

# Evolução dos retornos da escolaridade no Brasil<sup>1</sup>


Marcos J. RIBEIRO

mjribeiro@usp.br |  <https://orcid.org/0000-0002-3456-3982>

Luciano NAKABASHI

luciano.nakabashi@gmail.com |  <https://orcid.org/0000-0002-4937-5669>

Fernando BARROS JR.

fabarrosjr@usp.br |  <https://orcid.org/0000-0002-9073-7684>

## Resumo

Neste artigo, estimamos os retornos da escolaridade para o Brasil de 1995 a 2023, utilizando dados da Pesquisa Nacional por Amostra de Domicílios (PNAD) e da PNAD Contínua (PNADC). Os resultados mostram redução nos retornos salariais da escolaridade, sobretudo entre trabalhadores com menor escolaridade. Essa trajetória de declínio é atribuída a dois determinantes principais: (i) a qualidade do sistema educacional; e (ii) a expansão da oferta de mão de obra qualificada, que resultou em saturação do mercado e alocação ineficiente de trabalhadores, onde trabalhadores menos qualificados foram, pelo menos parcialmente, substituídos por trabalhadores mais qualificados em ocupações semelhantes.

## Palavras Chave

Retornos da escolaridade, Educação, Salários.

## Evolution of returns to schooling in Brazil

### Abstract

In this paper, we estimate the returns to schooling in Brazil from 1995 to 2023, using data from the National Household Sample Survey (PNAD) and the Continuous PNAD (PNADC). The results show a reduction in wage returns to schooling, particularly among workers with lower educational attainment. This declining trend is attributed to two main determinants: (i) the quality of the educational system; and (ii) the expansion of the supply of skilled labor, which led to market saturation and inefficient worker allocation, where less-educated workers were at least partially replaced by more-educated workers in similar occupations.

### Keywords

Returns to schooling, Education, Wages.

### JEL Classification

J01, J21, J23, I26.

<sup>1</sup> Universidade de São Paulo (USP), Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto (FEA/RP), Ribeirão Preto, SP, Brasil.

Recebido: 10/04/2024.

Revisado: 27/11/2024.

Aceito: 06/12/2024.

DOI: <https://doi.org/10.1590/1980-53575522mlf>



This Article is Distributed Under the Terms of the Creative Commons Attribution 4.0 International License.

## 1. Introdução

A teoria econômica sustenta que níveis mais elevados de escolaridade desempenham papel fundamental no processo de desenvolvimento e crescimento econômico, pois facilitam a criação e disseminação de tecnologia, além de elevarem a produtividade do trabalho e os salários (Perez-Silva e Partridge 2020; Bosi *et al.* 2021; Deming, 2022). No Brasil, observamos uma aceleração nos investimentos em educação após a promulgação da Constituição de 1988. Por exemplo, dados do Banco Mundial indicam que os investimentos públicos em educação passaram de 4,6% do PIB, em 1995, para 5,8%, em 2020.<sup>1</sup> Esses investimentos, somados a outros fatores, resultaram em um aumento significativo na escolaridade média dos brasileiros, que passou de 6,0 anos, em 1995, para 10,3 anos, em 2023.<sup>2</sup>

A expansão do acesso ao sistema educacional brasileiro pode ter sido acompanhada por uma deterioração na qualidade da educação pública. Segundo Oliveira e Araujo (2005), e Curi e Menezes-Filho (2014), apesar da ampliação do acesso à educação com a promulgação da Constituição de 1988, o sistema educacional público brasileiro ainda não havia conseguido garantir uma educação de qualidade. De acordo com Curi e Menezes-Filho (2014), os resultados do exame de desempenho aplicado como parte do Sistema de Avaliação da Educação Básica (SAEB) documentam queda no desempenho escolar entre 1995 e 1999. Em contrapartida, Brotherhood *et al.* (2019) encontraram forte correlação positiva entre as notas do SAEB e os retornos da escolaridade estadual, sugerindo que os retornos podem ser utilizados como medida de qualidade da educação.

