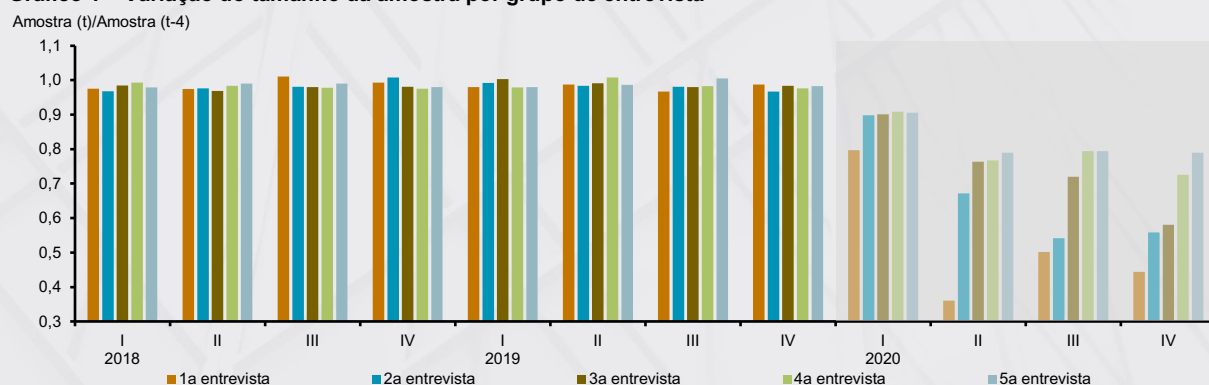


Estimativas da Pesquisa Nacional por Amostra de Domicílios Contínua ajustadas para a redução da taxa de respostas na pandemia

A Pesquisa Nacional por Amostra de Domicílios Contínua (PNAD Contínua) tem passado por desafios inerentes à mudança na forma de coleta de informações de presencial para telefone em decorrência da pandemia.¹ A redução da taxa de respostas, especialmente nos grupos com indivíduos nas primeiras entrevistas² a partir do segundo trimestre de 2020, destaca-se como uma das maiores adversidades. Esse fato, apontado por analistas de mercado de trabalho e pelo próprio Instituto Brasileiro de Geografia e Estatística (IBGE)³, é ilustrado no Gráfico 1.

Gráfico 1 – Variação do tamanho da amostra por grupo de entrevista



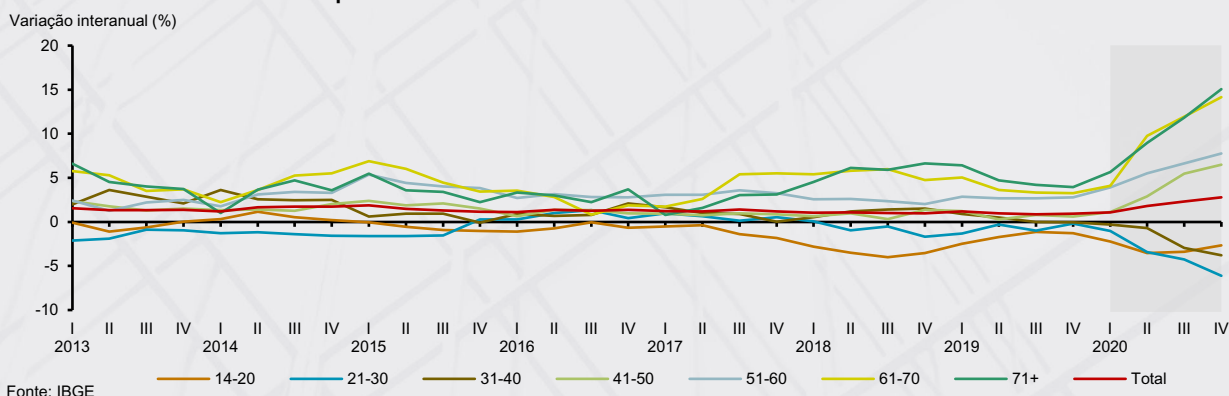
Este boxe analisa os efeitos da redução da taxa de respostas da PNAD Contínua em importantes indicadores econômicos da pesquisa. Uma motivação básica foi a aceleração da população em idade de trabalhar⁴ (PIT) no decorrer da pandemia, em desalinhamento com as projeções populacionais do IBGE (Gráfico 2). Adicionalmente, houve aumento do crescimento da população de 40 anos ou mais e intensificação da redução da população com menos de 40 anos (Gráfico 3).

