 e 2011 em 12 mil grupos de tamanho igual, ordenados por renda, e construir, a partir deles, as versões “sintéticas” de 2000 e 2010. A quantidade de grupos foi escolhida visando reproduzir com precisão a distribuição de renda de cada ano. Por exemplo, todas as estatísticas de renda, desigualdade e pobreza geram resultados idênticos até várias casas decimais, não importando se são aplicadas aos cerca de 100 mil domicílios de cada PNAD, ou se geradas a partir da agregação da PNAD em 12 mil grupos.

A comparação com as POFs é mais simples, pois, ao contrário do que ocorreu na década de 1970, durante a realização do ENDEF, não há mais interrupção da PNAD nestas ocasiões. Basta, então, escolher uma edição de fato existente da pesquisa. Tendo em vista que a data de referência da POF 2002 é janeiro de 2003, e a da POF 2008 é janeiro de 2009, a opção mais razoável é compará-las com as PNADs mais próximas às suas datas de referência, ou seja, 2002 e 2008.

3.5 Resumo dos filtros

A aplicação dos procedimentos descritos nos três primeiros tópicos desta seção garante a maior comparabilidade possível entre as três fontes de dados, ainda que a harmonização obviamente não seja perfeita. Em decorrência disso, o maior interesse da análise abaixo está em cotejar as versões originais e plenamente harmonizadas de cada base e em observar se há alguma convergência dos dados após estes procedimentos.

Contudo, também interessa saber qual dos procedimentos tem o maior impacto na promoção dessa convergência (ou divergência), principalmente porque os filtros apresentam características distintas. No caso dos filtros amostral e conceitual, há um possível *tradeoff* entre comparabilidade e viés com relação aos parâmetros populacionais, o que não ocorre no caso do filtro de tratamento. Por isso, a análise não se restringe aos dados originais – isto é, seguindo os passos descritos no início da seção e os procedimentos normalmente adotados pelos pesquisadores que usam cada banco de dados – e aos dados plenamente harmonizados: todos os exercícios foram replicados também para dados apenas parcialmente harmonizados.

Para cada comparação, foram criados, em cada ano, bancos de dados que representam todas as combinações possíveis dos filtros, como mostra o quadro 6. Com uma única exceção, as comparações sempre relacionam bancos de dados que receberam o mesmo tratamento: por exemplo, as PNADs com filtro amostral são cotejadas com os censos com o filtro amostral e assim por diante. Este princípio é violado apenas em uma das etapas de comparação entre PNADs e POFs, porque não é necessário aplicar o filtro conceitual às PNADs nem o filtro de tratamento às POFs. Nestes casos, as PNADs conceito e as POFs tratamento são idênticas às versões originais de cada pesquisa.

QUADRO 6

Resumo dos procedimentos de harmonização aplicados a cada fonte de dados

6A – Comparação entre PNAD e censo

Nome	Banco de dados	Período	Filtro amostral	Filtro conceitual	Filtro de tratamento
PNAD original	PNAD	2000 e 2010	Não	Não	Não
PNAD amostra	PNAD	2000 e 2010	Sim	Não	Não
PNAD tratamento	PNAD	2000 e 2010	Não	Não	Sim
PNAD harmonizada	PNAD	2000 e 2010	Sim	Não	Sim
Censo original	Censo	2000 e 2010	Não	Não	Não
Censo amostra	Censo	2000 e 2010	Sim	Não	Não
Censo tratamento	Censo	2000 e 2010	Não	Não	Sim
Censo harmonizada	Censo	2000 e 2010	Sim	Não	Sim

6B – Comparação entre PNAD e POF

Nome	Banco de dados	Período	Filtro amostral	Filtro conceitual	Filtro de tratamento
PNAD original	PNAD	2002 e 2008	Não	Não	Não
PNAD amostra	PNAD	2002 e 2008	Sim	Não	Não
PNAD conceito	PNAD	2002 e 2008	Não	Não	Não
PNAD tratamento	PNAD	2002 e 2008	Não	Não	Sim
PNAD harmonizada	PNAD	2002 e 2008	Sim	Não	Sim
POF original	POF	2002 e 2008	Não	Não	Não
POF amostra	POF	2002 e 2008	Sim	Não	Não
POF conceito	POF	2002 e 2008	Não	Sim	Não
POF tratamento	POF	2002 e 2008	Não	Não	Não
POF harmonizada	POF	2002 e 2008	Sim	Sim	Não

Apesar de numerosos, os ajustes aos dados originais são pouco intrusivos. A tabela 3 lista a porcentagem da população afetada por cada tipo de filtro, isto é, aqueles casos que são descartados ou têm sua renda modificada por imputação em cada etapa. A harmonização plena nem sempre equivale à soma de todos os filtros, pois há casos que são afetados por mais de um procedimento.

TABELA 3
Frações das populações originais expandidas afetadas por cada filtro de harmonização nos censos, nas PNADs e nas POFs

Banco de dados	População afetada – casos descartados ou rendimentos imputados (%)			
	Amostra	Conceitual	Tratamento	Harmonização plena
PNAD 2000 ¹	0,0	0,0	3,3	3,3
PNAD 2010 ¹	1,9	0,0	5,1	7,0
Censo 2000	37,6	0,0	2,3	39,3
Censo 2010	37,0	0,0	2,9	39,2
Banco de dados	População afetada – casos descartados ou rendimentos imputados (%)			
	Amostra	Conceitual	Tratamento	Harmonização plena
PNAD 2002 ¹	0,0	0,0	2,9	2,9
PNAD 2008 ¹	2,1	0,0	3,8	5,8
POF 2002	1,9	0,0	0,0	1,9
POF 2008	2,1	0,0	0,0	2,1

Fonte: microdados das PNADs 1999, 2001, 2002, 2008, 2009 e 2011; dos censos 2000 e 2010; das POFs 2002-2003 e 2008-2009.

Nota: ¹ Excepcionalmente, essa tabela considera como população total nos dados originais da Pesquisa Nacional por Amostra de Domicílio (PNAD) a soma dos indivíduos com rendimentos declarados e dos indivíduos com renda domiciliar *per capita* ignorada. Nas demais análises deste trabalho, os números referentes aos dados originais da PNAD sempre excluem os casos com rendimentos ignorados.

Tanto nas PNADs quanto nas POFs só uma fração minoritária dos indivíduos da amostra é afetada pela harmonização. Nas POFs, as porcentagens ficam em torno de 2%, representando apenas os casos descartados de moradores do Norte rural, exceto Tocantins. Nas PNADs, os números são um pouco maiores, variando entre 3% e 7%, em função dos indivíduos com renda domiciliar *per capita* ignorada. O procedimento-padrão – aplicado aos “dados originais” – consiste na exclusão deste grupo; o filtro de tratamento da renda, por sua vez, imputa rendimentos válidos para eles.

Nos censos, o contingente de indivíduos descartados ou com renda imputada é muito maior, chegando a quase 40% das amostras, mas isto era esperado: uma das etapas do filtro amostral é justamente a eliminação de todos os indivíduos em municípios não autorrepresentativos não pesquisados pelas PNADs de 2001 a 2009.

4 A COMPARAÇÃO ENTRE OS CENSOS, AS PNADS E AS POFs

4.1 As divergências nos dados originais

Para dimensionar as discrepâncias entre os dados originais dos censos, das PNADs e das POFs, as tabelas 4 e 5 exibem, respectivamente, informações sobre o nível, a desigualdade e a composição dos rendimentos em cada uma das três fontes de dados.

A comparação entre os censos e as PNADs revela números gerais muito diferentes, inclusive com aumento da divergência ao longo do tempo. A renda *per capita* nos censos já era mais elevada em 2000, e cresceu de forma mais rápida que nas PNADs até 2010, enquanto a desigualdade e a extrema pobreza apresentaram a tendência inversa. A diferença é particularmente grande no caso da desigualdade, que permanece no Censo 2010 em um patamar superior ao registrado pela PNAD há uma década.

No entanto, essas diferenças não parecem resultar de questões mais óbvias de má captação de alguma fonte específica de rendimentos. Tanto a composição da renda total quanto o recebimento de cada fonte são similares nas duas pesquisas. No caso da composição, inclusive, houve convergência entre censos e PNADs ao longo do tempo.

A comparação entre PNADs e POFs também revela diferenças substantivas. Em consonância com resultados anteriores (Barros, Cury e Ulyseia, 2006), a tabela 5

mostra que o nível de renda é muito mais alto na POF, mas os níveis de desigualdade são relativamente próximos. As duas variáveis – renda média e desigualdade – também apresentam comportamentos distintos ao longo do tempo: enquanto a renda nas POFs descolou-se ainda mais das PNADs, a queda na desigualdade teve quase a mesma intensidade nas duas pesquisas.