O aumento do tempo médio de estudo no Brasil trouxe mudanças significativas nos retornos da educação. Neste estudo, estimamos esses retornos e analisamos as possíveis causas das alterações observadas ao longo do tempo. Utilizando dados da Pesquisa Nacional por Amostra de Domicílios (PNAD) e da PNAD Contínua (PNADC), cobrimos o período de 1995 a 2023 e realizamos as estimativas em três dimensões: de forma agregada; por grupos de escolaridade; e por estado. As análises levam em conta

<sup>1</sup> Ver os dados do Banco Mundial: <https://data.worldbank.org>.

<sup>2</sup> Ver a Tabela C2 no Apêndice A. Destacamos também que importantes reformas educacionais influenciaram a permanência de crianças e jovens na escola e, consequentemente, no tempo médio de estudo (Menezes Filho 2003). Nesse sentido, destacam-se a Lei de Diretrizes e Bases da Educação Nacional (1996), a criação do Fundo de Manutenção e Desenvolvimento do Ensino Fundamental e de Valorização do Magistério (FUNDEF) e a criação do programa Bolsa Escola, em 2001.

diversas características socioeconômicas dos trabalhadores, como nível de escolaridade, sexo, raça, setor de atividade e zona de residência.

O método tradicional de estimação da equação de salários, com base em Mínimos Quadrados Ordinários (MQO), é amplamente utilizado, mas pode gerar vieses, por exemplo, vies de habilidade, retorno e erro de medida. Para lidar com essas limitações, adotamos o método proposto por Garen (1984), que corrige esses vieses, e o método de Heckman (1979), que corrige o viés de seleção amostral. Com isso, conseguimos obter estimativas mais robustas e confiáveis sobre os retornos da educação no Brasil.

Os resultados indicam que trabalhadores com maior nível de escolaridade tendem a obter retornos salariais mais elevados para cada ano adicional de estudo. Essa diferença pode ser explicada pela heterogeneidade entre os trabalhadores em diferentes níveis educacionais. Embora isso destaque a relevância dos investimentos em educação, observamos que, desde 1995, o retorno de um ano adicional de estudo tem diminuído gradualmente, independentemente do nível educacional. No entanto, essa tendência de queda tem se enfraquecido nos últimos anos, especialmente entre trabalhadores com maior escolaridade, o que tem ampliado o diferencial de retorno entre os mais e os menos escolarizados.

Argumentamos que a redução nos retornos da escolaridade está associada tanto à qualidade do sistema educacional quanto à expansão da mão de obra qualificada. Além disso, mostramos que o método de MQO subestima os retornos da escolaridade quando há endogeneidade e erros de medida. Por outro lado, ele tende a superestimá-los na presença de vies de seleção amostral.

Para o Brasil, alguns estudos empíricos sobre os retornos da escolaridade estimaram essa medida para um ano específico ou poucos anos (Van Zaist *et al.* 2010; Silva *et al.* 2022; Souza *et al.* 2022; Santiago *et al.* 2023). Nós estendemos a análise para o período que vai de 1995 a 2023. Para isso, utilizamos dados da PNAD e PNADC que, embora sejam semelhantes em alguns aspectos, possuem diferenças metodológicas relevantes. Sendo assim, buscamos minimizar as possíveis lacunas existentes entre as bases de dados para obter estimativas mais precisas. A importância de se analisar um período mais longo reside no fato de obtermos uma compreensão mais abrangente dos efeitos do considerável aumento na escolaridade dos trabalhadores sobre os retornos salariais.

Alguns estudos analisam a relação entre características socioeconômicas dos trabalhadores e os retornos da escolaridade. Por exemplo, Pereira *et al.* (2013) concluíram que os retornos da escolaridade são mais elevados no meio urbano, o que é atribuído às diferenças na qualidade educacional entre zonas rurais e urbanas. De forma semelhante, os resultados de Sachsida *et al.* (2004) indicam que os retornos da escolaridade são mais altos para indivíduos brancos. Em outra análise, Vaz (2017) constatou que os trabalhadores do setor de serviços modernos experimentam maiores retornos da escolaridade em relação aos demais setores da economia.