Gráfico 2 – Crescimento da PIT



- 1/ Desafios correlatos são enfrentados em todo o mundo. Para mais detalhes, vide o Relatório da Organização Internacional do Trabalho sobre o tema (ILO, 2020).
- 2/ O esquema de rotação amostral da pesquisa consiste no acompanhamento dos entrevistados por cinco trimestres consecutivos através de grupos de entrevista (1 a 5).
- 3/ Queda na taxa de respostas também é referida como redução do número de entrevistas realizadas na amostra. Vide, por exemplo, as Notas Técnicas 8/2020 (“Informações referentes à divulgação dos dados do 2º trimestre de 2020”) e 2/2021 (“Sobre o processo de ponderação da PNAD Contínua”) da PNAD Contínua do IBGE e a Carta de Conjuntura do Instituto de Pesquisa Econômica Aplicada (Ipea) Número 50 – Nota de Conjuntura 22 – 1º trimestre de 2021 (“A redução no número de entrevistas na PNAD Contínua durante a pandemia e sua influência para a evolução do emprego formal”).
- 4/ População com idade igual ou superior a 14 anos.

Gráfico 3 – Crescimento da PIT por faixa etária



É usual em pesquisas domiciliares que a estrutura de idade e sexo dos entrevistados seja diferente da encontrada na população, fenômeno que a literatura denomina “viés de disponibilidade”. Em particular na PNAD Contínua há maior disponibilidade de mulheres e indivíduos mais velhos – são encontrados com maior frequência – com o passar dos trimestres. Logo, em um mesmo trimestre, em geral, tende a haver mais mulheres e, especialmente, mais pessoas mais velhas nas amostras com indivíduos já entrevistados. Conforme ilustram os Gráficos 4 e 5, o viés de disponibilidade – especialmente de indivíduos mais velhos (com 40 anos ou mais) – aumentou de forma nítida a partir do segundo trimestre de 2020 em todos os grupos de entrevista, mas particularmente nos grupos de entrevista iniciados já na pandemia, em que o primeiro contato foi por telefone⁵. A mudança abrupta na proporção de mulheres e indivíduos mais velhos pode refletir algum viés de seleção da amostra relacionado, por exemplo, ao acesso a números de telefone ou à propensão de responder às tentativas de contato pelo IBGE.

Gráfico 4 – Proporção de pessoas com mais de 40 anos na amostra por grupo de entrevista

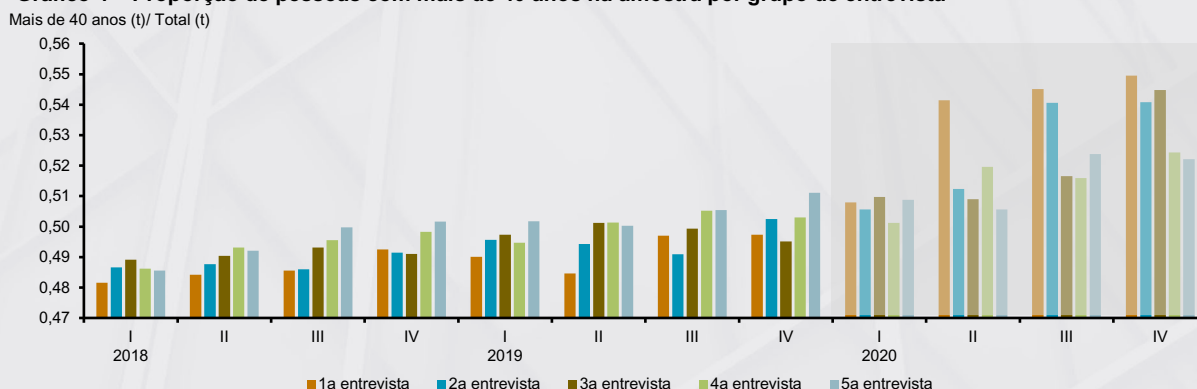
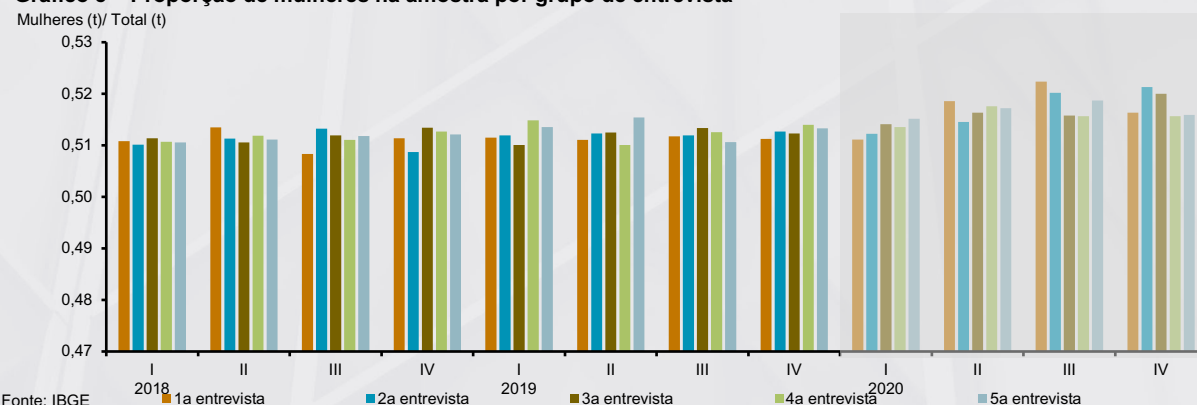