TABELA 4
Estatísticas gerais sobre nível, distribuição e composição dos rendimentos nas PNADs e nos censos (2000 e 2010)

Estatísticas gerais	PNAD 2000 (A)	Censo 2000 (B)	Diferença (B - A) (%)	PNAD 2010 (C)	Censo 2010 (D)	Diferença (D - C) (%)
Renda <i>per capita</i> (R\$ 2011)	571	629	10	732	828	13
Gini (x100)	59,3	64,6	9	53,4	60,9	14
Extrema pobreza (%)	9,5	13,5	4,0	4,4	8,9	4,5
Composição da renda (%)	PNAD 2000 (A)	Censo 2000 (B)	Diferença (B - A)	PNAD 2010 (C)	Censo 2010 (D)	Diferença (D - C)
Trabalho	78	78	0	77	74	-2
Com carteira	39	30	-9	43	43	0
Sem carteira	11	16	5	10	8	-2
Conta própria	17	20	3	15	17	3
Empregador	12	12	0	9	6	-3
Demais rendas	22	22	0	23	26	2
Pessoas em domicílios que recebem (%)	PNAD 2000 (A)	Censo 2000 (B)	Diferença (B - A)	PNAD 2010 (C)	Censo 2010 (D)	Diferença (D - C)
Trabalho	90	86	-5	89	83	-5
Com carteira	48	43	-5	54	55	1
Sem carteira	35	37	3	31	26	-5
Conta própria	35	31	-3	31	27	-4
Empregador	6	4	-2	5	2	-3
Demais rendas	41	41	1	56	54	-2
Previdência	30	32	2	29	30	1
Programas sociais	3	2	-1	26	26	1

Fonte: microdados das PNADs 1999, 2001, 2009 e 2011; e dos censos 2000 e 2010.

O principal problema fica óbvio na comparação da composição e recebimento da renda. Dado que as POFs registram rendimentos não monetários – como o consumo não monetário e o aluguel imputado –, os quais as PNADs ignoram, os outros rendimentos têm um peso muito maior naquela pesquisa. Uma comparação adequada entre ambas requer a harmonização destes conceitos. Se, como parece plausível, estes rendimentos não monetários forem apropriados principalmente pelos mais pobres, então, talvez, a concordância entre PNADs e POFs seja apenas ilusória.

TABELA 5
Estatísticas gerais sobre nível, distribuição e composição dos rendimentos nas PNADs e nas POFs (2002 e 2008)

Estatísticas gerais	PNAD 2002 (A)	POF 2002 (B)	Diferença (B - A) (%)	PNAD 2008 (C)	POF 2008 (D)	Diferença (D - C) (%)
Renda per capita (R\$ 2011)	576	709	23	691	941	36
Gini (x100)	58,7	61,2	4	54,4	56,0	3
Extrema pobreza (%)	7,9	8,2	0,3	4,5	2,3	-2,2
Composição da renda (%)	PNAD 2002 (A)	POF 2002 (B)	Diferença (B - A)	PNAD 2008 (C)	POF 2008 (D)	Diferença (D - C)
Trabalho	77	72	-5	77	63	-14
Com carteira	39	39	0	42	35	-6
Sem carteira	11	10	-1	10	9	-2
Conta própria	16	15	-1	14	13	0
Empregador	12	9	-3	11	6	-5
Demais rendas	23	28	5	23	37	14
Previdência	17	13	-5	18	15	-3
Aluguel	2	2	0	2	2	0
Doações e afins	2	3	0	2	2	0
Programas sociais	0	0	0	1	1	0
Outras rendas	1	10	9	1	17	16
Pessoas em domicílios que recebem (%)	PNAD 2002 (A)	POF 2002 (B)	Diferença (B - A)	PNAD 2008 (C)	POF 2008 (D)	Diferença (D - C)
Trabalho	91	98	8	90	91	1
Com carteira	49	54	5	54	52	-2
Sem carteira	36	44	8	34	40	6
Conta própria	34	46	11	31	39	8
Empregador	6	6	-1	7	4	-2
Demais rendas	47	84	37	54	96	42
Previdência	30	27	-2	29	30	1
Aluguel	4	5	2	4	5	1
Doações e afins	8	11	3	8	11	3
Programas sociais	12	14	2	23	23	-1
Outras rendas	38	73	35	38	94	57

Fonte: microdados das PNADs 2002 e 2008 e das POFs 2002-2003 e 2008-2009.

Em suma, as tabelas 4 e 5 indicam que, por um lado, PNADs e censos são muito discrepantes, mas não há culpados óbvios, e que, por outro, a aparente congruência entre PNADs e POFs quanto à desigualdade pode ser apenas ilusória.

4.2 A distribuição de renda

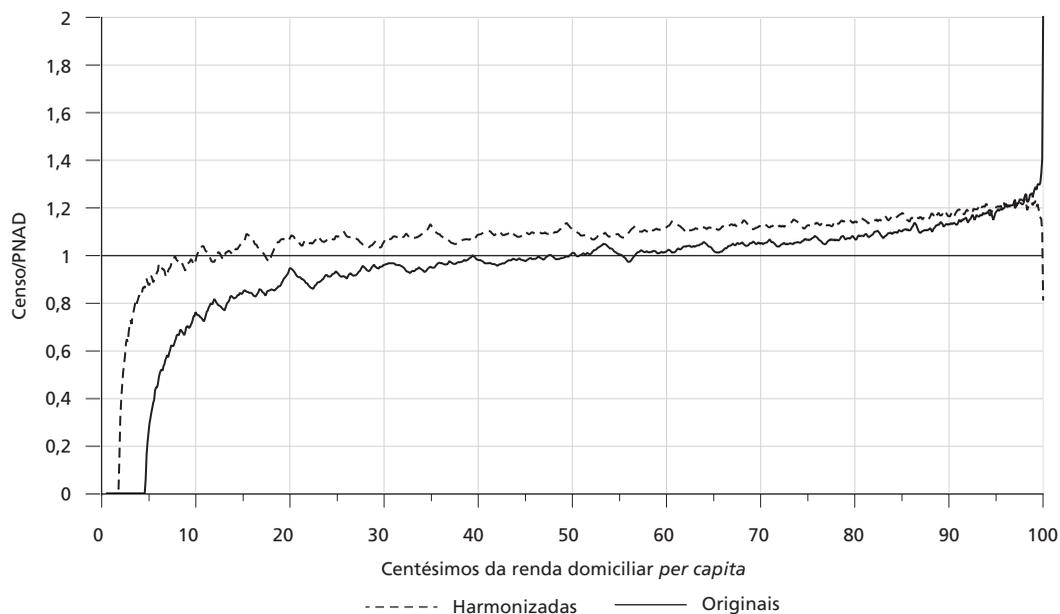
Se as divergências entre as três pesquisas derivarem principalmente dos fatores passíveis de harmonização, então é de se esperar que haja maior proximidade entre as distribuições plenamente harmonizadas que entre as distribuições originais. Os gráficos 9 e 10 proporcionam uma comparação visual: o primeiro mostra, para 2000 e 2010, a razão entre a renda de cada centésimo nos censos e nas PNADs tanto para os dados originais quanto para os dados harmonizados; o segundo faz a mesma comparação para as POFs e as PNADs. Em ambos os casos, se as distribuições fossem idênticas às das PNADs de referência, os valores dos eixos verticais seriam sempre iguais a 1.

No caso dos censos, percebe-se que as divergências entre os dados originais estão nos extremos: os pobres são muito mais pobres e os ricos são muito mais ricos nos censos originais que nas PNADs. No entanto, a harmonização promove grande convergência das distribuições, em especial na cauda inferior. Por exemplo, se nos dados originais as razões entre censos e PNADs se aproximam da unidade apenas quando se chega perto da mediana, nos dados harmonizados isto ocorre já no 10º percentil. No outro extremo, enquanto a renda do 1% mais rico nos dados originais dos censos era 41% e 67% maior que nas PNADs em 2000 e 2010, respectivamente, no caso dos dados harmonizados estas porcentagens caem para 5% e 15%.

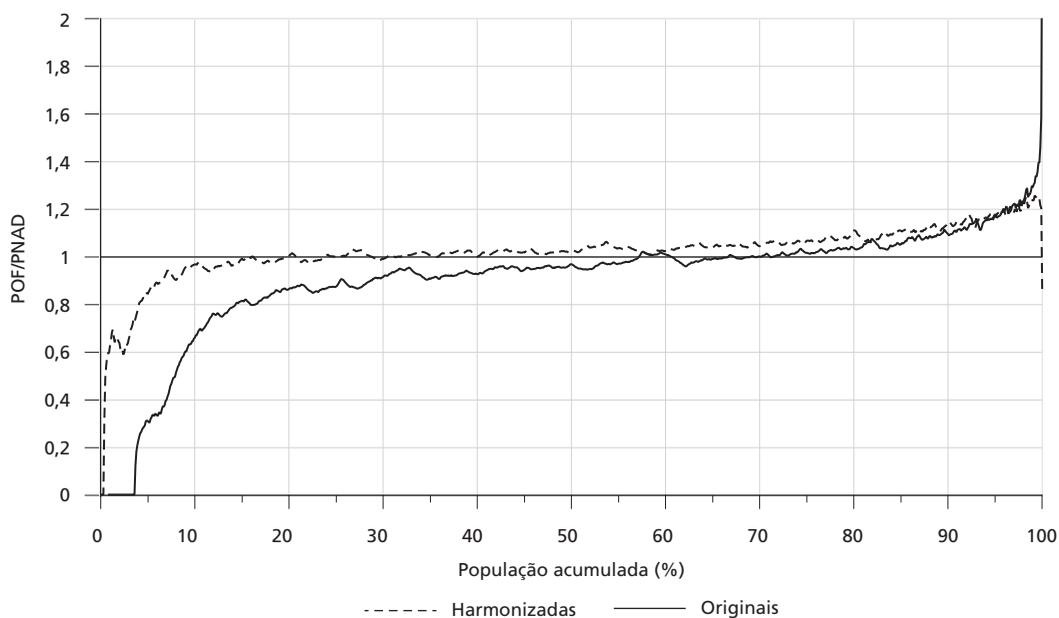
A comparação entre POFs e PNADs também mostra convergência após a harmonização dos dados. Neste caso, contudo, a harmonização afeta muito mais o nível que a desigualdade dos rendimentos. Com exceção dos primeiros centésimos, as distribuições harmonizadas correm paralelas, mas abaixo das originais.

