

A distribuição de renda nas pesquisas domiciliares brasileiras: harmonização e comparação entre Censos, PNADs e POFs*

Pedro Herculano Guimarães Ferreira de Souza**

O objetivo deste trabalho é documentar e explicar as diferenças nas distribuições de renda do Censo Demográfico, da Pesquisa Nacional por Amostra de Domicílios (PNAD) e da Pesquisa de Orçamentos Familiares (POF). A principal hipótese é a de que é possível promover grande convergência dos resultados entre as três pesquisas com procedimentos de harmonização *ex post*, que compatibilizam, na medida do possível, diferenças amostrais, conceituais e de coleta e tratamento dos dados. Os resultados confirmam, em boa medida, esta hipótese: de modo geral, a harmonização aproxima as três pesquisas e reduz significativamente as maiores discrepâncias entre as distribuições de renda, em especial na comparação entre Censo e PNAD. Embora persistam em alguns casos diferenças quanto aos níveis de renda, desigualdade e pobreza, sua evolução ao longo do tempo torna-se muito semelhante nas três pesquisas. Por fim, observa-se também que as discrepâncias remanescentes seguem um padrão, ou seja, mesmo após a harmonização, a distribuição de renda na PNAD tende a ser um pouco mais igualitária do que no Censo e na POF: os rendimentos dos mais pobres são mais altos e os dos mais ricos, mais baixos.

Palavras-chave: Pesquisas domiciliares. Mensuração da renda. Distribuição de renda. Desigualdade. Pobreza.

* O autor agradece os comentários de Rodolfo Hoffmann, Marcelo Medeiros, Fábio Veras Soares e dos dois pareceristas anônimos.

** University of Califórnia, Economics Department, Berkeley, Estados Unidos, e Instituto de Pesquisa Econômica Aplicada – Ipea (pedrosouza@gmail.com).

Introdução

O crescimento pró-pobre – isto é, com redução da desigualdade – na primeira década dos anos 2000 foi amplamente documentado e continua sendo objeto de inúmeras investigações (HOFFMANN, 2005; SOARES, 2006; BARROS; FOGUEL; ULYSSEA, 2006), o que só foi possível em função de outra novidade menos comentada: a profusão de bancos de dados de alta qualidade publicamente disponíveis. Essa abundância de dados é extremamente positiva, pois permite que os pesquisadores avaliem a robustez de suas descobertas. Contudo, ela também desperta dúvidas, já que nem sempre os números obtidos são compatíveis entre si.

Este trabalho procura entender as discrepâncias nas distribuições de renda nas três pesquisas domiciliares regularmente conduzidas pelo IBGE que têm cobertura nacional e coletam informações sobre rendimentos: o Censos Demográficos, a Pesquisa Nacional por Amostra de Domicílios (PNAD) e a Pesquisa de Orçamentos Familiares (POF).

Para isso, o artigo 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 outros indicadores importantes são comparados antes e depois da harmonização.

A principal hipótese é a de que um conjunto relativamente pequeno de ajustes é capaz de promover grande convergência nas distribuições de renda no Censo, na PNAD e na POF. Como a maior parte desses ajustes tem como objetivo tornar o Censo e a POF mais parecidos com a PNAD, a hipótese pode ser colocada em outros termos: a PNAD mede bem aquilo que seu desenho se propôs a medir e permite inferências adequadas – pelo menos dentro das limitações inerentes às pesquisas domiciliares –, principalmente para as tendências dos rendimentos ao longo do tempo. A convergência das distribuições será testada por meio da comparação das distribuições como um todo e por três aspectos centrais para o debate público: renda média; desigualdade medida pelo índice de Gini; e percentual de extrema pobreza.

Esse objetivo é diferente da abordagem comum da literatura, que tenta validar os rendimentos declarados nas pesquisas domiciliares, comparando-os com referências externas. As estratégias mais comuns são a comparação de agregados com registros administrativos ou com as contas nacionais

Em geral, os resultados dessa abordagem mostram que, no Brasil, as pesquisas domiciliares subestimam o nível dos rendimentos e a participação em programas sociais, como o Bolsa Família, e provavelmente também a desigualdade de renda, embora haja menos consenso nessa área (LLUCH, 1982; HOFFMANN, 1988; BARROS; CURY; ULYSSEA, 2006; HOFFMANN; NEY, 2008; SOUZA, 2013; MEDEIROS; SOUZA; CASTRO, 2014a, 2014b). A literatura internacional chega a resultados parecidos. Primeiro, a participação em programas governamentais e os rendimentos associados a eles são quase sempre

subestimados, com alguma piora ao longo do tempo, e com grande variação entre programas (MATHIOWETZ; BROWN; BOUND, 2001; WEINBERG et al., 1999; WHEATON, 2007; MEYER; MOK; SULLIVAN, 2009). Segundo, algumas fontes de rendimentos – como rendas de propriedade, juros e afins, ou rendimentos de empregadores e contas próprias – também são notoriamente mal captadas nas pesquisas domiciliares (CANBERRA GROUP, 2001; HURST, 2010). Terceiro, os resultados são sensíveis ao desenho das pesquisas e variam bastante entre levantamentos (CODER; SCOON-ROGERS, 1996; MOORE; STINSON; WELNIAK JR., 2000; GOUSKOVA; SCHOENI, 2007).

Neste trabalho, a ideia não é comparar os resultados com números “verdadeiros”, mas sim explicar diferenças entre pesquisas. Um dos pressupostos é que um fenômeno pode ser considerado bem estabelecido sempre que as três fontes de dados – Censo, PNAD e POF – concordarem entre si, independentemente da validação externa.

Naturalmente, é possível que haja problemas comuns às três fontes, mas sua investigação está além do escopo desse estudo. Por exemplo, há evidências baseadas em dados tributários que indicam que as pesquisas domiciliares brasileiras subestimam a concentração de renda no topo da distribuição e que houve estabilidade – e não queda – da desigualdade entre 2006 e 2012 (MEDEIROS; SOUZA; CASTRO, 2014a, 2014b).

Comparações parciais entre Censo, PNAD e POF já foram feitas por alguns autores, juntamente com o contraste com as contas nacionais, mas sempre em apenas um momento do tempo e nunca entre as três pesquisas simultaneamente (LLUCH, 1982; BARROS; CURY; ULYSSEA, 2006; HOFFMANN; NEY, 2008). Este trabalho cobre todas as edições realizadas desde o fim da hiperinflação, o que inclui dois censos (2000 e 2010), 15 PNADs (1995-2011) e duas POFs com cobertura nacional (2002-2003 e 2008-2009 – para simplificar, 2002 e 2008).

Breve caracterização das pesquisas

O Censo Demográfico de 1872 foi a primeira pesquisa domiciliar brasileira de cobertura nacional. Desde então, a periodicidade decenal foi mantida, salvo poucas exceções, mas o escopo dos dados coletados cresceu continuamente (MÉDICI, 1986; OLIVEIRA; SIMÕES, 2005).

A coleta de informações sobre rendimentos data de 1890. Depois, os rendimentos só reapareceram em 1960, como uma única questão com respostas fechadas em múltiplos do salário mínimo. A renda continuou a ser coletada nos censos subsequentes, em questões abertas com detalhamento crescente até 2000, quando foram pesquisadas oito diferentes fontes. O Censo 2010 discriminou os rendimentos em apenas três fontes (trabalho principal, demais trabalhos e outros rendimentos).

A PNAD surgiu como parte do esforço de aperfeiçoamento das estatísticas nacionais no pós-guerra (MARTINE et al., 1988; MÉDICI, 1988; MARTINE, 2005). Sua primeira edição foi conduzida em 1967, com caráter trimestral e abrangência geográfica restrita. Em 1974 e 1975, a PNAD foi interrompida para realização do Estudo Nacional de Despesas Familiares

(Endef). Seu retorno, em 1976, marcou uma nova etapa, com mudanças no questionário e periodicidade anual, exceto em anos censitários e em 1994 (SCHWARTZMAN, 1994). Não houve mudanças estruturais nos últimos 30 anos.