No presente estudo, avançamos nesse tema ao incluir nas especificações econométricas interações entre os anos de estudo e várias dessas características, por exemplo, sexo, raça, setor de atividade e zona de residência. Essas interações permitem verificar quais características socioeconômicas são importantes nos retornos da escolaridade. Nós agrupamos os trabalhadores em quatro grupos de acordo com o nível de escolaridade atingido e mostramos que, em média, os trabalhadores com retornos da escolaridade mais elevados possuem mais escolaridade, menos experiência no mercado de trabalho, são brancos, moram em regiões urbanas e trabalham no setor de serviços. Por outro lado, os trabalhadores com menos tempo de estudo possuem menores retornos da escolaridade, mais experiência no mercado de trabalho, são homens não brancos, boa parte deles trabalha no setor agrícola e residem em áreas rurais.<sup>3</sup>

Na análise por estados, nossos resultados sugerem duas possíveis explicações para a queda dos retornos da educação: queda na qualidade do sistema educacional; e a expansão da mão de obra mais escolarizada. Primeiro, estimamos três regressões com efeitos fixos estaduais, nas quais o retorno da escolaridade foi utilizado como variável dependente e algumas medidas de qualidade educacional (SAEB Matemática, SAEB Português e IDEB) como variáveis explicativas. Os resultados indicam que nos estados onde a qualidade da educação é maior, os trabalhadores experimentam maiores retornos. Adicionalmente, mostramos que, em média, há correlação positiva entre os retornos da educação e o nível de renda *per capita* dos estados. Esses resultados reforçam a relevância de investimentos em melhorias educacionais para potencializar os benefícios econômicos da escolaridade.

<sup>3</sup> Visando manter equivalência entre os anos de estudo da PNAD e PNADC, os grupos de escolaridade foram definidos conforme as seguintes categorias da PNAD: Grupo 1 (0 a 4 anos de estudo); Grupo 2 (5 a 8); Grupo 3 (9 a 11); e Grupo 4 (12 ou mais). Considerando a PNADC, temos: Grupo 1 (0 a 5 anos); Grupo 2 (6 a 9); Grupo 3 (10 a 12); e Grupo 4 (13 ou mais). Utilizamos a mesma nomenclatura em ambas as bases de dados para facilitar a compreensão.

Adicionalmente, regredimos os retornos da educação nos níveis estadual e por níveis de escolaridade contra a proporção de trabalhadores em cada respectivo nível de escolaridade. Os efeitos marginais indicam que, enquanto a oferta relativa de trabalhadores qualificados cresce, os retornos para esses trabalhadores diminuem, possivelmente devido ao excesso de oferta e ao descompasso educacional, onde trabalhadores mais qualificados acabam em funções em que são superqualificados, enquanto que para os trabalhadores menos qualificados o oposto ocorre. Essas descobertas sugerem a necessidade de um ambiente de negócios mais dinâmico que possa absorver a mão de obra qualificada, tornando a educação mais atraente do ponto de vista financeiro.

Além da presente introdução, na Seção 2 apresentamos as especificações econométricas e cada um dos métodos de estimação utilizados. Na Seção 3, apresentamos as bases de dados empregadas na análise (PNAD, PNADC e SAEB). Os principais resultados e implicações encontram-se na Seção 4. Por fim, a Seção 5 traz as principais conclusões.

## 2. Especificações e Métodos de Estimação

Ao estimar os retornos da escolaridade via Mínimos Quadrados Ordinários (MQO) surgem alguns vieses em sua estimação, sendo eles: vies de habilidade; vies de retorno; vies de erro de medida; e vies de seleção amostral (Heckman 1979; Card 2001; Blundell *et al.* 2001). Logo, estimamos os retornos por MQO e utilizamos dois métodos adicionais: o método de Heckman (1979) para corrigir o vies de seleção amostral; e o de Garen (1984) para corrigir os demais.