GRÁFICO 9
Razões entre renda domiciliar *per capita* dos centésimos
nos Censos e nas PNADs (2000 e 2010)

9A – Razão entre as rendas do Censo e da PNAD (2000)



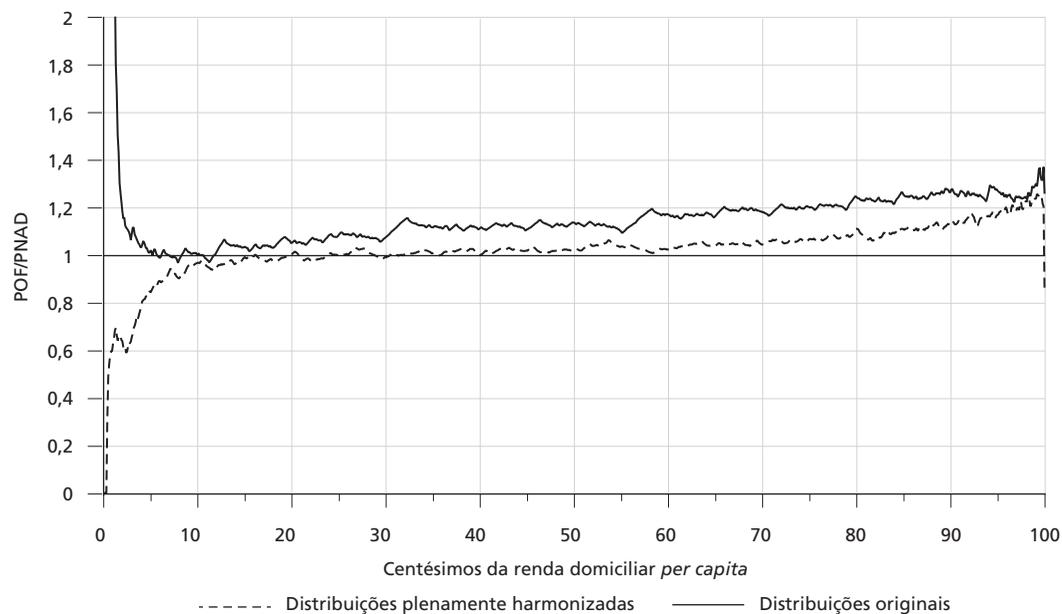
9B – Razão entre as rendas do Censo e da PNAD (2010)



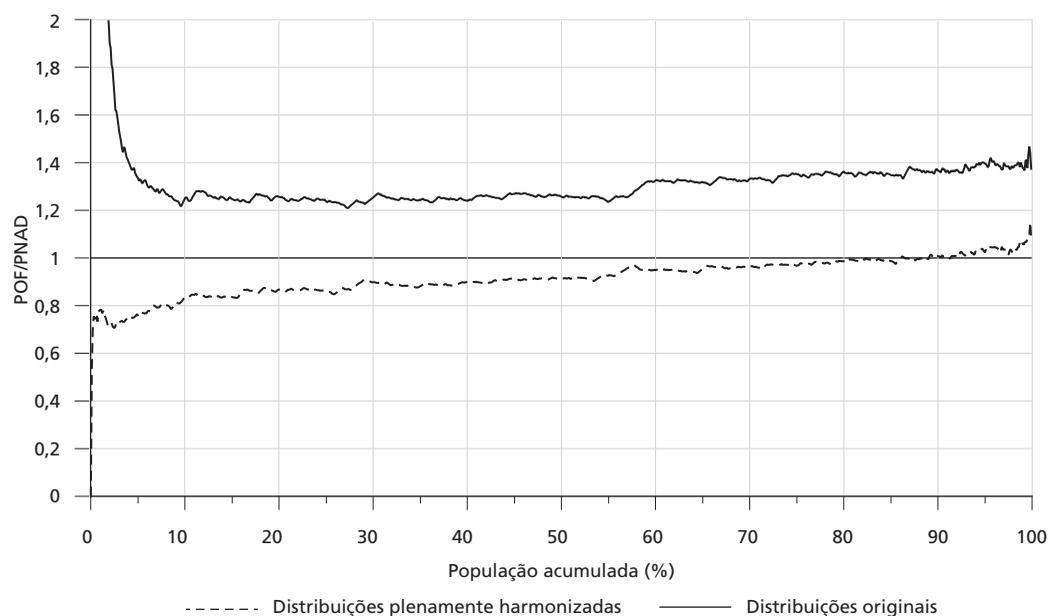
Fonte: microdados das PNADs 1999, 2001, 2009 e 2011; e dos censos 2000 e 2010.

GRÁFICO 10
Razões entre renda domiciliar *per capita* dos centésimos nas POFs e nas PNADs (2002 e 2008)

10A – Razão entre as rendas da POF 2002-2003 e da PNAD 2002



10B – Razão entre as rendas da POF 2008-2009 e da PNAD 2008



Fonte: microdados das PNADs 2002 e 2008 e das POFs 2002-2003 e 2008-2009.

A comparação visual, no entanto, é imprecisa por definição. Uma forma de quantificar com mais nitidez a distância original e a convergência entre as distribuições é o teste de Kolmogorov-Smirnov (K-S). Trata-se de um teste não paramétrico que compara as funções de distribuição acumuladas empiricamente obtidas de duas amostras de tamanhos n e m ($F_{1,n}(x)$ e $F_{2,m}(x)$). A estatística de K-S pode ser obtida por:

$$D_{n,m} = \sup | F_{1,n}(x) - F_{2,m}(x) |$$

Na prática, a estatística de K-S é igual ao maior valor absoluto da diferença entre as duas funções de distribuição empíricas ao longo dos valores possíveis de x e varia entre 0 (quando as distribuições são idênticas) e 1 (maior diferença possível, quando não há sobreposição de valores entre ambas). A hipótese nula de que ambas as amostras são provenientes de uma mesma população é testada contra os valores críticos apropriados da distribuição de Kolmogorov (K_a) e pode ser rejeitada se:

$$\sqrt{\frac{m}{n+m}} D_{n,m} > K_a$$

Na tabela 6, a aplicação do teste de K-S para as comparações entre censos e PNADs e entre POFs e PNADs revela que, em todos os casos, a harmonização torna as distribuições expressivamente mais similares, ainda que – principalmente por causa do tamanho das amostras – estatisticamente diferentes. Ou seja, embora a hipótese nula de que se trata de amostras retiradas da mesma população seja sempre rejeitada, é inegável que os três filtros de harmonização dos dados promovem a convergência das distribuições.

TABELA 6
Estatísticas de Kolmogorov-Smirnov para os dados originais e harmonizados

Comparação	Originais	Harmonizadas	Diferença
Censo 2000 versus PNAD 2000	0,049	0,036	-0,013
Censo 2010 versus PNAD 2010	0,048	0,042	-0,006
POF 2002 versus PNAD 2002	0,067	0,050	-0,017
POF 2008 versus PNAD 2008	0,117	0,047	-0,071

Fonte: microdados das PNADs 1999, 2001, 2002, 2008, 2009 e 2011; dos censos 2000 e 2010; e das POFs 2002-2003 e 2008-2009.

A tabela 7 decompõe a variação no teste de K-S para mostrar as contribuições absolutas e relativas de cada um dos três filtros para a convergência das distribuições. No caso dos censos, há uma inversão entre 2000 e 2010, o que reflete as mudanças operacionais e

conceituais entre as duas edições. Em 2000, o filtro amostral é responsável por quase três quartos da convergência observada, mas em 2010 seu efeito é bem pequeno e na direção oposta, de modo que o filtro de tratamento é quem responde por toda a aproximação entre os censos e as PNADs. Tamanha mudança sugere, portanto, que a própria série do censo sofre problemas de comparabilidade nas suas duas últimas edições.

Nas POFs, entretanto, os resultados são mais consistentes em função da homogeneidade temporal da pesquisa: o que importa mesmo é o filtro conceitual, ou seja, a harmonização dos rendimentos captados nas POFs e nas PNADs.