No caso dos rendimentos, desde 1981, ocorreram apenas pequenas mudanças nas perguntas sobre o trabalho e maior desagregação dos “outros rendimentos”, elevando para 11 as fontes pesquisadas. Apesar de positiva, tamanha consistência também produziu efeitos indesejados: até 2011, a PNAD continuou captando os rendimentos do abono de permanência – extinto em 1993 –, mas nunca incluiu no questionário básico perguntas sobre programas como o Bolsa Família e seus predecessores (ROCHA, 2003; MÉDICI, 1988). Isso só deverá ocorrer nos próximos anos, com a reformulação das pesquisas domiciliares do IBGE.

Finalmente, as origens da POF remontam ao Endef, a primeira pesquisa de grande abrangência a trazer dados de rendimentos e despesas e a coletar informações antropométricas e relativas ao consumo alimentar. Sua complexidade e alto custo fizeram com que uma nova edição viesse a ser realizada somente em 1987-1988, já com novo nome – Pesquisa de Orçamentos Familiares – e escopo mais restrito. As edições subsequentes reverteram esse movimento: tanto o consumo alimentar quanto a antropometria voltaram em 1995-1996, e as edições de 2002-2003 e 2008-2009 recuperaram os aluguéis estimados para famílias em residência própria e tiveram abrangência nacional (DINIZ et al., 2007; VAZ, 2012).

A coleta dos rendimentos é mais completa na POF do que na PNAD e no Censo. Os microdados da primeira edição com cobertura nacional, em 2002-2003, discriminam mais de 80 tipos de rendimentos monetários. Em 2008-2009, são cerca de 110.

Tratamento dos dados

Para harmonizar as fontes de dados, as divergências entre elas foram agrupadas em três categorias: plano amostral; definições conceituais; e coleta e tratamento dos dados.

A seleção das unidades de análise constituiu etapa preliminar: os resultados apresentados dizem respeito à renda domiciliar *per capita* medida para os moradores de domicílios particulares permanentes (exceto pensionistas e afins). Os rendimentos foram deflacionados pelo INPC para outubro de 2011, data da PNAD mais recente quando da elaboração deste texto. A linha de pobreza extrema é de R\$ 70 *per capita*, seguindo a definição do Programa Bolsa Família em 2011. Todas as estimativas consideram os pesos amostrais das pesquisas.

Diferenças no desenho amostral

Representatividade geográfica

As populações de pesquisa dos levantamentos são semelhantes. A harmonização se faz necessária apenas em função da PNAD: até 2004, sua cobertura não incluía áreas rurais

da Região Norte (exceto Tocantins), ao contrário do Censo, POF e PNADs recentes. Assim, o primeiro ajuste para harmonização amostral implicou descartar essas áreas, de modo que a população considerada nas análises subsequentes consiste nos residentes no território nacional, exclusive moradores das áreas rurais da Região Norte (exceto Tocantins). Os pesos amostrais não foram recalculados.

Tamanho das amostras

Entre 1995 e 2011, as PNADs ampliaram de 85 mil para 112 mil domicílios entrevistados, enquanto as duas POFs visitaram com sucesso 49 mil e 56 mil domicílios. As amostras dos Censos são entre 50 e 100 vezes maiores, uma vez que incluem porcentagens relativamente constantes do total de domicílios – 11,7% em 2000 e 10,7% em 2010 (IBGE, 2003; 2012a).

Para a maior parte das variáveis, isso não representa problema, mas o caso da renda é especial. Os rendimentos não têm limite superior necessário e a distribuição de renda aproxima-se de uma distribuição de Pareto no topo. Amostras de grande tamanho, ao alcançar melhor os muito ricos, podem apresentar indicadores distintos de amostras menores.

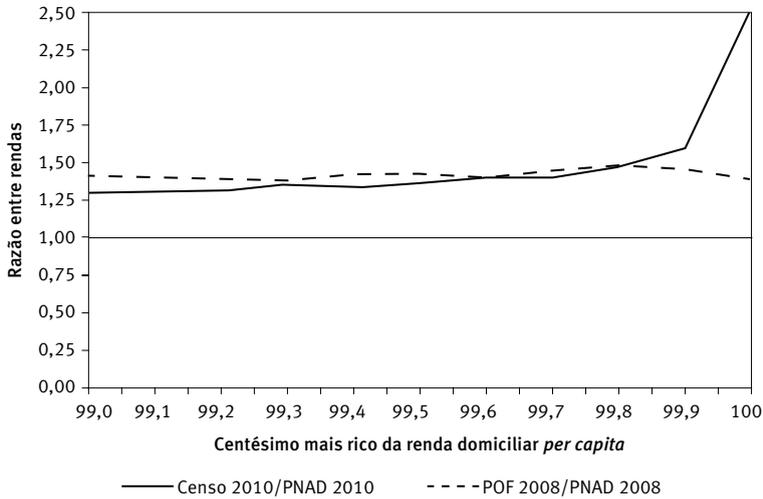
A maior capacidade em localizar e entrevistar os muito ricos é uma vantagem dos Censos: a probabilidade de PNADs ou POFs representarem corretamente grupos tão diminutos é tão irrisória que, na prática, pode-se considerar que eles não fazem parte do universo representado por ambas.

O Gráfico 1 mostra as razões entre as rendas médias de cada milésimo da população pertencente ao centésimo mais rico da população no Censo 2010 e na PNAD 2010 (ver a seção *Seleção dos anos para comparação*) e na POF 2008 e na PNAD 2008. No primeiro caso, há uma descontinuidade quando se chega ao milésimo mais rico; no segundo, isso não ocorre. Assim, no Censo 2010, a maior parte do 1% mais rico da população ganha entre 25% e 50% a mais que nas PNADs, mas quando se chega ao 0,1% mais rico, esta porcentagem dispara para algo entre 100% e 150% a mais.

Como isso não ocorre na comparação com as POFs, é razoável concluir que a influência do tamanho da amostra dos Censos se manifesta, no topo, sobretudo a partir do centésimo mais rico e, indiscutivelmente, acima do milésimo mais rico. Os mesmos padrões são encontrados quando se comparam o Censo 2000 com a PNAD 2000 sintética e a POF 2002 com a PNAD 2002.

Logo, o segundo procedimento de harmonização das amostras consistiu na eliminação do 0,1% mais rico nos Censos. Vale notar que esse procedimento visa, exclusivamente, aumentar a comparabilidade entre pesquisas, e não aproximá-las do “mundo real”: pelos motivos discutidos anteriormente, é muito provável que as informações do Censo para o topo da distribuição sejam mais precisas do que as das POFs e PNADs. Além disso, cabe notar também que a escolha do ponto de corte poderia ser diferente.

GRÁFICO 1
Razão entre as rendas dos milésimos dentro do centésimo mais rico da renda domiciliar *per capita*:
Censo 2010/PNAD 2010 e POF 2008/PNAD 2008



Fonte: IBGE. PNADs 2008, 2009 e 2011; Censo 2010; POF 2008-2009.

Seleção das amostras

Os Censos visitam todos os municípios do país. Em cada setor censitário, os domicílios são sorteados com equiprobabilidade, com fração amostral constante para setores no mesmo município. Em 2000, havia duas frações amostrais, 10% e 20%; em 2010, foram quatro faixas, de 5% a 50% (IBGE, 2003; 2012a).

A PNAD segue amostragem em três estágios, na qual municípios são as unidades primárias, setores censitários são as secundárias e domicílios, as terciárias. Os municípios são classificados em três estratos: o primeiro engloba as nove regiões metropolitanas mais antigas e o Distrito Federal (estrato das RM); o segundo é composto pelos municípios autorrepresentativos (AR), isto é, municípios não metropolitanos de grande porte; e o terceiro abarca os não autorrepresentativos (NAR), reunindo os demais municípios.