### 2.1. MQO

Para a análise empírica, utilizamos a especificação da equação minceriana com variáveis que são comuns na literatura. Os variáveis anos de estudo, experiência e experiência ao quadrado são provenientes da especificação minceriana original (Mincer 1958; Mincer 1974). As *dummies* de raça, zona de residência, sexo, setor de ocupação, região de residência e participação em sindicato são variáveis que captam as heterogeneidades dos indivíduos e são comuns na literatura empírica (Suliano e Siqueira 2012; Nakabashi e

Assahide 2017; Brotherhood *et al.* 2019; Souza *et al.* 2022; Santiago *et al.* 2023). A equação base para as estimativas do presente estudo é a seguinte:

$$\begin{aligned} \ln Y_i = & \alpha_0 + \beta_0 S_i + \beta_1 Exp_i + \beta_2 Exp_i^2 + \beta_3 Raça_i + \beta_4 Urb_i + \beta_5 Agr_i + \beta_6 Serv_i \\ & + \beta_7 Sexo_i + \beta_8 (Exp_i \times S_i) + \beta_9 (Raça_i \times S_i) + \beta_{10} (Urb_i \times S_i) \\ & + \beta_{11} (Agr_i \times S_i) + \beta_{12} (Serv_i \times S_i) + \beta_{13} (Sexo_i \times S_i) + \gamma X \epsilon_i \end{aligned} \quad (1)$$

em que  $\ln Y_i$  é o logaritmo natural do salário hora do indivíduo  $i$ ;  $S_i$  representa os anos de estudo;  $Exp_i$  é a experiência no mercado de trabalho.<sup>4</sup> As demais variáveis são *dummies*.  $Raça_i$  é 1 se branco e 0 caso contrário;  $Urb_i$  é igual a 1 caso o indivíduo more em áreas urbanas;  $Agr_i$  tem o valor 1 caso o setor de atividade seja na agricultura e  $Serv_i$  tem o valor 1 caso esteja no setor de serviços;  $Sexo_i$  é igual a 1 caso seja homem;  $X$  é um conjunto de variáveis de controle que inclui *dummies* indicando a região do Brasil em que o indivíduo reside (Sudeste, Nordeste, Centro Oeste e Sul) e se é sindicalizado. Por fim,  $\epsilon_i$  é um termo de erro aleatório.

Há evidências que mostram diferenças nos retornos da escolaridade de acordo com a zona de residência (urbana ou rural) (Pereira *et al.* 2013; Kassouf 1997), sexo (Zhang *et al.* 2005; Psacharopoulos e Patrinos 2018), raça (Sachsida *et al.* 2004) e setores de atividade econômica (Vaz 2017). O que diferencia nossa especificação, expressa na Equação (1), de outras pesquisas é a adição de interações entre os anos de estudo e algumas das variáveis (Experiência, Raça, Urb, Agr, Serv, Sexo). Considerar essas interações permite analisar a relevância das características socioeconômicas nos retornos da escolaridade.

Ao derivarmos a Equação (1) em relação a  $S_i$ , obtemos o retorno marginal da escolaridade:

$$\frac{\partial \ln Y_i}{\partial \ln S_i} = \beta_0 + \beta_8 Exp_i + \beta_9 Raça_i + \beta_{10} Urb_i + \beta_{11} Agr_i + \beta_{12} Serv_i + \beta_{13} Sexo_i \quad (2)$$

Essa medida fornece o aumento percentual médio no salário hora para cada ano adicional de estudo.

<sup>4</sup> Seguimos Heckman *et al.* (2003) e calculamos a experiência como: idade - anos de estudo - 6.

## 2.2. O Método de Garen

O método MQO produz estimativas tendenciosas dos retornos da escolaridade. Os vieses que afetam as estimativas são o de habilidade, retorno e de erro de medida (Blundell *et al.* 2001; Card 2001). Apenas para visualizar como cada um deles afeta as estimativas, vamos assumir que a equação de salário tem como variável independente apenas o tempo de estudo  $S_i$  e que os retornos da escolaridade são heterogêneos. Temos então que:

$$Y_i = \alpha_i + \beta_i S_i + u_i \quad (3)$$

Nesse caso, temos um intercepto e um retorno da escolaridade para cada indivíduo  $i$ . No entanto, é comum na literatura estimar um  $\alpha$  e um  $\beta$  que são as médias populacionais de  $\alpha_i$  e  $\beta_i$ . Dessa forma, podemos reescrever a Equação (3) como:

$$Y_i = \alpha + (\alpha_i - \alpha) + \beta S_i + (\beta_i - \beta) S_i + u_i \quad (4)$$

Ao agrupar os termos não observáveis obtemos:

$$Y_i = \alpha + \beta S_i + \epsilon_i \quad (5)$$

onde  $\epsilon_i = (\alpha_i - \alpha) + (\beta_i - \beta) S_i + u_i$ .