TABELA 7
Decomposição da variação das estatísticas de Kolmogorov-Smirnov
entre os dados originais e os harmonizados

Contribuições absolutas	Filtro amostral	Filtro conceitual	Filtro de tratamento	Harmonização completa
Censo 2000 <i>versus</i> PNAD 2000	-0.010	-	-0.004	-0.013
Censo 2010 <i>versus</i> PNAD 2010	0.001	-	-0.007	-0.006
POF 2002 <i>versus</i> PNAD 2002	0.000	-0.018	0.001	-0.017
POF 2008 <i>versus</i> PNAD 2008	0.000	-0.071	0.000	-0.071
Contribuições relativas (%)	Filtro amostral	Filtro conceitual	Filtro de tratamento	Harmonização completa
Censo 2000 <i>versus</i> PNAD 2000	72	-	28	100
Censo 2010 <i>versus</i> PNAD 2010	-20	-	120	100
POF 2002 <i>versus</i> PNAD 2002	-2	105	-3	100
POF 2008 <i>versus</i> PNAD 2008	0	100	0	100

Fonte: microdados das PNADs 1999, 2001, 2002, 2008, 2009 e 2011; dos censos 2000 e 2010; e das POFs 2002-2003 e 2008-2009.

4.3 Os rendimentos médios

O efeito da harmonização sobre a renda domiciliar *per capita* em cada pesquisa não é unívoco, como se vê na tabela 8. Mesmo após a harmonização, os rendimentos médios nos censos continuam cerca de 10% mais altos que nas PNADs, mas as taxas de crescimento ficam mais próximas – a da PNAD continua em 2,8% ao ano (a.a.), e a do censo sobe para 2,7%. Contrariamente ao senso comum e às evidências internacionais, a coleta mais agregada dos dados de renda no Censo 2010 não produziu nenhuma inconsistência gritante com relação aos números de 2000 e não gerou nenhum viés de subestimação dos rendimentos médios, seja nos dados originais, seja nos harmonizados.

Nas POFs, há uma convergência quase perfeita das rendas médias após a harmonização. Tanto os níveis em 2002 quanto as taxas de crescimento até 2008 passam a ficar muito próximos, apesar de a harmonização não ter como contornar algumas diferenças fundamentais entre as pesquisas.

Em suma, o nível da renda varia um pouco entre as três pesquisas harmonizadas, mas sua taxa de crescimento é basicamente a mesma nos censos, nas PNADs e nas POFs – em torno de 3% a.a. Trata-se de uma velocidade um pouco maior que a do crescimento do produto interno bruto (PIB) *per capita*, de acordo com as contas nacionais,²³ mas tamanha concordância entre as três principais pesquisas domiciliares brasileiras é tranquilizadora.

TABELA 8
Renda domiciliar *per capita* nos dados originais e harmonizados:
comparações entre censos e PNADs e entre POFs e PNADs

Comparação	Dados originais			Dados plenamente harmonizados		
	2000	2010	Variação anual (%)	2000	2010	Variação anual (%)
Censo (A)	629	828	2,5	645	848	2,7
PNAD (B)	571	732	2,8	588	767	2,8
A/B (%)	110	113	-	110	111	-

Comparação	Dados originais			Dados plenamente harmonizados		
	2002	2008	Variação anual (%)	2002	2008	Variação anual (%)
POF (C)	709	941	4,8	588	721	3,3
PNAD (D)	576	691	3,1	592	719	3,5
C/D (%)	123	136	-	99	100	-

Fonte: microdados das PNADs 1999, 2001, 2002, 2008, 2009 e 2011; dos censos 2000 e 2010; e das POFs 2002-2003 e 2008-2009.

A tabela 9 traz a decomposição das razões entre as rendas domiciliares *per capita* nos censos e nas PNADs e nas POFs e PNADs. As três primeiras colunas mostram em pontos percentuais (p.p.) quanto cada filtro contribui individualmente para aumentar ou diminuir estas razões. A quarta coluna mostra o efeito combinado dos três filtros, isto é, da harmonização completa. As duas últimas colunas apresentam, respectivamente, os resíduos – ou seja, as diferenças remanescentes após a harmonização – e as diferenças originais.

23. Entre 2000 e 2010, o produto interno bruto (PIB) *per capita* cresceu 2,4% ao ano; entre 2002 e 2008, 2,9%. Para mais informações, ver: <<http://www.ipeadata.gov.br>>.

A harmonização praticamente não tem efeito sobre a comparação entre os censos e as PNADs. Em 2000, o efeito é quase zero, porque os filtros amostral e de tratamento se anulam. Em 2010, os dois filtros ajudam a diminuir esta vantagem, mas a redução total é relativamente pequena. Nas POFs, o quadro é mais consistente: o filtro conceitual é o grande responsável por fazer evaporar a diferença original entre as pesquisas, contando com uma pequena ajuda do filtro de tratamento.

Neste último caso, a própria artificialidade do filtro conceitual sugere que as PNADs de fato subestimam o nível da renda no país, mas as causas disso dependem de um julgamento de valor sobre a qualidade das informações nos censos e nas POFs: se for admitido que as POFs contêm as melhores informações, então a subestimação nas PNADs decorreria majoritariamente de questões conceituais, isto é, das próprias definições de rendimentos que as PNADs pretendem pesquisar; se se eleger os censos como as pesquisas com melhores dados, a subestimação nas PNADs decorreria da má captação dos rendimentos pesquisados entre os indivíduos no topo da distribuição.

De todo modo, a convergência das três pesquisas harmonizadas quanto à taxa de crescimento da renda sugere que, pelo menos no que diz respeito à renda monetária das famílias, o aumento médio de cerca de 3% a.a. na última década pode ser considerado um fato empiricamente bem estabelecido.

TABELA 9

Decomposição das razões entre a renda domiciliar *per capita* nos censos e nas PNADs e nas POFs e nas PNADs

Comparação	Contribuição de cada filtro (p.p.)				Razão entre dados harmonizados	Razão entre dados originais
	Filtro amostral	Filtro conceitual	Filtro de tratamento	Harmonização completa		
Censo 2000 versus PNAD 2000	1	-	-1	0	110	110
Censo 2010 versus PNAD 2010	-2	-	-1	-2	111	113
POF 2002 versus PNAD 2002	1	-22	-3	-24	99	123
POF 2008 versus PNAD 2008	0	-33	-3	-36	100	136

Fonte: microdados das PNADs 1999, 2001, 2002, 2008, 2009 e 2011; dos censos 2000 e 2010; e das POFs 2002-2003 e 2008-2009.

4.4 O índice de Gini

A tabela 10 replica para o índice de Gini a análise feita na tabela 8 para a renda média. De novo, a harmonização faz que os resultados mais discrepantes fiquem muito mais próximos. Originalmente, o nível de desigualdade nos censos era maior em 2000 e a queda, menor. Depois da harmonização, a desigualdade nas PNADs fica quase a mesma, enquanto os índices de Gini dos censos caem bastante. Com isso, a variação anual também converge: as PNADs registram uma diminuição de -0,6 ponto de Gini por ano e os censos, -0,5.

Os efeitos da harmonização sobre as POFs são mais ambíguos. Originalmente, os Ginis já eram muito próximos, com níveis e ritmo de queda superiores nas POFs. Com a harmonização, o Gini, em 2002, fica mais próximo, mas o de 2008 fica mais distante, de modo que a queda da desigualdade passa a ser um pouco mais lenta que nas PNADs, embora esta diferença seja mínima.

De modo mais geral, é possível chegar a conclusões semelhantes às anteriores: assim como no caso da renda, a comparação entre as três bases de dados harmonizadas apresenta pequenas discordâncias quanto ao nível da desigualdade, mas grande convergência no que diz respeito ao seu ritmo de queda.

TABELA 10
Índices de Gini da renda domiciliar *per capita* nos dados originais e harmonizados:
comparações entre censos e PNADs e entre POFs e PNADs

Comparação	Dados originais			Dados plenamente harmonizados		
	2000 (x100)	2010 (x100)	Variação anual (Δ Gini)	2000 (x100)	2010 (x100)	Variação anual (Δ Gini)
Censo (A)	64,6	60,9	-0,4	60,8	56,2	-0,5
PNAD (B)	59,3	53,4	-0,6	59,2	53,3	-0,6
A/B (%)	109	114	-	103	106	-

Comparação	Dados originais			Dados plenamente harmonizados		
	2002 (x100)	2008 (x100)	Variação anual (Δ Gini)	2002 (x100)	2008 (x100)	Variação anual (Δ Gini)
POF (C)	61,2	56,0	-0,9	60,7	57,2	-0,6
PNAD (D)	58,7	54,5	-0,7	58,7	54,2	-0,7
C/D (%)	104	103	-	103	106	-

Fonte: microdados das PNADs 1999, 2001, 2002, 2008, 2009 e 2011; dos censos 2000 e 2010; e das POFs 2002-2003 e 2008-2009.