Gráfico 5 – Proporção de mulheres na amostra por grupo de entrevista



5/ Ou seja, no grupo de entrevista 1 no segundo trimestre de 2020, nos grupos de entrevista 1 e 2 no terceiro e nos grupos de entrevista 1 a 3 no quarto.

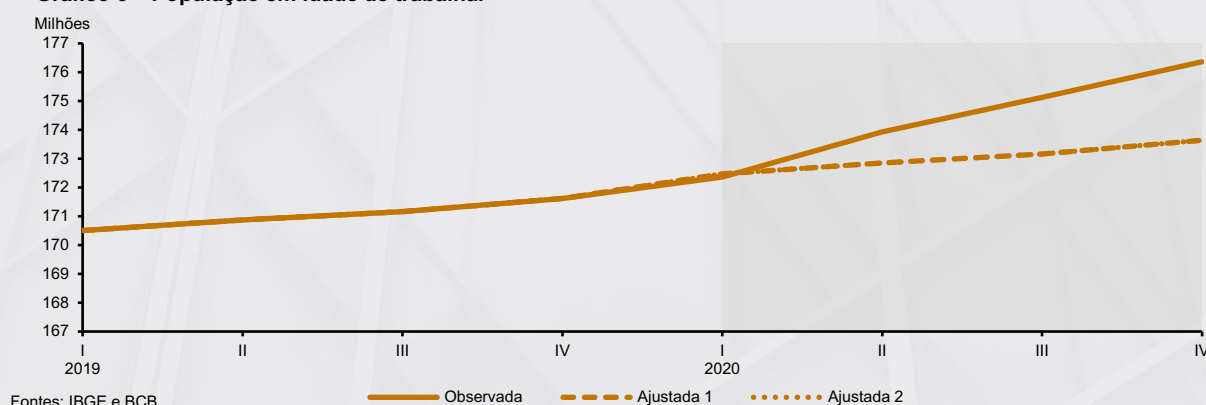
A PNAD Contínua utiliza a hipótese de perda amostral aleatória e aplica redistribuição dos pesos amostrais uniformemente dentre os indivíduos de uma mesma unidade primária de amostragem. A aplicação dessa hipótese em contexto de aumento do viés de disponibilidade pode explicar a aceleração do crescimento da PIT, na medida em que se interpretaria o aumento na proporção de indivíduos mais velhos como mudança demográfica subjacente à população.

O visível impacto da mudança na forma de coleta sobre estatísticas demográficas suscita interesse em avaliar em que medida o mesmo pode ocorrer nas estimativas das variáveis de mercado de trabalho, tais como população ocupada, taxa de desocupação e taxa de participação. Para avaliar essa questão, são aplicadas duas estratégias empíricas correlatas utilizando os microdados da PNAD Contínua trimestral. Ambas objetivam mitigar o viés de disponibilidade associado à redução da taxa de respostas utilizando as amostras de 2019 e a projeção populacional do IBGE de 2020 como referências para ajustar as estimativas da pesquisa.