A regressão por MQO produz estimativas tendenciosas devido à correlação entre  $\epsilon_i$  e  $S_i$ . O viés de habilidade surge devido à correlação entre o termo  $(\alpha_i - \alpha)$  e  $S_i$ . O viés de retorno é proveniente da correlação entre  $(\beta_i - \beta)$  e  $S_i$ . Por fim, o viés de erro de medida surge devido aos erros de medida em  $S_i$ . Para corrigi-los, utilizamos o método proposto por Garen (1984).<sup>5</sup>

Esse método é bastante comum na literatura sobre retornos de escolaridade e pode ser visto nos estudos de Sachsida *et al.* (2004) e Suliano e Siqueira (2012). A ideia principal é estimar, no primeiro estágio, a variável endógena (anos de estudo) em função de alguns instrumentos e, no

<sup>5</sup> O método de Garen é uma alternativa ao método de variáveis instrumentais. Optamos pelo primeiro, pois a escolha do instrumento nem sempre é trivial (Card 2001). Adicionalmente, dentro do presente contexto em que analisamos um horizonte temporal de 29 anos, variáveis comumente utilizadas como instrumento, por exemplo, a escolaridade da mãe, não estão disponíveis. Além disso, o uso de variáveis instrumentais pouco correlacionadas com a variável explicativa pode levar a inconsistências ainda maiores (Bound *et al.* 1995).

segundo estágio, utilizar os erros desta estimação como regressor adicional na especificação principal (Equação 1). No primeiro estágio, estimamos a seguinte equação:

$$S_i = \kappa_0 + \kappa_1 Raça_i + \kappa_2 Mãe_i + \kappa_3 Urb_i + \eta_i, \quad (6)$$

sendo que  $S_i$  representa os anos de escolaridade do indivíduo  $i$ ;  $Raça_i$  toma o valor 1 se é branco e 0 caso contrário;  $Mãe_i$  assume o valor 1 caso a mãe esteja viva;  $Urb_i$  tem valor 1 caso more em área urbana. Também incluímos *dummies* que indicam a região do Brasil onde o indivíduo reside.<sup>6</sup> Por fim,  $\eta_i$  é o termo de erro aleatório.

Variáveis como raça e região onde reside podem ser vistas nos estudos de Garen (1984) e Sachsida *et al.* (2004). A zona de residência (rural ou urbana) também influencia nas decisões de estudar, uma vez que indivíduos que moram na zona rural podem ter pior saúde e nutrição, além de maiores dificuldades no acesso à escola (Gollin *et al.* 2014). Já a mãe do indivíduo, estando viva, pode influenciar na educação provendo os meios para que ele estude, sobretudo na infância (Suliano e Siqueira 2012).

No segundo estágio, utilizamos o termo de erro da Equação (6) e estimamos a Equação (1) acrescida dos seguintes termos:

$$\beta_{14} \hat{\eta}_i + \beta_{15}(\hat{\eta}_i \times S_i) \quad (7)$$

Dessa forma, estamos assumindo que o termo de erro da Equação (1) pode ser capturado por uma relação linear  $\epsilon_i = \beta_{14} \hat{\eta}_i + \beta_{15}(\hat{\eta}_i \times S_i) + \omega_i$ , e que  $E(S_i, \hat{\eta}_i, \hat{\eta}_i S_i) = 0$ . Portanto, ao incluir a Equação (7) na Equação (1), estamos produzindo um novo termo de erro  $\omega_i$  que não está correlacionado com  $S_i$ , resolvendo o problema da endogeneidade.