A tabela 11, por sua vez, mostra o efeito de cada filtro sobre a diferença entre o Gini registrado nos censos e nas PNADs e nas POFs e nas PNADs. No primeiro caso, o filtro amostral é responsável por quase toda a queda entre as razões observadas, enquanto o filtro de tratamento contribui para diminuí-las em cerca de 2 p.p. Vale notar que a principal etapa do filtro amostral que contribui para esta queda é a exclusão do 0,1% mais rico nos censos.

No segundo caso, a harmonização não muda muito a razão entre os Ginis. Apenas o filtro conceitual, em 2008, contribui para um aumento, ou seja, após este filtro, a desigualdade sobe mais nas POFs que nas PNADs. De qualquer modo, tanto os níveis quanto a dinâmica do Gini nas POFs harmonizadas ficam bem próximos das PNADs.

Em outras palavras, as PNADs captam relativamente bem o que pretendem captar. Suas diferenças em relação aos censos e às POFs decorrem, respectivamente, da capacidade dos censos de chegar aos mais ricos e da capacidade das POFs de captar outros rendimentos que as PNADs não pesquisam.

Em termos substantivos, tanto os dados originais quanto os harmonizados concordam, *grosso modo*, quanto ao ritmo da queda da desigualdade no Brasil. O que as PNADs parecem subestimar são os níveis de desigualdade, muito embora seja difícil culpá-las por isto. Afinal, não é possível conduzir anualmente uma pesquisa com amostras grandes como as dos censos e detalhadas como as das POFs.

TABELA 11
Decomposição das razões dos índices de Gini nos censos
e nas PNADs e nas POFs e nas PNADs

Comparação	Contribuição de cada filtro (p.p.)				Razão entre dados harmonizados	Razão entre dados originais
	Filtro amostral	Filtro de tratamento	Filtro conceitual	Harmonização completa		
Censo 2000 versus PNAD 2000	-5	-2	-	-6	103	109
Censo 2010 versus PNAD 2010	-6	-2	-	-8	106	114
POF 2002 versus PNAD 2002	-1	0	0	-2	103	104
POF 2008 versus PNAD 2008	-1	0	3	2	106	103

Fonte: microdados das PNADs 1999, 2001, 2002, 2008, 2009 e 2011; dos censos 2000 e 2010; e das POFs 2002-2003 e 2008-2009.

4.5 A extrema pobreza

Os números relativos à extrema pobreza, apresentados na tabela 12, confirmam as tendências anteriores: após a harmonização, as porcentagens nas três fontes de dados convergem, em especial no caso dos censos e das PNADs, que, afinal, são muito mais parecidos entre si que com a POF.

TABELA 12
**Incidência da extrema pobreza nos dados originais e harmonizados:
 comparações entre censos e PNADs e entre POFs e PNADs**

Comparação	Dados originais			Dados plenamente harmonizados		
	2000 (%)	2010 (%)	Variação anual (p.p.)	2000 (%)	2010 (%)	Variação anual (p.p.)
Censo (A)	13,5	8,9	-0,5	9,2	4,5	-0,5
PNAD (B)	9,5	4,4	-0,5	8,5	3,4	-0,5
A-B (p.p.)	4,0	4,5	-	0,7	1,1	-

Comparação	Dados originais			Dados plenamente harmonizados		
	2002 (%)	2008 (%)	Variação anual (p.p.)	2002 (%)	2008 (%)	Variação anual (p.p.)
POF (C)	8,2	2,3	-1,0	11,5	5,7	-1,0
PNAD (D)	7,9	4,5	-0,6	7,1	3,7	-0,6
C-D (p.p.)	0,3	-2,2	-	4,4	2,0	-

Fonte: microdados das PNADs 1999, 2001, 2002, 2008, 2009 e 2011; dos censos 2000 e 2010; e das POFs 2002-2003 e 2008-2009.

Originalmente, a principal diferença entre censos e PNADs estava nos níveis, não na trajetória de queda da extrema pobreza: embora ambos registrassem um decréscimo anual de -0,5 p.p., em 2010, a incidência da extrema pobreza no censo era duas vezes maior que na PNAD. Após a harmonização, a pobreza extrema diminui nas duas pesquisas, mas a queda é bem maior nos censos, de modo que esta diferença cai de 4,5 p.p. para apenas 1,1 p.p. em 2010. O ritmo de queda, no entanto, permanece quase idêntico em ambas as fontes de dados.

Nas POFs a situação é distinta. Originalmente, POF e PNAD apresentavam porcentagens bem próximas em 2002, mas a queda até 2008 era muito mais rápida na primeira. Depois da harmonização, a extrema pobreza continua caindo mais rapidamente nas POFs, mas parte de um patamar bem mais elevado que nas PNADs.

De modo geral, os resultados relativos à extrema pobreza são um pouco menos unívocos que aqueles referentes à renda média e ao índice de Gini, o que não surpreende

tanto, uma vez que é plausível que todas as diferenças não harmonizáveis já enumeradas afetem mais os mais pobres que os demais. Dos baixos níveis de escolaridade à própria instabilidade dos rendimentos, vários fatores contribuem para que diferenças no treinamento, nos questionários e afins sejam particularmente importantes para este grupo.

A tabela 13 mostra o efeito de cada filtro sobre as diferenças nas porcentagens de pobreza extrema de cada pesquisa com as PNADs. Na comparação com os censos, o filtro amostral, sozinho, reduz a diferença em 2000 e 2010, respectivamente, em 2,1 p.p. e 1,4 p.p. Ou seja, metade (em 2000) e um terço (em 2010) das diferenças decorrem apenas de questões amostrais, principalmente a escolha de municípios não autorrepresentativos nas PNADs de 2001 a 2009. O filtro de tratamento, por sua vez – em especial a imputação de renda para os indivíduos com renda *per capita* igual a zero, mas sem perfil de pobreza extrema –, responde por algo entre 31% (em 2000) e 44% (em 2010) das diferenças. Com isso, os dois filtros em conjunto explicam 75% ou mais do “excesso” de extrema pobreza nos censos.

A comparação entre POFs e PNADs revela que praticamente todas as mudanças ocorrem por causa do filtro conceitual; em particular, pela exclusão dos rendimentos não monetários.

TABELA 13
Decomposição das diferenças entre a incidência de extrema pobreza entre os censos e as PNADs e entre as POFs e as PNADs

Comparação	Contribuição de cada filtro (p.p.)				Diferença entre dados harmonizados (p.p.)	Diferença entre dados originais (p.p.)
	Filtro amostral	Filtro de tratamento	Filtro conceitual	Harmonização completa		
Censo 2000 versus PNAD 2000	-2,1	-1,2	-	-3,3	0,7	4,0
Censo 2010 versus PNAD 2010	-1,4	-2,0	-	-3,4	1,1	4,5
POF 2002 versus PNAD 2002	-0,4	0,8	3,6	4,1	4,4	0,3
POF 2008 versus PNAD 2008	-0,1	0,8	3,6	4,1	2,0	-2,2

Fonte: microdados das PNADs 1999, 2001, 2002, 2008, 2009 e 2011; dos censos 2000 e 2010; e das POFs 2002-2003 e 2008-2009.

Tudo isso leva a três conclusões substantivas. A primeira delas diz respeito à velocidade da queda na extrema pobreza, que também pode ser considerada um fato empiricamente bem estabelecido, pois há grande concordância entre as três pesquisas

tanto nos seus dados originais quanto nos seus dados harmonizados. Na última década, a extrema pobreza parece ter recuado a um ritmo de pelo menos 0,5 p.p. por ano. Os números da POF chegam a ser até um pouco maiores, possivelmente em função do período de referência mais longo.

A segunda conclusão refere-se à eficiência das PNADs para captar as informações sobre os rendimentos possíveis de ser captados, dado seu plano amostral e suas definições conceituais. Depois das harmonizações, mesmo com inúmeros aspectos ainda discrepantes entre as pesquisas, os censos, as PNADs e as POFs apresentam números muito semelhantes.

A terceira conclusão, mais uma vez, é que as PNADs parecem ter maiores problemas quanto aos níveis que quanto às trajetórias, mas a direção do viés depende de julgamento sobre qual a medida de renda relevante: se a renda não monetária for incluída, então o saldo líquido será que as PNADs – e, em maior grau ainda, os censos – superestimam a incidência da pobreza extrema; se a definição for restrita à renda monetária, então o viés amostral da PNAD, que favorece municípios maiores, implicará uma subestimação da pobreza extrema.

No primeiro caso, a extrema pobreza no Brasil já seria residual, sendo hoje possivelmente inferior aos 2% registrados na POF 2008. No segundo, a incidência “real” de extrema pobreza estaria em um meio-termo entre os números originais do censo e os da PNAD, ou seja, entre 6% e 7% em 2010.

5 CONCLUSÃO

Os censos, as PNADs e as POFs são as três principais fontes de dados para o estudo da distribuição de renda no Brasil, mas, muitas vezes, os resultados obtidos em cada uma delas parecem incompatíveis uns com os outros. O objetivo deste texto foi tentar entender estas diferenças, tomando as PNADs como parâmetro.