Na primeira abordagem, todos os grupos de entrevista iniciados em 2020 são excluídos. Em seguida, as observações que continuam na amostra são estratificadas por sexo, faixa etária⁶ e UF. A motivação para essa exclusão se baseia em duas hipóteses: i) os grupos de entrevista iniciados em 2020 aumentaram o viés de disponibilidade da pesquisa; e ii) os indivíduos nos grupos de entrevista iniciados em 2019 são mais parecidos com aqueles das amostras trimestrais de 2020, caso não houvesse redução da taxa de respostas. Em seguida, calibra-se a população em idade de trabalhar através da razão entre a PIT da projeção populacional do IBGE e a PIT obtida pela exclusão citada acima, de modo a manter os totais populacionais por grupos de sexo x faixas etárias x UF de acordo com o projetado pelo IBGE.⁷

A segunda abordagem ajusta diretamente os pesos amostrais – de forma similar ao sugerido, e em implementação, pelo IBGE na Nota Técnica 2/2021 da PNAD Contínua – como procedimento para correção

Gráfico 6 – População em idade de trabalhar



6/ Foram definidas cinco faixas etárias: 14-24 anos, 25-39 anos, 40-54 anos, 55-69 anos e 70 ou mais anos. O resultado das estimativas é robusto a definições alternativas de faixas.

7/ A estimativa simulada de alguma variável X de mercado de trabalho tem a seguinte definição formal:

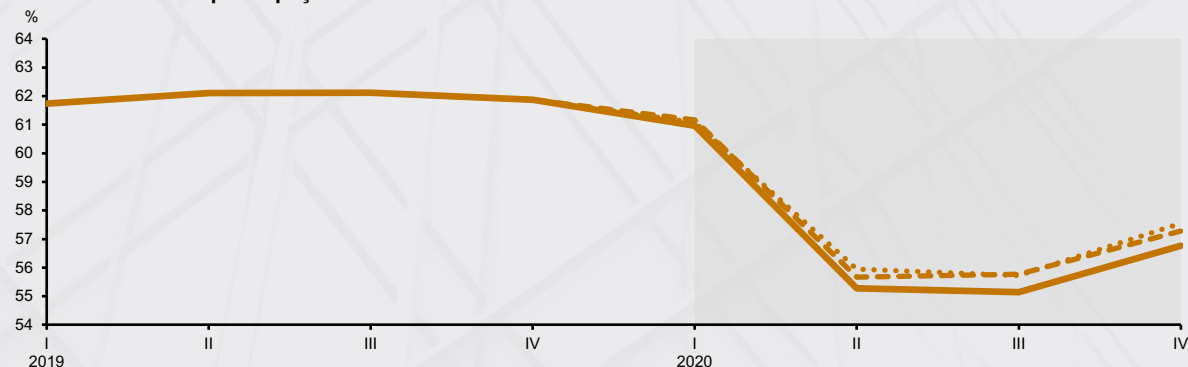
$$X_t^{s_1} = \sum_{G(g)} \sum_{A(a)} \sum_{U(u)} X_{g,a,u,t}^{s_1} \left(\frac{PIT_{g,a,u,t}^{IBGE}}{PIT_{g,a,u,t}^{s_1}} \right), \quad (1)$$

$$\text{onde } X_{g,a,u,t}^{s_1} = \sum_{i \in F(g,a,u,E(t))} w_{i,t} \mathbb{I}(X).$$