No primeiro estágio, os municípios RM e AR são necessariamente selecionados, isto é, sorteados com probabilidade igual a 1. Os municípios NAR são agrupados principalmente por critérios geográficos e sorteados sem reposição, em seus grupos, com probabilidades proporcionais à população. No segundo estágio, os setores censitários são selecionados, em cada município, com probabilidade proporcional ao tamanho e com reposição; e no terceiro estágio, os domicílios são sorteados por amostragem sistemática simples dentro de cada setor (SILVA; PESSOA; LILA, 2002; IBGE, 2010a).

O plano amostral da POF é conglomerado em dois estágios, com setores censitários como unidades primárias e domicílios particulares permanentes como secundárias. O

sorteio do primeiro estágio é feito a partir da estratificação geográfica e socioeconômica das unidades primárias, que são selecionadas, dentro de cada estrato, com probabilidade proporcional ao seu tamanho; o segundo estágio sorteia os domicílios por amostragem aleatória simples (IBGE, 2004). A POF 2008, ao contrário da anterior, foi feita com base na Amostra Mestra, que deverá municiar as pesquisas do IBGE nos próximos anos (FREITAS et al., 2007; IBGE, 2010b).

Em suma, a PNAD visita menos municípios do que a POF e o Censo, e os municípios visitados são mais concentrados geograficamente (IBGE, 2007). Não há o que ser feito para minimizar essa diferença entre as pesquisas. Felizmente, essas diferenças devem ser pouco significativas, pois os pesos amostrais da PNAD levam em conta as probabilidades de inclusão dos municípios.¹

Diferenças conceituais

Definição geral da renda

A definição geral da renda depende de duas decisões: se os rendimentos são brutos ou líquidos; e se devem se restringir aos rendimentos monetários. O padrão internacional recomenda o uso da renda disponível – rendimentos monetários e não monetários líquidos de tributos diretos – como aproximação mais fidedigna da capacidade de consumo das famílias (CANBERRA GROUP, 2001). Quase toda a literatura brasileira, entretanto, limita-se à renda monetária bruta, porque nem Censos nem PNADs trazem informações sobre rendimentos não monetários e tributos.

A questão passa a ser como definir a renda nas POFs. A opção foi apresentar somente dados relativos aos rendimentos brutos – mesmo para tabulações dos dados originais – e considerar como primeira etapa do filtro conceitual a exclusão dos rendimentos não monetários.

Períodos de referência

No Censo e na PNAD, a referência é o mês de realização da pesquisa, enquanto a POF não se restringe a um mês de referência fixo: seu questionário coleta, para os diversos tipos de rendimentos, o valor bruto do último rendimento e o número de vezes que a pessoa recebeu aquele rendimento no período de referência da pesquisa, isto é, os 12 meses anteriores à entrevista.² Além disso, tanto o Censo quanto a PNAD dão espaço para subjetividade ao perguntar qual o rendimento “habitualmente” recebido nos meses de referência, o que gera um sem-número de dificuldades adicionais (MÉDICI, 1988; ROCHA, 2003). Infelizmente, não há como fazer qualquer harmonização dessas diferenças.

¹ As versões anteriores deste texto aplicavam um procedimento para manter nos Censos apenas os municípios sorteados nas PNADs, com efeitos pequenos e/ou insignificantes sobre a maior parte dos resultados. O autor agradece os esclarecimentos prestados por um dos pareceristas anônimos.

² Como as entrevistas da POF são conduzidas ao longo de um ano, os 12 meses que constituem o período de referência não são idênticos para todas as famílias.

Definição dos rendimentos captados

Grosso modo, Censo e PNAD captam um conjunto semelhante de rendimentos, ainda que em níveis de agregação diferentes, enquanto a POF é mais detalhada. Toda a harmonização consistiu, então, em descartar os rendimentos não captados nos Censos e nas PNADs, como pagamentos de 13^o salário e férias, saques do Fundo de Garantia por Tempo de Serviço (FGTS), rendimentos não monetários e afins.³

A Tabela 1 mostra que a renda média na POF cai 18% em 2002 e 24% em 2008 com a exclusão dos rendimentos não captados nas outras pesquisas. Como esperado, os principais responsáveis pela queda são os rendimentos de empregados formais e os outros rendimentos (que incluem os não monetários).

Diferenças na coleta e tratamento dos dados

Se as diferenças conceituais são mais graves e menos tratáveis que as amostrais, as diferenças na coleta e tratamento dos dados são ainda maiores. Há diversas questões irreversíveis relacionadas ao treinamento dos entrevistadores, à duração da entrevista, à ordem das perguntas, ao conhecimento dos entrevistados sobre as pesquisas, etc.

TABELA 1
Estimativas da renda domiciliar *per capita* nas POFs antes e depois da harmonização conceitual, segundo fontes de rendimento
Brasil – 2002-2008

Fontes de rendimento	POF 2002			POF 2008		
	Original (em reais de 2011)	Harmonizado (em reais de 2011)	Variação (%)	Original (em reais de 2011)	Harmonizado (em reais de 2011)	Variação (%)
Trabalho	514	453	-12,0	591	528	-11,0
Com carteira	275	230	-17,0	333	285	-14,0
Sem carteira	73	66	-9,0	80	73	-9,0
Conta própria	106	99	-6,0	125	120	-4,0
Empregador	60	58	-4,0	53	50	-7,0
Outras fontes	195	128	-34,0	349	184	-47,0
Previdência pública	89	89	0,0	145	136	-6,0
Aluguéis	15	15	0,0	16	16	0,0
Doações e afins	19	19	0,0	21	21	-3,0
Programas sociais	2	2	0,0	7	7	0,0
Outros rendimentos	71	4	-94,0	159	3	-98,0
Total	709	581	-18,0	941	712	-24,0

Fonte: IBGE. POFs 2002/2003 e 2008/2009.

Nota: Cada linha representa a soma populacional da renda total de cada fonte dividida pela estimativa da população total.

A harmonização diz respeito, então, a dois pontos principais: indivíduos com rendimentos ignorados – que não quiseram ou não souberam declarar renda em alguma fonte – e indivíduos com renda domiciliar *per capita* igual a zero.

³ A lista de variáveis utilizadas na construção dos rendimentos na POF pode ser obtida com o autor.

No primeiro caso, tanto o Censo quanto a POF imputam rendimentos válidos via *hot deck* ou semelhante:⁴ grosso modo, os indivíduos são estratificados segundo características socioeconômicas e recebem valores imputados a partir de doadores selecionados aleatoriamente dentro do seu estrato (IBGE, 2003; 2004; 2010b; 2012b). Na PNAD não há imputação. A maioria dos pesquisadores simplesmente descarta os domicílios em que pelo menos um membro possui algum rendimento ignorado. Além de prejudicar a comparação, o problema maior é que houve aumento dos casos nessa situação, que flutuaram em torno de 2% da população até 2005 e chegaram a mais de 5% em 2011. Nos Censos e nas POFs, os números oscilam entre 3% (POF 2002) e 9% (Censo 2000).

Os domicílios com renda igual a zero são praticamente inexistentes nas POFs e muito mais numerosos nas PNADs (cerca de 1% da população desde 1995) e nos Censos (cerca de 4%). O caso da POF é relativamente fácil de entender, pois não há um mês de referência fixo e a coleta de informações sobre despesas provavelmente induz a declaração de rendimentos positivos. Já na PNAD e no Censo, isso causa mais estranhamento, pois, como a pergunta diz respeito aos rendimentos habituais, seria de se esperar que apenas poucas famílias tenham rendimentos tão sazonais a ponto de ser “normal” não ter nenhuma renda nos meses de referência.