Para isso, o texto partiu de uma breve recapitulação sobre as origens, as trajetórias e os propósitos de cada pesquisa, e, em seguida, procurou documentar as principais diferenças entre elas, agregando-as em três grandes grupos: *i)* plano amostral; *ii)* plano conceitual; e *iii)* plano da coleta e tratamento dos dados. Em cada um dos três casos, foram desenvolvidos procedimentos de harmonização para tornar os censos, as PNADs

e as POFs mais próximos. Assim como em qualquer harmonização *ex post*, estes ajustes são muito imperfeitos, mas necessários para que a comparação entre as pesquisas possa ser feita de forma apropriada.

O chamado filtro amostral é o mais simples, com três etapas: *i*) a exclusão dos moradores do Norte rural (exceto Tocantins), para compatibilizar os censos, as POFs e as PNADs 2004-2011 com as PNADs anteriores; *ii*) a exclusão do 0,1% da população com maior renda domiciliar *per capita* dos censos, pois se trata de um público que não consegue ser captado em amostras menores, como as das PNADs e das POFs; e *iii*) a exclusão, nos censos, dos municípios não autorrepresentativos não sorteados pelo desenho amostral das PNADs de 2001 a 2009. Infelizmente, não foi possível replicar esta última etapa nas POFs.

O filtro conceitual é mais problemático, porque não há como resolver perfeitamente todas as discrepâncias entre as pesquisas. Por motivos óbvios, nem os períodos de referência nem o grau de agregação das perguntas sobre rendimentos podem ser harmonizados. Com isso, este filtro se limitou a manter nas POFs apenas os rendimentos também captados pelos censos e pelas PNADs, excluindo, principalmente, rendimentos não monetários e alguns relacionados ao trabalho assalariado formal – abono de férias, 13^a salário e afins.

O filtro de tratamento também se limita a harmonizar apenas uma parte das diferenças de coleta e tratamento. No caso dos indivíduos com rendimentos ignorados – isto é, que não quiseram ou não souberam responder a alguma pergunta sobre renda –, foi utilizado nas PNADs um método de imputação por *hot deck* parecido com aquele que o IBGE já aplica nos censos e nas POFs. No caso dos indivíduos com renda domiciliar *per capita* igual a zero, foi empregado um modelo de análise de conglomerados nos censos e nas PNADs para dividi-los em dois grupos: aqueles em domicílios com perfil de pobreza extrema e aqueles em domicílios com perfil claramente não pobre. Em seguida, estes últimos tiveram rendimentos imputados também por *hot deck*.

A etapa seguinte do trabalho foi a comparação entre os dados originais, parcialmente e totalmente harmonizados, dos censos, das PNADs e das POFs para testar se os ajustes são capazes de promover uma convergência significativa entre as três fontes de dados. A expectativa era de que, caso isto acontecesse, então seria lícito supor que as

PNADs captam bem aquilo que se destinam a captar e que as discrepâncias em relação aos censos e às POFs seriam facilmente compreensíveis.

A análise empírica confirma, ao menos parcialmente, essa hipótese principal. Após a harmonização, de modo geral, os resultados obtidos nos censos, nas PNADs e nas POFs são bastante convergentes. Embora persistam variações, todas as grandes discrepâncias somem ou são minimizadas.

Assim, no que diz respeito à distribuição em geral, a convergência entre dados originais e harmonizados ocorre porque, por um lado, os ajustes diminuem as discrepâncias entre os extremos das distribuições nos censos e nas PNADs e, por outro, aproximam-se, ao longo da distribuição nas POFs, daqueles observados nas PNADs. Mais especificamente, a renda domiciliar *per capita* continua cerca de 10% maior nos censos que nas PNADs, mas se torna basicamente idêntica nas POFs e nas PNADs, principalmente por causa do filtro conceitual.

Algo semelhante ocorre com o índice de Gini. Mesmo depois da harmonização, as PNADs tendem a registrar níveis um pouco mais baixos que os censos e as POFs, mas a velocidade da queda é quase a mesma. Esta subestimação da desigualdade decorre principalmente do fato de as amostras das PNADs não serem capazes de chegar até os muito, muito ricos, como as amostras dos censos, e as perguntas sobre rendimentos não captarem tantas fontes quanto as POFs.

Finalmente, a extrema pobreza apresenta resultados um pouco menos unívocos. Há grande convergência entre censos e PNADs depois dos ajustes, mas as POFs e as PNADs passam a divergir mais acentuadamente. Isto ocorre, ao menos em parte, porque os números originais de POFs e PNADs já eram – por pura coincidência – muito próximos, e os ajustes fazem a diferença aumentar um pouco. De qualquer modo, mais uma vez, as tendências são as mesmas nas três pesquisas, antes e depois da harmonização. A queda da extrema pobreza ocorre em ritmo quase idêntico nos censos e nas PNADs – em torno de 0,5 p.p. ao ano –, e é ainda mais rápida nas POFs.

Em resumo, o sucesso da harmonização em aproximar os resultados das três pesquisas confirma que algumas diferenças sistemáticas entre planos amostrais, conceitos e tratamento dos dados são responsáveis por grande parte das discrepâncias observadas nos dados originais. Os dados harmonizados apresentam variações temporais quase

idênticas de renda, desigualdade e pobreza nas três fontes de dados, ainda que os níveis originais não sejam exatamente os mesmos.

Nenhuma pesquisa jamais será capaz de retratar com perfeição o que é o “mundo real”. Escolhas conceituais e operacionais, além dos inevitáveis erros de medida, sempre influenciam os resultados. O importante é entender quais os limites de cada pesquisa e quanto compatíveis entre si elas de fato são. Nesse sentido, a convergência dos resultados após a harmonização é bastante encorajadora: o retrato que emerge do Brasil nos censos, nas PNADs e nas POFs é coerente e robusto. É possível afirmar que, pelo menos de acordo com os dados disponíveis até o momento, o crescimento da renda, a queda da pobreza e a desigualdade podem ser considerados fenômenos bem estabelecidos, bem como concluir que a maior parte das discrepâncias entre censos, PNADs e POFs decorre de questões amostrais, conceituais e de tratamento relativamente intuitivas.

REFERÊNCIAS

BARROS, R. P. A efetividade do salário mínimo em comparação à do programa Bolsa Família como instrumento de redução da pobreza e da desigualdade. *In*: BARROS, R. P.; FOGUEL, M. N.; ULYSSEA, G. (Eds.). **Desigualdade de renda no Brasil**: uma análise da queda recente. Brasília: Ipea, 2007. v. 2.

BARROS, R. P. *et al.* A queda recente da desigualdade de renda no Brasil. *In*: BARROS, R. P.; FOGUEL, M. N.; ULYSSEA, G. (Eds.). **Desigualdade de renda no Brasil**: uma análise da queda recente. Brasília: Ipea, 2006. v. 1.

BARROS, R. P.; CARVALHO, M.; FRANCO, S. O papel das transferências públicas na queda recente da desigualdade de renda brasileira. *In*: BARROS, R. P.; FOGUEL, M. N.; ULYSSEA, G. (Eds.). **Desigualdade de renda no Brasil**: uma análise da queda recente. Brasília: Ipea, 2007. v. 2.

BARROS, R. P.; CURY, S.; ULYSSEA, G. A desigualdade de renda no Brasil encontra-se subestimada? Uma análise comparativa usando PNAD, POF e Contas Nacionais. *In*: BARROS, R. P. DE; FOGUEL, M. N.; ULYSSEA, G. (Eds.). **Desigualdade de renda no Brasil**: uma análise da queda recente. Brasília: Ipea, 2006. v. 1.

CANBERRA GROUP. **Expert group on household income statistics**: final report and recommendations. Ottawa: Canberra Group, 2001. (Final Report and Recommendations).

CAVALCANTE, A. A. M.; PRIORE, S. E.; FRANCESCHINI, S. C. Estudos de consumo alimentar: aspectos metodológicos gerais e o seu emprego na avaliação de crianças e adolescentes. **Revista brasileira de saúde materna e infantil**, v. 4, n. 3, p. 229-240, 2004.

CODER, J.; SCOON-ROGERS, L. **Evaluating the quality of income data collected in the annual supplement to the march current population survey and the survey of income and program participation**. Washington, D.C.: Housing and Household Economic Statistics Division; Bureau of the Census, 1996.

CORSEUIL, C. H.; FOGUEL, M. N. **Uma sugestão de deflatores para rendas obtidas a partir de algumas pesquisas domiciliares do IBGE**. Rio de Janeiro: 2002. (Texto para Discussão, n. 897).

COSTA, R. A. Comentário sobre “A mensuração da subjetividade: notas sobre a variável renda nas PNADs”. *In*: SAWYER, D. (Ed.). **PNADs em foco: anos 80**. Belo Horizonte: ABEP, 1988.

DAVERN, M. *et al.* The effect of income question design in health surveys on family income, poverty and eligibility estimates. **Health services research**, v. 40, n. 5, p. 1.534-1.552, 2005.