Ou seja, o total da variável X simulada (simulação 1: s_1) do sexo g , faixa etária a , UF u , no trimestre t é igual à soma do peso amostral (w) de todos os indivíduos i na condição da variável X (identificados pela função indicadora $\mathbb{I}(X)$) e que estão no conjunto F , que depende do sexo, faixa etária, UF e grupos de entrevista $E(t)$, que, por sua vez, dependerá do trimestre de 2020: $E(t) = \sum_{s=t-1}^{s=t+1} e_s$, em que e refere-se ao grupo de entrevista. Portanto, em I 2020 ($t = 1$) apenas os grupos de indivíduos na 2ª entrevista a 5ª entrevista são considerados, em II 2020, os grupos da 3ª a 5ª entrevistas e assim por diante até IV 2020, que considera apenas o grupo da 5ª entrevista (que foi iniciado – era grupo da 1ª entrevista – em IV 2019). Logo, todos grupos de entrevistas de 2020 foram iniciados em 2019. Note ainda que a definição de $PIT_{g,a,u,t}^{s_1}$ é a mesma de $X_{g,a,u,t}^{s_1}$ e que $PIT_{g,a,u,t}^{IBGE}$ é a PIT do trimestre t de 2020 que utiliza as taxas de crescimento anuais da projeção populacional do IBGE por sexo, faixa etária e UF aplicadas uniformemente às PITs dos quatro trimestres anteriores em 2019. O total agregado da variável X simulada, $X_t^{s_1}$, soma $X_{g,a,u,t}^{s_1}$ multiplicada pela razão das PITs por UF ($U(u)$), faixa etária ($A(a)$) e sexo ($G(g)$).

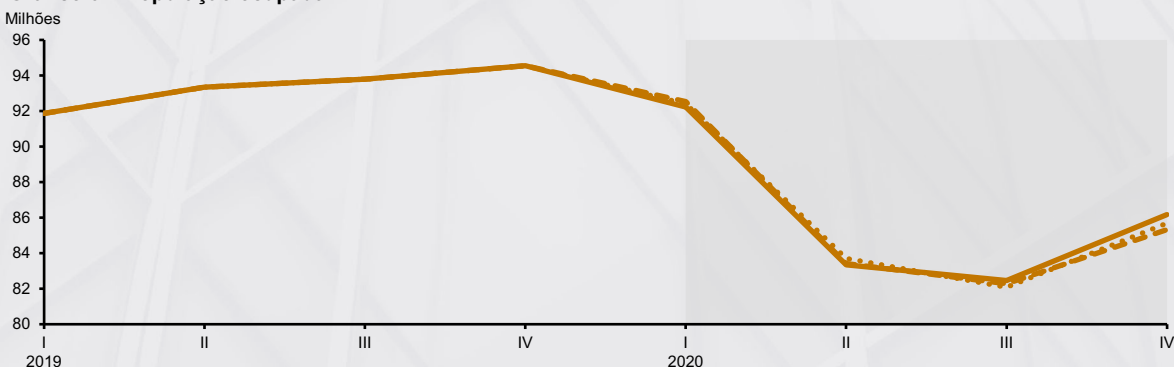
de possíveis vieses na pesquisa. Primeiramente são previstas as propensões (ou probabilidades) de resposta individuais às entrevistas da PNAD Contínua em cada trimestre de 2020 e para cada grupo de entrevista utilizando uma regressão logística com uma série de covariadas demográficas e geográficas com referência aos mesmos grupos de entrevista nos trimestres de 2019. A propensão de resposta ajusta o peso amostral individual em cada trimestre de 2020. Intuitivamente, o objetivo desse ajuste por propensão de resposta é, para cada grupo de entrevista, dar mais (menos) peso a indivíduos em cada trimestre de 2020 que são mais (menos) parecidos com indivíduos nas amostras trimestrais de 2019, segundo as características definidas pelas covariadas na regressão logística. Em seguida, os totais populacionais são calibrados de modo similar à primeira abordagem.⁸

Gráfico 7 – Taxa de participação



Fontes: IBGE e BCB

Gráfico 8 – População ocupada



Fontes: IBGE e BCB

Os resultados das duas implementações empíricas são, em geral, semelhantes, como esperado. Na primeira abordagem, excluem-se amostras de grupos de entrevistas iniciados em 2020, restando apenas amostras de grupos de entrevistas iniciados em 2019. Na segunda, aproxima-se a estrutura demográfica das amostras trimestrais de 2020 da estrutura demográfica observada em 2019. Logo, ambas são tentativas de mitigar o viés de seleção em variáveis observáveis (demográficas) em 2020.