Osorio, Soares e Souza (2011) argumentam que boa parte desses domicílios nas PNADs possui perfil incompatível com a ausência de renda, sendo “não pobres” em situação temporária. Hoffmann e Ney (2008) também identificaram casos muito atípicos no Censo 2000. É provável que o problema seja ainda mais agudo no Censo 2010, pois há fortes suspeitas de que “a opção ‘não tem rendimento’ serviu [...], em um número significativo de situações, como alternativa para a categoria ‘ignorado’” (IBGE, 2012b).

Imputação dos rendimentos ignorados nas PNADs

A imputação de valores válidos de renda para os indivíduos com rendimentos ignorados deu-se via *hot deck*. Para os rendimentos do trabalho – responsáveis por cerca de 80% da não declaração – e abono de permanência, o *hot deck* foi estratificado por Unidade da Federação, área censitária (região metropolitana, município autorrepresentativo, município não autorrepresentativo), educação (nenhuma, fundamental completo, médio completo, superior completo) e posição na ocupação (formal, informal, conta própria, empregador). Para as demais rendas, a estratificação foi por Unidade da Federação, área censitária, gênero e educação. Especificações alternativas foram testadas, sem mudanças significativas nos resultados.

⁴ Mais precisamente, no Censo 2000 as variáveis de rendimento passaram por processo de crítica dentro do sistema DIA (Detección e Imputación Automática de Errores para Datos Cualitativos), e a imputação foi feita pela técnica de árvores de regressão (IBGE, 2003). No Censo 2010, a crítica e a imputação foram feitas pelo sistema Canceis (Canadian Census Edit and Imputation System) (IBGE, 2012b). Nas POFs 2002 e 2008, a imputação da não resposta para despesas e rendimentos deu-se via *hot deck* (IBGE, 2004; 2010b). De modo geral, todos esses métodos consistem em atribuir aos casos problemáticos valores observados em doadores com características semelhantes.

A imputação foi precedida pela análise do perfil dos domicílios com renda ignorada ao longo do tempo. Os resultados não serão discutidos em detalhe, mas, de modo geral, pode-se afirmar que o aumento da não declaração de renda foi concomitante à diminuição do viés: antes, indivíduos com alta escolaridade, empregadores e contas próprias tinham probabilidade muito maior de não declarar rendimentos, mas nos últimos anos houve estreitamento dos diferenciais.

Identificação e imputação de rendimentos para domicílios com renda igual a zero e sem perfil de extrema pobreza

O tratamento dos domicílios com renda zero foi feito em duas etapas. Primeiro, aplicou-se modelo de análise de conglomerados para identificar quais não tinham perfil socioeconômico compatível com a extrema pobreza (OSORIO et al., 2011). Em seguida, foi feita imputação de rendimentos para esses domicílios via *hot deck*.

Quatro variáveis foram utilizadas na análise de conglomerados: *dummies* para domicílios urbanos; cujas pessoas de referência eram homens; com membro(s) com ensino médio completo; e índice socioeconômico do domicílio. O índice foi construído a partir da análise de componentes principais de variáveis relacionadas às características do domicílio (presença ou não de geladeira, máquina de lavar, televisão, banheiro exclusivo, esgotamento sanitário, coleta de lixo, água de rede geral de distribuição, densidade de mais de dois moradores por dormitório; variável contínua de moradores por cômodos) e padronizado para ter média igual a zero em todos os anos.

O *hot deck* para imputar rendimentos foi estratificado por Unidade da Federação, área censitária, maior nível educacional entre os moradores e número de moradores por cômodo.

Os resultados confirmam que Censos e PNADs superestimam a quantidade de indivíduos com renda domiciliar *per capita* zero. No Censo 2000, 50% das pessoas com renda *per capita* igual a zero não tinham perfil de pobreza; em 2010, já eram 80%. Nas PNADs, esse percentual cresce de 55%, em 1995, para quase 85%, em 2011.

Os dois grupos identificados entre os domicílios com renda igual a zero são muito distintos. O índice socioeconômico para os pobres extremos oscila, nos Censos e nas PNADs, entre -1,5 e -3,0; já o dos “não pobres” fica sempre pouco acima de zero, ou seja, da média geral do Brasil.

Seleção dos anos para comparação

Como não há PNADs em anos censitários, optou-se por criar PNADs sintéticas para 2000 e 2010, de modo que o nível e a distribuição dos rendimentos em cada caso representam a média entre a PNAD do ano anterior e a do posterior. Com isso, evitam-se escolhas arbitrárias entre 1999 ou 2001 e 2009 ou 2011.

Há diversas formas mais ou menos sofisticadas para construir essas PNADs sintéticas. Neste trabalho, optou-se 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 de forma a 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, seja se aplicadas aos cerca de 100 mil domicílios de cada PNAD, seja se geradas a partir da agregação da PNAD em 12 mil grupos.

A comparação com as POFs é simples: 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, 2002 e 2008.

Resumo

A aplicação desses procedimentos garante a maior comparabilidade possível entre as três fontes de dados. O maior interesse está em cotejar as versões originais e plenamente harmonizadas e observar se há convergência dos dados. Contudo, também interessa saber qual procedimento tem o maior impacto, 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 com o filtro de tratamento.

As comparações sempre relacionam bancos que receberam o mesmo tratamento: as PNADs com filtro amostral são cotejadas com os Censos com o filtro amostral e assim por diante. Como não é necessário aplicar o filtro conceitual às PNADs nem o de tratamento às POFs, nesses casos foram usados os dados originais.

A comparação entre Censos, PNADs e POFs

Divergências nos dados originais

Para dimensionar as discrepâncias entre os dados originais dos Censos, das PNADs e das POFs, a Tabela 2 exibe informações sobre renda média, desigualdade e extrema pobreza em cada pesquisa.

A comparação entre Censos e PNADs revela números muito diferentes, com aumento da divergência ao longo do tempo. A renda *per capita* nos Censos era mais elevada em 2000, crescendo de forma mais rápida até 2010, enquanto a desigualdade e a extrema pobreza eram mais altas e caíram menos do que nas PNADs. A diferença é particularmente grande para o coeficiente de Gini, que permanece no Censo 2010 em patamar superior ao registrado pela PNAD há uma década. Essas diferenças não parecem resultar da 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.

Em consonância com resultados anteriores (BARROS; CURY; ULYSSEA, 2006), a comparação entre PNADs e POFs revela que o nível de renda é mais alto nas POFs, mas

os níveis de desigualdade são relativamente próximos. Os dois aspectos – renda e desigualdade – apresentam comportamentos distintos ao longo do tempo: enquanto a renda nas POFs descolou-se ainda mais das PNADs, a queda na desigualdade teve intensidade semelhante nas duas. Já a extrema pobreza parte de patamar semelhante nas duas pesquisas, mas cai mais rapidamente nas POFs.

TABELA 2
Renda domiciliar *per capita* média, índice de Gini e extrema pobreza nos Censos, PNADs e POFs
Brasil – 2000-2010

Indicadores	Comparação: PNADs e Censos					
	PNAD 2000	Censo 2000	Diferença (%)	PNAD 2010	Censo 2010	Diferença (%)
Renda <i>per capita</i> (em reais de 2011)	571	629	10,0	732	828	13,0
Gini (x 100)	59,3	64,6	9,0	53,4	60,9	14,0
Extrema pobreza (%)	9,5	13,5	4,0 p.p.	4,4	8,9	4,5 p.p.
Indicadores	Comparação: PNADs e POFs					
	PNAD 2002	POF 2002	Diferença (%)	PNAD 2008	POF 2008	Diferença (%)
Renda <i>per capita</i> (em reais de 2011)	576	709	23,0	691	941	36,0
Gini (x 100)	58,7	61,2	4,0	54,4	56,0	3,0
Extrema pobreza (%)	7,9	8,2	0,3 p.p.	4,5	2,3	-2,2 p.p.

Fonte: IBGE. PNADs 1999, 2001, 2002, 2008, 2009 e 2011; Censos 2000 e 2010; POFs 2002-2003 e 2008-2009.