DINIZ, B. C. *et al.* As pesquisas de orçamentos familiares no Brasil. *In*: SILVEIRA, F. G. *et al.* (Eds.). **Gasto e consumo das famílias brasileiras contemporânea**. Brasília: Ipea, 2007. v. 2.

FREITAS, M. P. S. DE *et al.* **Amostra Mestra para o Sistema Integrado de Pesquisas Domiciliares**. Rio de Janeiro: IBGE, 2007. (Texto para Discussão, n. 23).

GOUSKOVA, E.; SCHOENI, R. **Comparing the estimates of family income in the PSID and the March Current Population Survey, 1968-2005**. Ann Arbor, Michigan: Institute for Social Research; Survey Research Center-University of Michigan, 2007. (PSID Technical Series, n. 07-01).

HALLAK NETO, J. *et al.* Trabalho e rendimentos na nova série do sistema de contas nacionais. **Revista de economia contemporânea**, v. 12, n. 3, p. 571-592, 2008.

HOFFMANN, R. A subdeclaração dos rendimentos. **Revista São Paulo em perspectiva**, v. 2, n. 1, p. 50-54, 1988.

_____. A distribuição da renda no Brasil no período 1992-2001. **Economia e sociedade**, v. 11, n. 19, p. 213-235, 2002.

_____. As transferências não são a causa principal da redução da desigualdade. **Econômica**, v. 7, n. 2, p. 335-341, 2005.

HOFFMANN, R.; NEY, M. G. A recente queda da desigualdade de renda no Brasil: análise de dados da PNAD, do censo demográfico e das contas nacionais. **Econômica**, v. 10, n. 1, p. 7-39, 2008.

HURST, E. **Are household surveys like tax forms?** Evidence from income underreporting of the self-employed. Cambridge, Mass.: National Bureau of Economic Research, 2010. (NBER Working Paper, n. 16433).

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Censo Demográfico 2000 – Trabalho e rendimentos**: resultados da amostra. Rio de Janeiro: IBGE, 2003.

_____. **Pesquisa de orçamentos familiares 2002-2003: primeiros resultados – Brasil e Grandes Regiões.** Rio de Janeiro: IBGE, 2004.

_____. **Sistema integrado de pesquisas domiciliares.** Rio de Janeiro: IBGE, 2007. (Texto para Discussão, n. 24).

_____. **Manual de entrevista da Pesquisa Nacional por Amostra de Domicílios 2009 (Corpo básico).** 2009. Rio de Janeiro: IBGE, 2009.

_____. **Pesquisa de orçamentos familiares 2008-2009: despesas, rendimentos e condições de vida.** Rio de Janeiro: IBGE, 2010.

_____. **Censo Demográfico 2010: resultados gerais da amostra.** Rio de Janeiro: IBGE, 2012a.

_____. **Pesquisa Nacional por Amostra de Domicílios: síntese de indicadores 2011.** Rio de Janeiro: IBGE, 2012b.

_____. **Estudos e tratamento da variável rendimento no Censo Demográfico 2010.** Rio de Janeiro: IBGE, 2012c.

IBRE – INSTITUTO BRASILEIRO DE ECONOMIA; FGV – FUNDAÇÃO GETULIO VARGAS. **Atualização da estrutura de ponderação do IPC-BR.** Rio de Janeiro: Editora da FGV, 2012.

INHABER, H.; CARROLL, S. **How rich is too rich?** Income and wealth in America. New York: Praeger, 1992.

LLUCH, C. Sobre medições de renda a partir dos censos e das contas nacionais do Brasil. **Pesquisa e planejamento econômico**, v. 12, n. 1, p. 133-148, 1982.

MARTINE, G. *et al.* A PNAD: notas para uma avaliação. *In:* SAWYER, D. (Ed.). **PNADs em foco: anos 80.** Belo Horizonte: ABEP, 1988. p. 281-305.

MARTINE, G. O papel dos organismos internacionais na evolução dos estudos populacionais no Brasil: notas preliminares. **Revista brasileira de estudos de população**, v. 22, n. 2, p. 257-275, 2005.

MATHIOWETZ, N.; BROWN, C.; BOUND, J. Measurement error in surveys of the low-income population. *In:* PLOEG, M. V.; MOFFITT, R.; CITRO, C. F. (Eds.). **Studies of welfare populations: data collection and research issues.** Washington, D.C.: The National Academy Press, 2001.

MÉDICI, A. Notas interpretativas sobre a variável “renda” nos censos demográficos brasileiros. **Revista brasileira de estatística**, v. 47, n. 187, p. 305-348, 1986.

_____. A mensuração da subjetividade: notas sobre a variável renda nas PNADs. *In:* SAWYER, D. (Ed.). **PNADs em foco: anos 80.** Belo Horizonte: ABEP, 1988. p. 121-151.

MENEZES, R. C. E.; OSÓRIO, M. M. Inquéritos alimentares e nutricionais no Brasil: perspectiva histórica. **Nutrire**, v. 34, p. 161-177, 2009.

MEYER, B.; MOK, W.; SULLIVAN, J. **The under-reporting of transfers in household surveys: its nature and consequences**. National Bureau of Economic Research, 2009. (NBER Working Paper, n. 15.181).

MONDINI, L.; MONTEIRO, C. A. Mudanças no padrão de alimentação da população urbana brasileira (1962-1988). **Revista de saúde pública**, v. 28, n. 6, p. 433-439, 1994.

MOORE, J.; STINSON, L.; WELNIAK JUNIOR, E. Income measurement error in surveys: a review. **Journal of official statistics**, v. 16, n. 4, p. 331-361, 2000.

OLIVEIRA, L. A. P.; SIMÕES, C. C. S. O IBGE e as pesquisas populacionais. **Revista brasileira de estudos de população**, v. 22, n. 2, p. 291-302, 2005.

ONU – ORGANIZAÇÃO DAS NAÇÕES UNIDAS. **Principles and recommendations for population and housing censuses**. New York: ONU, 2008.

OSORIO, R. G. **A desigualdade racial de renda no Brasil: 1976-2006**. Brasília: Editora da UnB, 2009.

OSORIO, R. G.; SOARES, S.; SOUZA, P. H. G. F. **Erradicar a pobreza extrema: um objetivo ao alcance do Brasil**. Brasília: Ipea, 2011. (Texto para Discussão, n. 1.619).

ROCHA, S. A investigação da renda nas pesquisas domiciliares. **Economia e sociedade**, v. 12, n. 2, p. 205-224, 2003.

SCHWARTZMAN, S. **O presente e o futuro do IBGE**. Rio de Janeiro: IBGE, 1994.

SILVA, P. L. N.; PESSOA, D. G. C.; LILA, M. F. Análise estatística de dados da PNAD: incorporando a estrutura do plano amostral. **Ciência & saúde coletiva**, v. 7, n. 4, p. 659-670, 2002.

SOARES, F. V. *et al.* Programas de transferência de renda no Brasil: impactos sobre a desigualdade. *In*: BARROS, R. P.; FOGUEL, M. N.; ULYSSEA, G. (Eds.). **Desigualdade de renda no Brasil: uma análise da queda recente**. Brasília: Ipea, 2007. v. 2.

SOARES, S. **Distribuição de renda no Brasil de 1976 a 2004 com ênfase no período entre 2001 e 2004**. Brasília: Ipea, 2006. (Texto para Discussão, n. 1.166).

_____. **Volatilidade de renda e a cobertura do Programa Bolsa Família**. Brasília: Ipea, 2009. (Texto para Discussão, n. 1.459).

SOUZA, P. H. G. F. **Uma metodologia para decompor diferenças entre dados administrativos e pesquisas amostrais, com aplicação para o Programa Bolsa Família e o Benefício de Prestação Continuada na PNAD**. Brasília: Ipea, 2010. (Texto para Discussão, n. 1.517).

SOUZA, P. H. G. F.; OSORIO, R. G.; SOARES, S. **Uma metodologia para simular o Programa Bolsa Família**. Brasília: Ipea, 2011. (Texto para Discussão, n. 1.654).

VASCONCELLOS, M. T. L. **Análise crítica dos métodos de avaliação nutricional de populações, a partir de dados de consumo familiar de energia.** Rio de Janeiro: Escola Nacional de Saúde Pública; Fundação Oswaldo Cruz, 2001.

VAZ, F. M. **Escalas de equivalência e demanda do consumidor.** Brasília: Editora da UnB, 2012.

WEINBERG, D. *et al.* Fifty years of U.S. income data from the current populatoin survey: alternatives, trends, and quality. **The American economic review**, v. 89, n. 2, p. 18-22, 1999.

WHEATON, L. Underreporting of means-tested transfer programs in the CPS and SIPP. **Proceedings of the American Statistical Association**, Alexandria, Social Statistics Section, p. 3.622-3.629, 2007. CD-ROM.

YUBA, T. Y. **Evolução dos preços relativos e da estrutura de gastos com alimentos no município de São Paulo:** uma aplicação do banco de dados do IPC-Fipe de 1939 a 2010. São Paulo: Editora USP, 2011.