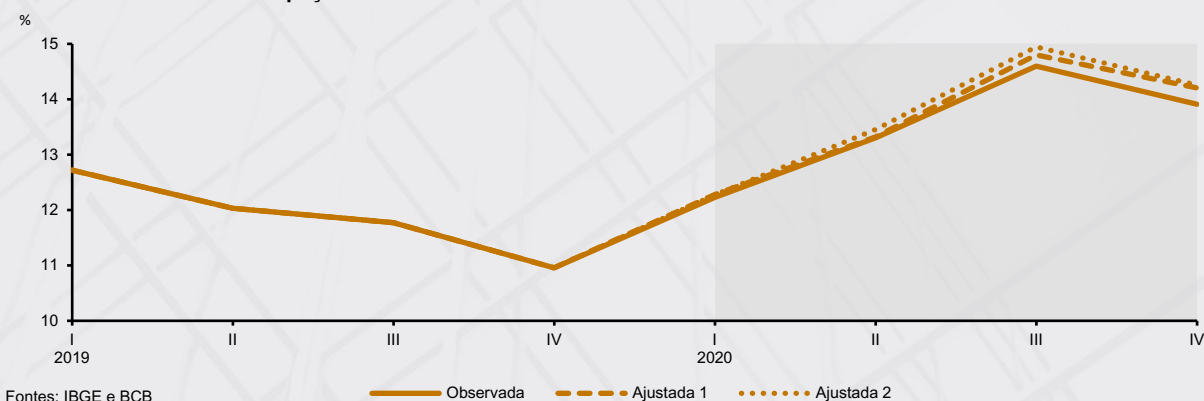
8/ A definição formal da simulação 2 (s_2) de alguma variável de mercado de trabalho segue:

$$X_t^{s_2} = \sum_{\hat{g}} \sum_{A(a)} \sum_{U(u)} X_{g,au,t}^{s_2} \left(\frac{\text{PI}\Gamma_{g,au,t}^{\text{IBGE}}}{\text{PI}\Gamma_{g,au,t}^{s_2}} \right), \quad (2)$$

$$\text{onde } X_{g,au,t}^{s_2} = \sum_{i \in F(g,au)} \hat{\theta}_{i,t|2019} w_{i,t} \mathbb{I}(X);$$

tal que $\hat{\theta}_{i,t|2019}$ é a estimativa da propensão de resposta do indivíduo i no trimestre t de 2020, dada a estrutura demográfica de 2019, de acordo com um conjunto de covariadas de uma regressão logística. A mesma é definida como $\text{logit}(\theta_{i,e,t|2019}(y_{i,e,t} = 1 | x_{i,e,t})) = x_{i,e,t}^T \beta_{e,t}$, tal que na implementação $y_{i,e,t} = 1$ se o indivíduo está em algum trimestre de 2019 no grupo de entrevista e , e $y_{i,e,t} = 0$ se o indivíduo está no trimestre t de 2020 e no grupo de entrevista e . x é um vetor de covariadas (*dummies* para região metropolitana, frequenta escola, faixas de escolaridade, condição no domicílio, faixa etária, grande região, sexo, rural, espécie de domicílio, grupos por quantidade de moradores com idade mínima de 14 anos e UFs) e β é o vetor de coeficientes. A escolha das covariadas tem como referência Almeida et al. (2019).