Nota: Para renda e Gini, usa-se a diferença relativa e, para a extrema pobreza, emprega-se a diferença absoluta (em pontos percentuais).

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 do que entre as originais.

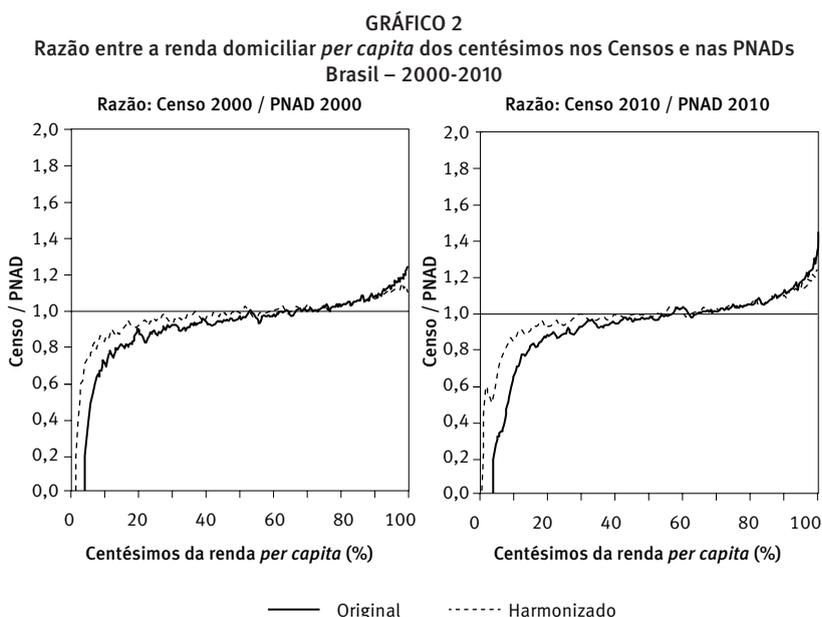
Os Gráficos 2 e 3 proporcionam comparações visuais: o primeiro mostra, para 2000 e 2010, a razão entre a renda domiciliar *per capita* média de cada centésimo nos Censos e nas PNADs tanto para os dados originais quanto para os plenamente harmonizados; o segundo faz o mesmo para as POFs. Em ambos, se as distribuições fossem idênticas às das PNADs de referência, os valores das curvas seriam constantes e iguais a 1.

Nos Censos, as divergências originais estão nos extremos: os pobres são muito mais pobres e os ricos são muito mais ricos do que nas PNADs. A harmonização promove convergência, em especial na cauda inferior. Nos dados originais, as razões entre Censos e PNADs se aproximam da unidade apenas quando se chega perto da mediana; nos harmonizados isso ocorre antes. Por exemplo, nos dados originais, a renda média dos 20% mais pobres nos Censos era apenas 70% da renda média desse grupo nas PNADs; já nos dados harmonizados, esse percentual sobe para cerca de 85%. No outro extremo, a renda média do centésimo mais rico nos dados originais dos Censos era 41% e 67%

maior do que nas PNADs 2000 e 2010; nos dados harmonizados essas porcentagens caem para 3% e 14%.⁵

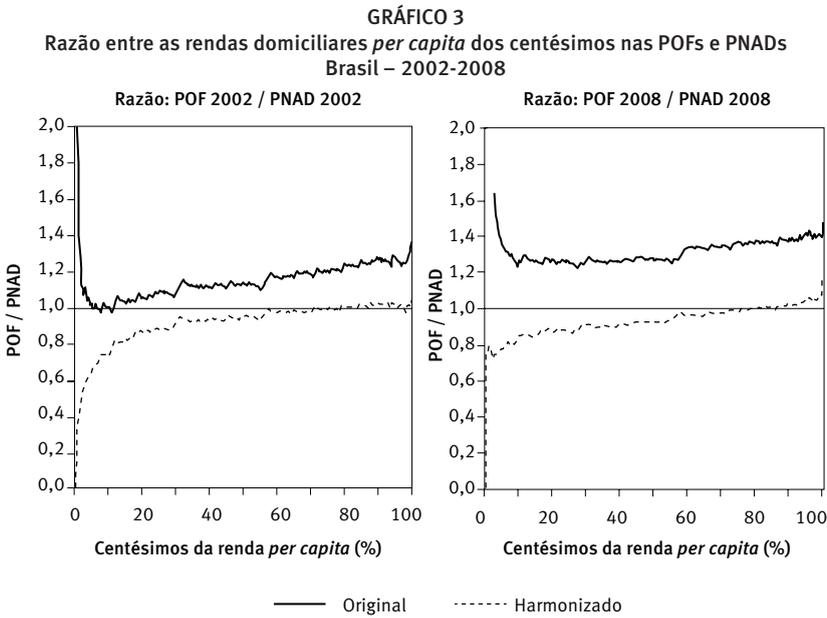
A comparação entre POFs e PNADs também mostra convergência após a harmonização dos dados. Nesse caso, a harmonização afeta mais o nível do que a desigualdade dos rendimentos. Com exceção dos primeiros centésimos, as distribuições harmonizadas correm paralelas, mas abaixo das originais e mais próximas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

Naturalmente, em nenhum dos dois casos a convergência é perfeita. Mesmo nos dados harmonizados, as PNADs permanecem com rendas mais elevadas para os mais pobres e com rendas mais baixas para os mais ricos, em especial na comparação com os Censos. De todo modo, a magnitude das diferenças diminui perceptivelmente.



Fonte: IBGE. PNADs 1999, 2001, 2009 e 2011; Censos 2000 e 2010.

⁵ A razão no topo após a harmonização parece mais elevada no gráfico do que esses números sugerem, pois o 1% mais rico – que concentra parcela expressiva da renda nacional – tem, por definição, pouco destaque visual quando se mostra a distribuição completa. Infelizmente, caso os gráficos fossem sensíveis à distribuição desigual da renda, as diferenças na cauda inferior da distribuição ficariam imperceptíveis.



Fonte: IBGE. PNADs 2002 e 2008; POFs 2002-2003 e 2008-2009.

Podemos quantificar a convergência pela estatística de Kolmogorov-Smirnov (K-S), que permite calcular a proximidade entre distribuições ao comparar as funções de distribuição empíricas acumuladas de duas amostras de tamanhos n e m . A estatística K-S é igual ao maior valor absoluto da diferença entre as duas funções de distribuição empíricas ao longo de x e varia entre 0 (distribuições idênticas) e 1 (sem sobreposição de valores entre elas):

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

Na Tabela 3, a aplicação da estatística 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 mais similares.⁶

TABELA 3
Estatísticas de Kolmogorov-Smirnov para rendas domiciliares *per capita* originais e harmonizadas:
comparações entre Censos e PNADs e entre POFs e PNADs
Brasil – 2000-2010

Comparação	Originais	Harmonizadas	Diferença
Censo 2000 <i>versus</i> PNAD 2000	0,049	0,033	-0,017
Censo 2010 <i>versus</i> PNAD 2010	0,048	0,029	-0,019
POF 2002 <i>versus</i> PNAD 2002	0,067	0,050	-0,017
POF 2008 <i>versus</i> PNAD 2008	0,117	0,047	-0,071

Fonte: IBGE. PNADs 1999, 2001, 2002, 2008, 2009 e 2011; Censos 2000 e 2010; POFs 2002-2003 e 2008-2009.

⁶ O teste de K-S não é adequado quando os dados são provenientes de amostras complexas. Por isso, apenas a estatística é apresentada como referência, sem ênfase na parte inferencial e em testes de significância. O autor agradece a um dos pareceristas anônimos por levantar este ponto.