APÊNDICE

APÊNDICE A

Construção do índice socioeconômico dos domicílios

QUADRO A.1

Definição dos componentes do índice socioeconômico dos domicílios nas PNADs e nos censos

Variável	Descrição
Geladeira	<i>Dummy</i> para presença de geladeira
Máquina de lavar	<i>Dummy</i> para presença de máquina de lavar roupa
Televisão	<i>Dummy</i> para televisão em cores
Esgoto	<i>Dummy</i> para fossa séptica ou rede geral de esgoto
Lixo	<i>Dummy</i> para coleta direta ou indireta de lixo
Densidade	<i>Dummy</i> para mais de dois moradores por dormitório
Morador por cômodo	Número de moradores por cômodo (mínimo de 0,1; máximo de 1,5)
Água	<i>Dummy</i> para água canalizada em pelo menos um cômodo
Banheiro	<i>Dummy</i> para banheiro de uso exclusivo do domicílio

Elaboração do autor.

TABELA A.1

Médias populacionais dos componentes do índice socioeconômico dos domicílios nas PNADs e nos censos

PNADs	Geladeira	Máquina de lavar	Televisão	Esgoto	Lixo	Densidade	Morador por cômodo	Água	Banheiro
1995	74,9	26,7	61,0	48,4	72,1	30,1	0,75	80,5	85,9
1996	78,4	30,5	69,4	51,3	73,3	28,8	0,74	83,6	87,2
1997	80,3	31,7	74,1	50,8	76,3	28,2	0,74	83,3	87,6
1998	81,9	32,3	78,1	52,7	78,3	26,4	0,72	84,7	88,8
1999	82,8	32,8	79,7	52,9	80,0	25,9	0,71	85,6	89,7
2001	85,1	33,6	83,0	54,6	83,2	25,1	0,70	87,3	90,7
2002	86,6	33,9	85,0	56,0	84,8	23,8	0,68	88,7	91,9
2003	87,3	34,3	85,9	56,7	85,6	22,3	0,67	89,2	92,7
2004	87,3	34,3	86,8	56,3	84,6	22,4	0,67	89,2	93,2
2005	87,8	35,6	88,5	56,4	85,5	21,5	0,66	89,9	93,5
2006	89,0	37,3	91,2	57,1	86,3	20,4	0,64	91,0	94,1
2007	90,6	39,2	93,4	58,9	87,1	18,8	0,64	91,7	94,5
2008	91,9	41,5	94,4	59,2	87,7	17,7	0,61	92,5	95,2
2009	93,3	44,3	95,2	59,0	88,4	17,4	0,61	93,0	95,6
2011	95,8	51,0	96,6	62,6	88,8	15,6	0,60	94,2	96,4
Censos	Geladeira	Máquina de lavar	Televisão	Esgoto	Lixo	Densidade	Morador por cômodo	Água	Banheiro
2000	83,4	32,9	87,2	62,7	79,1	27,7	0,73	83,2	83,6
2010	93,7	47,3	95,1	67,2	87,4	18,1	0,64	90,7	93,7

Fonte: microdados das PNADs 1995-2011 e dos censos 2000 e 2010.

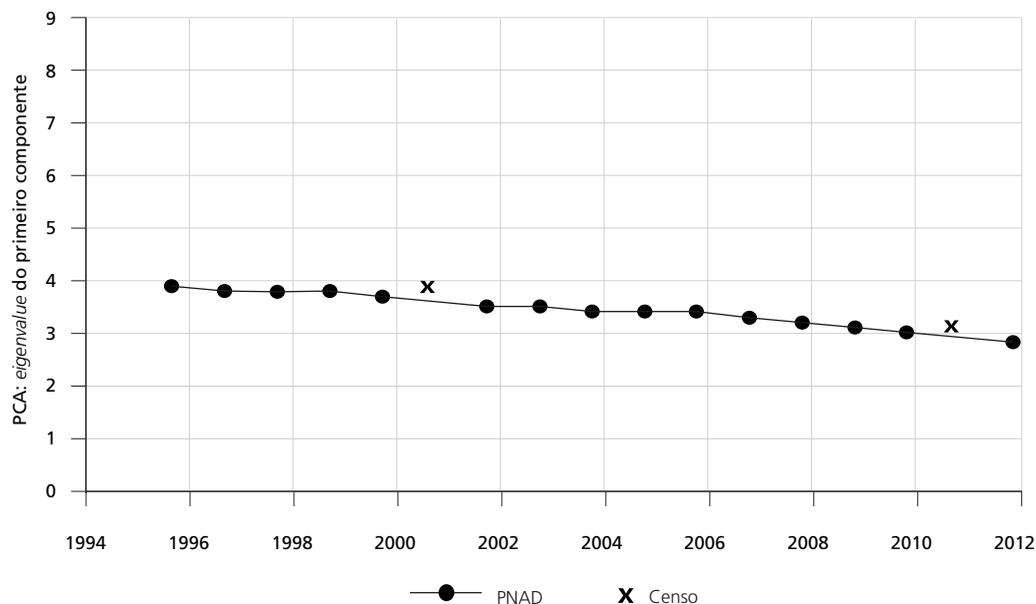
TABELA A.2
Médias dos componentes do índice socioeconômico dos domicílios

PNADs	Geladeira	Máquina de lavar	Televisão	Esgoto	Lixo	Densidade	Morador por cômodo	Água	Banheiro
1995	74,9	26,7	61,0	48,4	72,1	30,1	0,75	80,5	85,9
1996	78,4	30,5	69,4	51,3	73,3	28,8	0,74	83,6	87,2
1997	80,3	31,7	74,1	50,8	76,3	28,2	0,74	83,3	87,6
1998	81,9	32,3	78,1	52,7	78,3	26,4	0,72	84,7	88,8
1999	82,8	32,8	79,7	52,9	80,0	25,9	0,71	85,6	89,7
2001	85,1	33,6	83,0	54,6	83,2	25,1	0,70	87,3	90,7
2002	86,6	33,9	85,0	56,0	84,8	23,8	0,68	88,7	91,9
2003	87,3	34,3	85,9	56,7	85,6	22,3	0,67	89,2	92,7
2004	87,3	34,3	86,8	56,3	84,6	22,4	0,67	89,2	93,2
2005	87,8	35,6	88,5	56,4	85,5	21,5	0,66	89,9	93,5
2006	89,0	37,3	91,2	57,1	86,3	20,4	0,64	91,0	94,1
2007	90,6	39,2	93,4	58,9	87,1	18,8	0,64	91,7	94,5
2008	91,9	41,5	94,4	59,2	87,7	17,7	0,61	92,5	95,2
2009	93,3	44,3	95,2	59,0	88,4	17,4	0,61	93,0	95,6
2011	95,8	51,0	96,6	62,6	88,8	15,6	0,60	94,2	96,4

Censos	Geladeira	Máquina de lavar	Televisão	Esgoto	Lixo	Densidade	Morador por cômodo	Água	Banheiro
2000	83,4	32,9	87,2	62,7	79,1	27,7	0,73	83,2	83,6
2010	93,7	47,3	95,1	67,2	87,4	18,1	0,64	90,7	93,7

Fonte: microdados das PNADs 1995-2011 e dos censos 2000 e 2010.

GRÁFICO A.1
Eigenvalues do primeiro componente da análise de componentes principais



Fonte: microdados das PNADs 1995-2011 e dos censos 2000 e 2010.

EDITORIAL

Coordenação

Cláudio Passos de Oliveira

Supervisão

Everson da Silva Moura

Reginaldo da Silva Domingos

Revisão

Andressa Vieira Bueno

Clícia Silveira Rodrigues

Idalina Barbara de Castro

Laeticia Jensen Eble

Leonardo Moreira de Souza

Luciana Dias

Marcelo Araújo de Sales Aguiar

Marco Aurélio Dias Pires

Olavo Mesquita de Carvalho

Celma Tavares de Oliveira (estagiária)

Patricia Firmina de Oliveira Figueiredo (estagiária)

Editoração

Aline Rodrigues Lima

Bernar José Vieira

Daniella Silva Nogueira

Daniilo Leite de Macedo Tavares

Jeovah Herculano Szervinsk Junior

Leonardo Hideki Higa

Daniel Alves de Sousa Júnior (estagiário)

Diego André Souza Santos (estagiário)

Capa

Luís Cláudio Cardoso da Silva

Projeto Gráfico

Renato Rodrigues Bueno

Livraria do Ipea

SBS – Quadra 1 - Bloco J - Ed. BNDES, Térreo.

70076-900 – Brasília – DF

Fone: (61) 3315-5336

Correio eletrônico: livraria@ipea.gov.br

Composto em adobe garamond pro 12/16 (texto)
Frutiger 67 bold condensed (títulos, gráficos e tabelas)
Impresso em offset 90g/m²
Cartão supremo 250g/m² (capa)
Brasília-DF

Missão do Ipea

Produzir, articular e disseminar conhecimento para aperfeiçoar as políticas públicas e contribuir para o planejamento do desenvolvimento brasileiro.