Gráfico 9 – Taxa de desocupação

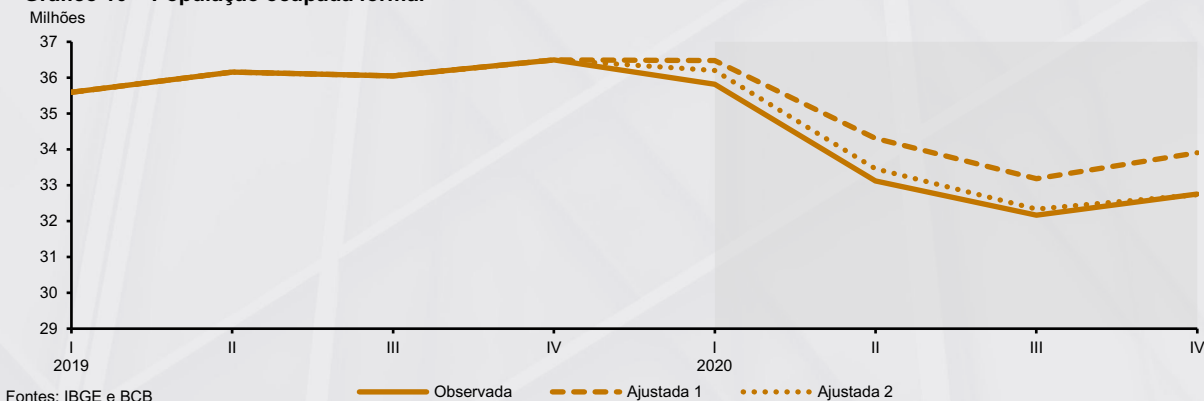


Os resultados presentes nos Gráficos 6 a 9, a princípio, podem surpreender, pois um ajuste na PIT da ordem de 3,8 milhões de pessoas ao longo de 2020 implica alterações de magnitude limitada nas variáveis econômicas do mercado de trabalho. As novas estimativas possivelmente estão dentro do intervalo de confiança para os domínios de estimação da PNAD Contínua. Por exemplo, os ajustes sugerem que a estimativa publicada pelo IBGE sobrestimou a população ocupada em cerca de 0,5 milhão no quarto trimestre de 2020. A taxa de desocupação teria sido cerca de 0,4 p.p. maior do que a divulgada (13,9%) e a taxa de participação teria reduzido menos, aproximadamente 0,8 p.p. (57,6% contra 56,8%), ambos no quarto trimestre de 2020.

Os resultados dos ajustes sugerem que eventual seleção em variáveis observáveis ligadas à demografia, que está associado ao aumento do viés de disponibilidade devido à redução de amostra com a mudança de coleta por telefone na pandemia, não parece ter gerado mudanças significativas nas principais estimativas agregadas ligadas a ocupação e participação no mercado de trabalho na PNAD Contínua. Isso não descarta, entretanto, a possibilidade de que os problemas de coleta tenham implicado viés de seleção em variáveis não-observáveis *ex-ante* pela pesquisa.⁹

Um possível exemplo desse tipo de viés está ilustrado nos Gráficos 10 e 11. O Gráfico 10 com a população ocupada formal¹⁰ indica que – segundo a primeira implementação empírica do ajuste, que exclui da amostra os grupos de entrevistas iniciados em 2020 – a queda da ocupação formal teria sido entre 0,7 e 1,2 milhão menor do que a divulgada em 2020. Ao excluir da amostra os grupos de entrevistas iniciados em 2020, a primeira implementação pode mitigar tanto viés de seleção em observáveis demográficas quanto viés em não-observáveis. Já a segunda estratégia empírica de ajuste, que é estritamente demográfico e corrige o

Gráfico 10 – População ocupada formal

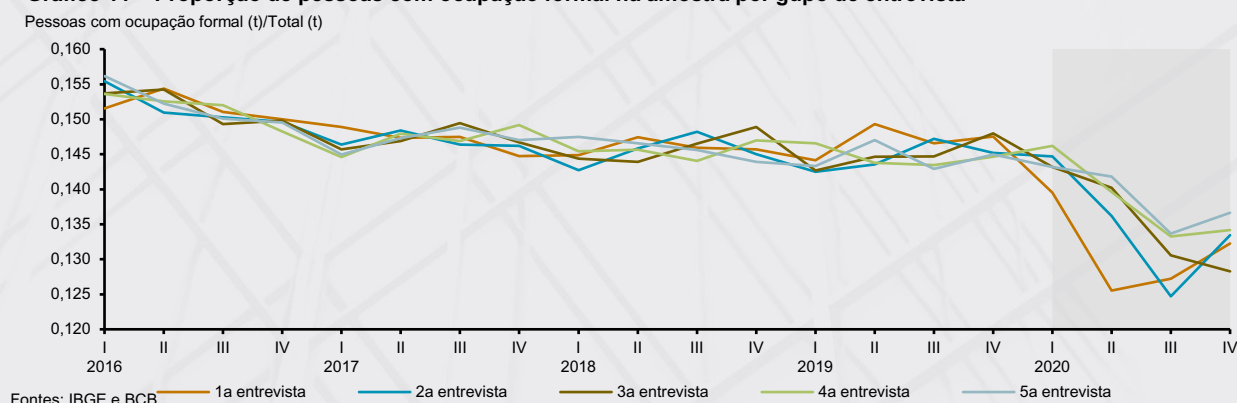


9/ Viés de seleção em variáveis não-observáveis *ex-ante* pela pesquisa indica que a pesquisa pode estar involuntariamente selecionando indivíduos em condições específicas na força de trabalho ou fora dela, por exemplo. Por outro lado, na seleção em variáveis observáveis, a pesquisa seleciona indivíduos por características sabidas (observáveis) no desenho da amostra, como por exemplo, faixa etária e sexo.