A Tabela 4 decompõe a variação na estatística de K-S para mostrar as contribuições dos três filtros para a convergência. No caso dos Censos, o filtro de tratamento tem o maior efeito, inclusive com importância crescente entre 2000 e 2010, refletindo as mudanças operacionais e conceituais entre os Censos que resultaram em maior percentual de indivíduos com renda *per capita* igual a zero nos dados originais com relação às PNADs. Na comparação entre PNADs e POFs, os resultados são mais estáveis: o que importa mesmo é o filtro conceitual.

TABELA 4
Decomposição da variação das estatísticas de Kolmogorov-Smirnov entre rendas domiciliares *per capita* originais e harmonizadas: comparações entre Censos e PNADs e entre POFs e PNADs
Brasil – 2000-2010

Comparação	Amostral	Conceitual	Tratamento	Harmonização completa
Contribuições absolutas				
Censo 2000 <i>versus</i> PNAD 2000	-0.007	-	-0.010	-0.017
Censo 2010 <i>versus</i> PNAD 2010	-0.004	-	-0.016	-0.019
POF 2002 <i>versus</i> PNAD 2002	0.000	-0.018	0.000	-0.017
POF 2008 <i>versus</i> PNAD 2008	0.000	-0.070	0.000	-0.070
Contribuições relativas (%)				
Censo 2000 <i>versus</i> PNAD 2000	40,0	-	60,0	100,0
Censo 2010 <i>versus</i> PNAD 2010	19,0	-	81,0	100,0
POF 2002 <i>versus</i> PNAD 2002	-2,0	105,0	-2,0	100,0
POF 2008 <i>versus</i> PNAD 2008	0,0	100,0	0,0	100,0

Fonte: IBGE. PNADs 1999, 2001, 2002, 2008, 2009 e 2011; Censos 2000 e 2010; POFs 2002-2003 e 2008-2009.

Nota: As diferenças entre a coluna “harmonização completa” e a soma das demais decorrem de arredondamentos.

Os rendimentos médios

A harmonização promove maior convergência na renda domiciliar *per capita* nas POFs do que nos Censos, como se vê na Tabela 5. No primeiro caso, há convergência quase perfeita: se, nos dados originais, tanto o nível quanto o crescimento da renda eram maiores nas POFs do que nas PNADs, nos dados harmonizados ambos tornam-se substantivamente idênticos nas duas pesquisas.

Nos Censos, o efeito da harmonização é menor, mas não desprezível: a renda média nos dados originais era entre 10% e 13% maior do que nas PNADs, mas nos dados harmonizados esse percentual cai para algo em torno de 5%. As taxas de crescimento da renda também se tornam levemente mais próximas. A coleta mais agregada dos dados de renda no Censo 2010 não gerou viés de subestimação dos rendimentos médios, seja nos dados originais ou harmonizados.

TABELA 5
Renda domiciliar *per capita* nos dados originais e harmonizados: comparações entre Censos e PNADs e entre POFs e PNADs
Brasil – 2000-2010

Comparação	Dados originais			Dados plenamente harmonizados		
	2000	2010	Variação anual (%)	2000	2010	Variação anual (%)
Censo (A)	629	828	2,5	616	816	2,7
PNAD (B)	571	732	2,8	588	767	2,8
A/B (%)	110,0	113,0	-	105,0	106,0	-

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 (%)	123,0	136,0	-	99,0	100,0	-

Fonte: IBGE. PNADs 1999, 2001, 2002, 2008, 2009 e 2011; Censos 2000 e 2010; POFs 2002-2003 e 2008-2009.

A Tabela 6 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 para alterar essas razões. A quarta coluna mostra o efeito combinado dos três (harmonização completa). As duas últimas apresentam as diferenças após a harmonização e as diferenças originais.

Na comparação entre Censos e PNADs, verifica-se que o filtro amostral é o responsável pelas mudanças. Mais especificamente, a eliminação do milésimo mais rico nos Censos derruba a renda média o suficiente para aproximar essa pesquisa das PNADs. Na comparação entre POFs e PNADs, o filtro conceitual é o grande responsável por fazer evaporar a diferença original entre as pesquisas. A diferença original decorre exclusivamente do fato de que as POFs aplicam um conceito de renda mais abrangente.

Em conjunto, os dois resultados ajudam a explicar boa parte das diferenças entre as pesquisas, mas também reforçam a ideia de que as PNADs de fato subestimam o nível de renda no país, para além das limitações que afetam todas as pesquisas domiciliares. No limite, pode-se especular que uma pesquisa com uma amostra como a do Censo e com questionário abrangente como a POF revelaria níveis de renda ainda mais altos. De qualquer modo, há grande concordância entre Censos, PNADs e POFs no que diz respeito à taxa de crescimento da renda.

TABELA 6
Decomposição das razões entre a renda domiciliar *per capita* média: comparações entre Censos e PNADs e entre POFs e PNADs
Brasil – 2000-2010

Comparação	Contribuição de cada filtro (p.p.)				Razão entre dados harmonizados	Razão entre dados originais
	Amostral	Conceitual	Tratamento	Harmonização completa		
Censo 2000 <i>versus</i> PNAD 2000	-4	-	-1	-5	105	110
Censo 2010 <i>versus</i> PNAD 2010	-6	-	0	-7	106	113
POF 2002 <i>versus</i> PNAD 2002	1	-22	-3	-24	99	123
POF 2008 <i>versus</i> PNAD 2008	0	-33	-3	-36	100	136

Fonte: IBGE. PNADs 1999, 2001, 2002, 2008, 2009 e 2011; Censos 2000 e 2010; POFs 2002-2003 e 2008-2009.

Nota: As diferenças entre a coluna "harmonização completa" e a soma das demais decorrem de arredondamentos.

O coeficiente de Gini

A Tabela 7 replica para o coeficiente de Gini a análise da Tabela 5. De novo, a harmonização aproxima bastante os resultados, embora não elimine totalmente as discrepâncias entre eles. Originalmente, o nível de desigualdade nos Censos era maior em 2000 e a queda da desigualdade, menor. Depois da harmonização, tanto os níveis quanto o ritmo de queda ficam mais parecidos nas duas pesquisas.

Os efeitos da harmonização sobre as POFs são mais ambíguos. Originalmente, os coeficientes de Gini já eram muito próximos, com níveis e ritmo de queda superiores nas POFs. Com a harmonização, o coeficiente de 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 do que nas PNADs.

TABELA 7
Coefficiente de Gini da renda domiciliar *per capita* nos dados originais e harmonizados: comparações entre Censos e PNADs e entre POFs e PNADs
Brasil – 2000-2010

Comparação	Dados originais			Dados plenamente harmonizados		
	2000 (x 100)	2010 (x 100)	Varição anual (Δ Gini)	2000 (x 100)	2010 (x 100)	Varição anual (Δ Gini)
Censo (A)	64,6	60,9	-0,4	61,6	56,9	-0,5
PNAD (B)	59,3	53,4	-0,6	59,3	53,3	-0,6
A/B (%)	109,0	114,0	-	104,0	107,0	-

Comparação	Dados originais			Dados plenamente harmonizados		
	2000 (x100)	2010 (x100)	Varição anual (Δ Gini)	2000 (x100)	2010 (x100)	Varição anual (Δ Gini)
POF (C)	61,2	56,0	-0,9	60,7	57,2	-0,6
PNAD (D)	58,7	54,4	-0,7	58,7	54,2	-0,7
C/D (%)	104,0	103,0	-	103,0	106,0	-

Fonte: IBGE. PNADs 1999, 2001, 2002, 2008, 2009 e 2011; Censos 2000 e 2010; POFs 2002-2003 e 2008-2009.

A Tabela 8 mostra o efeito de cada filtro sobre as razões entre os coeficientes de Gini registrados nos Censos e PNADs e nas POFs e PNADs. No primeiro caso, o filtro amostral é um pouco mais importante do que o de tratamento, o que reflete o maior peso dos 0,1% mais ricos nos Censos do que nas PNADs. No segundo caso, a harmonização não muda muito a razão entre os coeficientes de Gini, mas o filtro conceitual é o que tem mais peso.