### 2.3. O Método de Heckman

Em alguns estudos, a equação minceriana é estimada utilizando-se somente dados dos indivíduos que estão no mercado de trabalho. Ao proceder dessa forma, a amostra está sendo selecionada de forma não aleatória causando o viés de seleção amostral. Uma das fontes desse viés é a existência de diferentes salários de reserva que afetam as decisões dos indivíduos de

<sup>6</sup> Por conveniência, não indicamos as *dummies* na Equação (5).



entrarem ou não no mercado de trabalho. Dessa forma, isso deve ser levado em consideração ao se estimar os retornos da escolaridade. Para corrigir esse viés, o método de Heckman (1979), que também é estimado em dois estágios, é o procedimento mais adotado e pode ser visto em Marcelo e Wyllie (2006), Moura (2008) e Nakabashi e Assahide (2017).

No primeiro estágio, estimamos a equação de seleção, que deve conter variáveis que afetam a decisão dos indivíduos em entrar no mercado de trabalho como, por exemplo, anos de estudo, experiência de trabalho e *dummies* que indicam raça, zona de residência, sexo e número de filhos. Essa equação é estimada no intuito de calcular a razão inversa de Mills. A equação de seleção que estimamos é a seguinte:

$$D_i = \psi_0 + \psi_1 Raça_i + \psi_2 Urb_i + \psi_3 Sexo_i + \psi_4 Filhos_i + \psi_5 S_i + \zeta_i, \quad (8)$$

Onde  $D_i$  é uma *dummy* com valor 1 caso o indivíduo  $i$  esteja empregado e 0 caso contrário;  $Raça_i$  é 1 se branco e 0 caso contrário;  $Urb_i$  tem o valor 1 se o indivíduo mora na área urbana;  $Sexo_i$  tem valor 1 se é do sexo masculino;  $Filhos_i$  é 1 se tem filhos menores de 14 anos; e  $S_i$  são os anos de estudo. Adicionalmente, inserimos *dummies* que indicam a região do Brasil que o indivíduo reside. Por fim,  $\zeta_i$  é um termo de erro aleatório.

As variáveis sexo, raça e filhos são comuns na literatura e podem influenciar a probabilidade de participação no mercado de trabalho. Por exemplo, mulheres com filhos podem ter salário de reserva maiores que os demais participantes do mercado de trabalho. Segundo Monte *et al.* (2011), o salário de reserva das mulheres é quase o dobro daquele dos homens em todas as regiões do Brasil. Já variáveis de região e zona de residência (urbana ou rural) servem para captar as heterogeneidades do mercado de trabalho brasileiro.

A Equação (8) é estimada utilizando-se o modelo Probit e a razão inversa de Mills é calculada da seguinte forma:

$$\lambda_i = \frac{\phi(\widehat{D}_i)}{\Phi(\widehat{D}_i)}, \quad (9)$$

sendo que  $\phi(\cdot)$  é a função densidade de probabilidade e  $\Phi(\cdot)$  é a função de distribuição acumulada da distribuição normal padrão. A Equação (1) omite a razão inversa de Mills como regressor de modo que o viés de

seleção amostral é resultado da omissão de variável explicativa. Portanto, no segundo estágio do método de Heckman, estimamos a Equação (1) acrescida da razão inversa de Mills:

$$\beta_{14}\lambda_i \quad (10)$$

Consequentemente, a inclusão da razão inversa de Mills produz estimativas dos retornos da escolaridade sem o viés de seleção amostral.

### 3. Bases de Dados

#### 3.1. PNAD e PNADC

Para estimar os retornos da educação no Brasil, utilizamos dados da Pesquisa Nacional por Amostra de Domicílios (PNAD) de 1995 a 2015 e da Pesquisa Nacional por Amostra de Domicílios Contínua (PNADC) de 2016 a 2023. A PNAD, conduzida pelo Instituto Brasileiro de Geografia e Estatística (IBGE), foi encerrada em 2015, enquanto a PNADC, iniciada em 2012, continua em vigor. A partir de 2004, a cobertura das pesquisas se estendeu a todas as regiões do Brasil. Ambas oferecem dados socioeconômicos abrangentes sobre a população brasileira, incluindo informações sobre renda, educação, fecundidade, habitação, entre outros. Embora sejam semelhantes em vários aspectos, a PNAD e PNADC possuem diferenças relevantes que podem ter efeitos sobre as estimativas dos retornos da escolaridade. Destacamos quatro delas.<sup>7</sup>