10/ A definição de ocupados formais segue: empregados do setor privado e público com carteira e empregados do setor público sem carteira contribuintes da previdência. Exclui empregados domésticos, estatutários e militares.

peso amostral pela propensão de resposta, indica queda da ocupação formal mais próxima da originalmente divulgada. O Gráfico 11 ilustra como a proporção de indivíduos com ocupação formal variou de forma diferenciada entre grupos de entrevista após a pandemia.¹¹

Gráfico 11 – Proporção de pessoas com ocupação formal na amostra por grupo de entrevista



Resumidamente, o boxe investigou os impactos da redução da taxa de respostas da PNAD Contínua em indicadores do mercado de trabalho. Os resultados de dois métodos distintos de ajuste indicam que o possível viés de seleção em variáveis observáveis ligadas à demografia não parece ter implicado mudanças relevantes em estimativas agregadas de ocupação e participação no mercado de trabalho. Todavia, os métodos são incapazes de corrigir plenamente a existência de viés de seleção em variáveis não-observáveis.¹² Se este existir, o acompanhamento do estado do mercado de trabalho durante a pandemia é ainda mais desafiador e não somente para os analistas, mas também para os institutos que conduzem pesquisas amostrais domiciliares em todo o mundo.

Referências bibliográficas

Almeida, P. A. *et al.* Pesos longitudinais para a Pesquisa Nacional por Amostra de Domicílios Contínua (PNAD Contínua). **Boletim Mercado de Trabalho: Conjuntura e Análise**, n. 67, out. 2019. Disponível em: <https://www.ipea.gov.br/portal/images/stories/PDFs/mercadodetrabalho/191101_bmt_67_nt_pesos_longitudinais.pdf>.

Corseuil, C. H, Russo, F. A redução no número de entrevistas na PNAD Contínua durante a pandemia e sua influência para a evolução do emprego formal. **Carta de Conjuntura: Mercado de Trabalho**, n. 50, Nota de conjuntura 22, 1º trimestre de 2021. Disponível em: <https://www.ipea.gov.br/portal/images/stories/PDFs/conjuntura/210318_cc_50_nota_22_amostra_da_pnad_continua.pdf>.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Pesquisa Nacional por Amostra de Domicílios Contínua (PNAD Contínua)**: informações referentes à coleta do mês de abril de 2020. Rio de Janeiro: IBGE, 2020. (Nota Técnica). Disponível em: <https://ftp.ibge.gov.br/Trabalho_e_Rendimento/Pesquisa_Nacional_por_Amostra_de_Domicilios_continua/Nota_Tecnica/Nota_Tecnica_Divulgacao_2Tri2020_Agosto_2020.pdf>.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Pesquisa Nacional por Amostra de Domicílios Contínua (PNAD Contínua)**: Sobre o processo de ponderação da PNAD Contínua. Rio de Janeiro: IBGE, 2021. (Nota Técnica). Disponível em: <https://ftp.ibge.gov.br/Trabalho_e_Rendimento/Pesquisa_Nacional_por_Amostra_de_Domicilios_continua/Nota_Tecnica/Nota_Tecnica_02_2021_Sobre_o_processo_de_ponderacao.pdf>.

11/ Destaca-se que este comportamento agregado da ocupação formal na amostra da PNAD Contínua no Gráfico 11 se replica por faixas etárias e sexo. Logo, não se trata de efeito composição desses grupos demográficos.

12/ O primeiro método de ajuste minimiza o problema de viés de seleção em variáveis não-observáveis. Contudo não pode ser utilizado a partir do primeiro trimestre de 2021, pois todos os indivíduos da amostra terão respondido a pesquisa apenas por telefone.

ILO – INTERNATIONAL LABOUR ORGANIZATION. Ilostat: COVID-19 impact on the collection of labour market statistics. Geneva: ILO, 2020. Disponível em: < <https://ilostat.ilo.org/topics/covid-19/covid-19-impact-on-labour-market-statistics/>>