De modo mais geral, é possível chegar a conclusões semelhantes às anteriores: a comparação entre as três bases harmonizadas apresenta resultados bem mais próximos do que os observados originalmente, sem eliminar totalmente as diferenças. Mais uma vez, verifica-se uma leve tendência de subestimação do nível de desigualdade nas PNADs em comparação com as outras duas pesquisas. Ainda assim, há grande convergência quanto ao ritmo de queda da desigualdade nos Censos, PNADs e POFs.

TABELA 8
Decomposição das razões dos coeficientes de Gini da renda domiciliar *per capita*: comparações entre Censos e PNADs e entre POFs e PNADs
Brasil – 2000-2010

Comparação	Contribuição de cada filtro (p.p.)				Razão entre dados harmonizados	Razão entre dados originais
	Amostral	Conceitual	Tratamento	Harmonização completa		
Censo 2000 <i>versus</i> PNAD 2000	-3	-2	-	-5	104	109
Censo 2010 <i>versus</i> PNAD 2010	-5	-2	-	-7	107	114
POF 2002 <i>versus</i> PNAD 2002	-1	0	0	-1	103	104
POF 2008 <i>versus</i> PNAD 2008	-1	0	3	2	106	103

Fonte: IBGE. PNADs 1999, 2001, 2002, 2008, 2009 e 2011; Censos 2000 e 2010; POFs 2002-2003 e 2008-2009.

Nota: As diferenças entre a coluna "harmonização completa" e a soma das demais decorrem de arredondamentos.

A extrema pobreza

Os números relativos à extrema pobreza, apresentados na Tabela 9, confirmam as tendências anteriores: após a harmonização, as porcentagens nas três fontes de dados tornam-se mais próximas, pelo menos entre Censos e PNADs. Originalmente, a principal diferença nesse caso estava nos níveis, não na trajetória de queda da extrema pobreza. Após a harmonização, a diferença nos níveis cai mais ou menos pela metade: de 4,0 e 4,5 p.p. para 2,3 e 2,1 p.p. em 2000 e 2010, respectivamente.

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

TABELA 9
Incidência da extrema pobreza nos dados originais e harmonizados: comparações entre Censos e PNADs e entre POFs e PNADs
Brasil – 2000-2010

Comparação	Dados originais			Dados plenamente harmonizados		
	2000 (%)	2010 (%)	Variação anual (p.p.)	2000 (%)	2010 (%)	Variação anual (p.p.)
Censo (A)	13,5	8,94,4	-0,5	10,8	5,5	-0,5
PNAD (B)	9,5	4,4	-0,5	8,5	3,4	-0,5
A/B (p.p.)	4,0	4,5	-	2,3	2,1	-

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: IBGE. PNADs 1999, 2001, 2002, 2008, 2009 e 2011; Censos 2000 e 2010; POFs 2002-2003 e 2008-2009.

A Tabela 10 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, a importância do filtro de tratamento chama a atenção: entre 30% (em 2000) e 43% (em 2010) das diferenças com relação às PNADs decorrem de questões de coleta e tratamento dos dados que são passíveis de harmonização. A explicação remete ao percentual elevado de indivíduos com renda *per capita* igual a zero nos Censos, em especial em 2010, com relação às PNADs.

A comparação entre POFs e PNADs revela que praticamente todas as mudanças ocorrem por causa do filtro conceitual. Mais especificamente, a exclusão dos rendimentos não monetários aumenta significativamente a incidência da extrema pobreza. Como as PNADs obviamente não captam esse tipo de rendimentos, pode-se concluir que a concordância entre pesquisas observada nos dados originais em 2002 era mera coincidência.

Em boa medida, os resultados para extrema pobreza corroboram as conclusões das seções anteriores: por um lado, as diferenças mais gritantes diminuem sensivelmente, como na comparação entre censos e PNADs; por outro, a harmonização não promove convergência perfeita entre as pesquisas e, mais um vez, os números das PNADs destoam um pouco das outras duas, indicando percentuais de extrema pobreza um pouco inferiores aos Censos e POFs.

TABELA 10
Decomposição das diferenças na incidência de extrema pobreza: comparações entre Censos e PNADs e entre POFs e PNADs
Brasil – 2000-2010

Comparação	Contribuição de cada filtro (p.p.)				Diferença (p.p.)	
	Amostral	Conceitual	Tratamento	Harmonização completa	Dados harmonizados	Dados originais
Censo 2000 <i>versus</i> PNAD 2000	-0,6	-1,2	-	-1,7	2,3	4,0
Censo 2010 <i>versus</i> PNAD 2010	-0,4	-1,9	-	-2,4	2,1	4,5
POF 2002 <i>versus</i> PNAD 2002	-0,4	0,8	3,7	4,1	4,4	0,3
POF 2008 <i>versus</i> PNAD 2008	-0,1	0,8	3,6	4,2	2,0	-2,2

Fonte: IBGE. PNADs 1999, 2001, 2002, 2008, 2009 e 2011; Censos 2000 e 2010; POFs 2002-2003 e 2008-2009.

Nota: As diferenças entre a coluna “harmonização completa” e a soma das demais decorrem de arredondamentos.

Conclusão

Os Censos, as PNADs e as POFs são as três principais fontes de dados sobre distribuição de renda no Brasil, mas, muitas vezes, os resultados obtidos em cada uma delas parecem incompatíveis entre si. O objetivo aqui foi entender essas diferenças, tomando as PNADs como referência.

Para isso, o texto partiu de uma breve caracterização das pesquisas e, em seguida, procurou harmonizar suas principais diferenças, agregando-as em três grandes grupos: diferenças amostrais; diferenças conceituais; e diferenças na coleta e tratamento dos dados. Como em qualquer harmonização *ex post*, os ajustes propostos são imperfeitos, mas necessários para tornar a comparação mais apropriada.

O filtro amostral é simples, com duas etapas: exclusão dos moradores do Norte rural (exceto Tocantins), para compatibilizar os Censos, as POFs e as PNADs 2004-2011 com as PNADs anteriores; e exclusão do milésimo da população com maior renda domiciliar *per capita* dos Censos, público que não é captado em amostras menores.

Já o filtro conceitual é mais problemático, porque não há como resolver perfeitamente as discrepâncias entre as pesquisas. Com isso, esse filtro limitou-se a manter nas POFs apenas os rendimentos também captados pelos Censos e pelas PNADs, excluindo, principalmente, rendimentos não monetários e alguns relacionados ao trabalho formal.

O filtro de tratamento também foi apenas parcial. No caso dos indivíduos com rendimentos ignorados, foi utilizado nas PNADs um método de imputação por *hot deck* parecido com o que o IBGE aplica aos Censos e às POFs. No caso dos domicílios com renda igual a zero, foi empregado modelo de análise de conglomerados nos Censos e nas PNADs para dividi-los entre aqueles com perfil de pobreza extrema e aqueles com perfil não pobre. Em seguida, estes últimos tiveram rendimentos imputados por *hot deck*.

A etapa seguinte foi a comparação entre dados originais e harmonizados dos Censos, das PNADs e das POFs. A expectativa era de que, caso ocorresse convergência, então as discrepâncias em relação a Censos e às POFs estariam explicadas e seria lícito supor que as PNADs captam bem aquilo que se propõem a captar.

A análise empírica confirma isso em boa medida. Após a harmonização, de modo geral, os resultados dos Censos, das PNADs e das POFs tornaram-se bem mais próximos do que os observados originalmente, sem a harmonização proposta.

Os filtros são particularmente úteis na comparação entre Censos e PNADs, o que era de se esperar, pois são pesquisas mais semelhantes entre si: no que diz respeito aos níveis de renda média, desigualdade e extrema pobreza, nos dados harmonizados as discrepâncias caem em geral pelo menos à metade. Nas três dimensões, também há convergência quanto às tendências ao longo do tempo, que já eram originalmente próximas e se tornam ainda mais semelhantes.