A primeira é devido às diferenças no método de coleta empregado. A PNADC abrange um número maior de municípios, totalizando 3.500 contra os 1.110 da PNAD. Além disso, sua amostra é substancialmente maior, compreendendo cerca de 211.000 domicílios particulares, em comparação aos, aproximadamente, 150.000 da PNAD. Em termos de periodicidade, a PNADC é realizada trimestralmente e anualmente, enquanto a PNAD ocorre apenas anualmente. Essas diferenças são inerentes aos métodos de coleta de dados e não podem ser modificadas. Outra distinção importante é a idade mínima para considerar as pessoas em idade de trabalhar, sendo 14 anos na PNADC e 10 anos na PNAD.

<sup>7</sup> A PNAD e a PNADC estão disponíveis no site do IBGE: <https://www.ibge.gov.br>. Detalhes sobre as principais diferenças metodológicas entre ambas podem ser vistos em IBGE (2015b).

A segunda diferença é que há duas variáveis *dummies* que empregamos em nossas especificações econométricas que estão presentes na PNAD, mas não na PNADC. A primeira, *Filhos<sub>i</sub>*, indica se o indivíduo possui filhos, enquanto a segunda, *Mãe<sub>i</sub>*, indica se a mãe do indivíduo está viva. Para estimar as duas variáveis, utilizamos redes neurais artificiais (RNA) e dados da PNAD. Obtivemos uma acurácia média de 69% para *Filhos<sub>i</sub>* e 75% para *Mãe<sub>i</sub>*. Detalhes adicionais sobre as estimações estão no Apêndice B.

A terceira diferença entre as pesquisas está nas definições das variáveis utilizadas, especialmente na contabilização dos anos de escolaridade. Na PNAD, categorizamos os trabalhadores em quatro grupos de escolaridade: Grupo 1 (0 a 4 anos de estudo), Grupo 2 (5 a 8 anos), Grupo 3 (9 a 11 anos) e Grupo 4 (12 anos ou mais). Os dois primeiros grupos correspondem aos ciclos iniciais do ensino fundamental, o terceiro ao ensino médio e o quarto ao ensino superior.

Na PNADC, a conclusão do ensino fundamental corresponde a 9 anos de estudo, o ensino médio completo a 12 anos, e o ensino superior a 16 anos. Para garantir consistência e equivalência com a PNAD, ajustamos os dados da PNADC minimizando essas discrepâncias. Agrupamos os anos de estudo da PNADC da seguinte forma: Grupo 1 (0 a 5 anos), Grupo 2 (6 a 9 anos), Grupo 3 (10 a 12 anos), e Grupo 4 (13 anos ou mais)<sup>8</sup>. Adotamos essa nomenclatura ao longo do artigo e nas figuras para facilitar a leitura. Essa simplificação entre ambas as bases de dados visa manter a clareza.

A quarta diferença é que, antes de 2004, a PNAD não incluía a área rural da região Norte do país. Para avaliar o impacto dessa região, estimamos os retornos da escolaridade sem considerar a área rural da região Norte a partir de 2004. Constatamos que as mudanças nos resultados são marginais, afetando sobretudo o Grupo 1.<sup>9</sup> Para esse grupo, houve um aumento médio nos retornos de escolaridade de 0,4%, 0,41% e 0,5% nas estimativas pelos métodos de Mínimos Quadrados Ordinários (MQQ), Heckman e Garen, respectivamente. Dessa forma, esses ajustes praticamente não alteraram os resultados encontrados.<sup>10</sup>

<sup>8</sup> Detalhes sobre a metodologia de cálculo da variável anos de estudo da PNAD podem ser consultados em IBGE (2015a).

<sup>9</sup> Parcela considerável da população rural da região Norte tem entre 0 a 4 anos de estudo.

<sup>10</sup> Não reportamos esses resultados apenas para manter a objetividade da análise, mas eles estão disponíveis com os autores.