Os efeitos da harmonização sobre a comparação entre POFs e PNADs são mais ambíguos. Por um lado, as grandes discrepâncias dos dados originais – relativas tanto ao nível quanto ao ritmo de crescimento da renda média – desaparecem por completo, dando lugar à plena convergência entre pesquisas nos dados harmonizados. Também há alguma convergência, em grau bem menor, quanto ao nível e ao ritmo de queda da desigualdade. Por outro lado, a harmonização fez aumentar as diferenças na incidência da pobreza extrema entre POFs e PNADs, que, por pura coincidência, eram mínimas nos dados originais.

Naturalmente, a convergência entre as pesquisas não é perfeita e, mais ainda, as discrepâncias remanescentes seguem mais ou menos o mesmo padrão: em comparação com os Censos e as POFs, as PNADs tendem a “contrair” um pouco a distribuição de renda, superestimando a renda dos mais pobres e subestimando a dos mais ricos.

Nenhuma pesquisa jamais será capaz de retratar com perfeição o que é o “mundo real”. O importante é entender os limites de cada pesquisa e quão compatíveis entre si elas são. Nesse sentido, a convergência após a harmonização é encorajadora: o retrato que emerge nos Censos, nas PNADs e nas POFs é relativamente coerente, confirmando que algumas diferenças sistemáticas entre planos amostrais, conceitos e tratamento dos dados são responsáveis por boa parte das discrepâncias observadas nos dados originais.

Referências

BARROS, R. P. de; 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. (Org.). **Desigualdade de renda no Brasil: uma análise da queda recente**. Brasília: Ipea, 2006.

BARROS, R. P. de; FOGUEL, M. N.; ULYSSEA, G. (Org.). **Desigualdade de renda no Brasil: uma análise da queda recente**. Brasília: Ipea, 2006.

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

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.

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 contemporâneas**. Brasília: Ipea, 2007.

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, 2007 (PSID technical series, n. 07-01).

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

_____. 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).

_____. **Pesquisa Nacional por Amostra de Domicílios – Síntese de Indicadores 2009**. Rio de Janeiro: IBGE, 2010a.

_____. **Pesquisa de Orçamentos Familiares 2008/2009: despesas, rendimentos e condições de vida**. Rio de Janeiro: IBGE, 2010b.

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

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

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.

_____. 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. VER; MOFFITT, R.; CITRO, C. F. (Eds.). **Studies of welfare populations: data collection and research issues**. Washington, D.C.: The National Academy Press, 2001.

MEDEIROS, M.; SOUZA, P. H. G. F.; CASTRO, F. A. **O topo da distribuição de renda no Brasil: primeiras estimativas com dados tributários e comparação com pesquisas domiciliares, 2006/2012**. Social Science Research Network Working Paper, 2014a. Disponível em: <<http://ssrn.com/abstract=2479685>>.

_____. **A estabilidade da desigualdade de renda no Brasil, 2006 a 2012: estimativa com**

dados do imposto de renda e pesquisas domiciliares. Social Science Research Network Working Paper, 2014b. Disponível em: <<http://ssrn.com/abstract=2493877>>.

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.

MEYER, B.; MOK, W.; SULLIVAN, J. *The under-reporting of transfers in household surveys: its nature and consequences*. Cambridge, Mass.: National Bureau of Economic Research, 2009 (NBER working paper, n. 15181).

MOORE, J.; STINSON, L.; WELNIAK JR, 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. da S. O IBGE e as pesquisas populacionais. *Revista Brasileira de Estudos de População*, v. 22, n. 2, p. 291-302, 2005.

ONU. *Principles and recommendations for population and housing censuses*. Nova York: Organização das Nações Unidas, 2008.

OSORIO, R. G. et al. *Perfil da pobreza no Brasil e sua evolução no período 2004-2009*. Brasília: Ipea, 2011 (Texto para discussão, n. 1647).

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. do 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, S. Análise de bem-estar e decomposição por fatores da queda na desigualdade entre 1995 e 2004. *Econômica*, v. 8, n. 1, p. 83-115, 2006.

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. *Revista Brasileira de Estudos de População*, v. 30, n. 1, p. 299-315, 2013.

VAZ, F. M. *Escalas de equivalência e demanda do consumidor*. Tese (Doutorado em Economia). Brasília: Universidade de Brasília, 2012.

WEINBERG, D. et al. Fifty years of U.S. income data from the Current Population Survey: alternatives, trends, and quality. *American Economic Review*, v. 89, n. 2, p. 18-22, 1999.

WHEATON, L. Underreporting of means-tested transfer programs in the CPS and SIPP. *2007 Proceedings of the American Statistical Association*, p. 3622-3629, 2007.

Sobre o autor

Pedro Herculano Guimarães Ferreira de Souza é mestre em Sociologia pelo Instituto Universitário de Pesquisas do Rio de Janeiro – IUPERJ e doutorando em Sociologia pela Universidade de Brasília, bolsista na University of California, Economics Department, Berkeley, USA.

Endereço para correspondência

2210 Curtis St. Berkeley
94702 – Califórnia, Estados Unidos

Abstract

Income distribution according to Brazilian household surveys: harmonization and comparison of Census, PNAD and POF data

The aim of this paper is to document and explain the differences in income distribution in three Brazilian household surveys: the Demographic Census, the National Household Sample Survey (PNAD - Pesquisa Nacional por Amostra de Domicílios) and the Family Budgets Survey (POF - Pesquisa de Orçamentos Familiares). The main hypothesis is that it is possible to achieve great convergence of results in the aforementioned surveys with ex post harmonization procedures that minimize, as far as possible, discrepancies in sampling design, in concepts, and in data collection and treatment. The results confirm, to a large extent, this hypothesis: in general, harmonization approximates the three surveys and significantly reduces the major discrepancies between income distributions, in particular concerning Census vs. PNAD comparisons. Although, in some cases, differences persist in the levels of income, inequality and poverty, their tendencies over time become remarkably similar in the three surveys. Finally, it is observed that the remaining discrepancies follow a pattern: even after harmonization, income distribution in PNADs tends to be a little more egalitarian than in Censuses and in POFs, that is, the poorest families have higher incomes and the richest families have lower incomes.

Keywords: Household surveys. Income measurement. Income distribution. Inequality. Poverty.

Resumen

La distribución del ingreso en las encuestas de hogares brasileñas: armonización y comparación entre los censos, las PNAD y las POF

El objetivo de este trabajo es documentar y explicar las diferencias en la distribución del ingreso que surgen del censo demográfico, la Pesquisa Nacional por Amostra de Domicílios (PNAD) y la Pesquisa de Orçamentos Familiares (POF). La hipótesis principal que se propone es que es posible promover una gran convergencia de los resultados entre las tres encuestas con los procedimientos de armonización ex post, que compatibilizan, en la medida de lo posible, las diferencias muestrales, conceptuales y de recogida y tratamiento de los datos. Los resultados confirman en buena medida esta hipótesis: de modo general, la armonización aproxima las tres encuestas y reduce significativamente las mayores discrepancias entre las distribuciones del ingreso, especialmente cuando se comparan el censo y la PNAD. Aunque en algunos casos persisten las diferencias en los niveles del ingreso, la desigualdad y la pobreza, su evolución en el tiempo llega a ser muy similar en las tres encuestas. Por último, también se observó que las discrepancias remanentes siguen un patrón, es decir, incluso después de la armonización, la distribución del ingreso en la PNAD tiende a ser un poco más igualitaria que en el censo y en la POF: en ese instrumento, los ingresos de los más pobres son más altos y los de los más ricos, más bajos.

Palabras clave: Encuestas de hogares. Medición de los ingresos. Distribución del ingreso. Desigualdad. Pobreza.

Recebido para publicação em 09/11/2014

Aceito para publicação em 24/05/2015