Para filtrar as informações relevantes, tornando as duas bases de dados comparáveis e permitir a correta estimação dos retornos da escolaridade, realizamos alguns ajustes nos dados. Como a PNAD não foi realizada nos anos de 2000 e 2010 devido à realização do Censo Demográfico, esses anos foram excluídos da análise. Restringimos a amostra aos trabalhadores com idade entre 25 e 64 anos, que não estudam, trabalham mais de uma hora por semana e não são militares ou servidores públicos. Também excluímos indivíduos que se autodeclararam indígenas ou amarelos, os que trabalham para o próprio consumo, bem como aqueles com dados faltantes nas variáveis analisadas.<sup>11</sup> É importante destacar que, em dois dos métodos econométricos descritos na Seção 2 (MQO e Garen), utilizamos apenas dados de indivíduos empregados. No entanto, para implementar o método de Heckman, também incluímos os dados de indivíduos não ocupados.

Conforme destacamos na seção anterior, a variável dependente da equação minceriana (Equação 1) é o logaritmo do salário-hora em reais de 2015. Calculamos essa medida dividindo o rendimento mensal em reais, recebido no trabalho principal, pela quantidade de horas trabalhadas semanalmente, multiplicada por 4,2.<sup>12</sup> Adicionalmente, excluímos os trabalhadores que recebem mais de 100 salários mínimos.<sup>13</sup>

Por fim, ressaltamos que a PNAD e PNADC são pesquisas amostrais com estratificação, conglomeração e probabilidades desiguais de seleção. Portanto, em nossa análise, consideramos o desenho amostral de ambas. A não consideração do desenho amostral pode resultar em estimativas tendenciosas dos retornos da escolaridade, além de testes de hipóteses e intervalos de confiança tendenciosos (Lumley 2004; Moura 2008). No Apêndice C, fornecemos as médias das variáveis que utilizamos na equação minceriana.

<sup>11</sup> Vaz e Barreira (2021) também realizaram ajustes semelhantes no intuito de compatibilizar a PNAD e PNADC.

<sup>12</sup> O salário reportado na PNAD e PNADC é mensal, enquanto as horas trabalhadas são semanais. Portanto, é necessário ajustar as duas medidas para que estejam na mesma unidade temporal. Pressupomos 252 dias úteis no ano, 21 dias úteis no mês e que os indivíduos trabalham 5 dias por semana. Como a PNAD e PNADC reportam as horas trabalhadas semanalmente, dividimos 21 por 5, obtendo 4,2 “semanas” trabalhadas no mês.

<sup>13</sup> Para esse ajuste, utilizamos o salário mínimo em reais de 2015, conforme o ano de referência da análise.

### 3.2. SAEB e IDEB

Para comparar os retornos da escolaridade com as medidas de qualidade da educação em nível estadual, utilizamos dados do Sistema Nacional de Avaliação da Educação Básica (SAEB) e do Índice de Desenvolvimento da Educação Básica (IDEB).<sup>14</sup> O SAEB é um exame que permite avaliar a qualidade do ensino brasileiro. Já o IDEB é calculado a partir dos dados de aprovação escolar, do Censo Escolar, e das médias dos testes de desempenho em matemática e português do SAEB, que são aplicados aos estudantes no final de cada ciclo de ensino. Como o SAEB é aplicado a cada dois anos, utilizamos os dados dos anos ímpares, no período de 2005 a 2015. Ambos indicadores são fornecidos pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP).<sup>15</sup>

Esses indicadores de qualidade são realizados para alunos de escolas públicas e privadas. Apesar das grandes diferenças existentes entre os dois sistemas de ensino, não entramos em detalhes sobre essas distinções, pois focamos apenas no desempenho médio estadual. Explorar esse tipo de heterogeneidade entre os diferentes sistemas de ensino sobre o retorno salarial é um tema de pesquisa promissor para futuros estudos.

## 4. Resultados

Nesta seção, apresentamos os resultados das estimativas dos retornos da educação e discutimos os possíveis determinantes da tendência de queda observada ao longo do período analisado. Na subseção 4.1, apresentamos as estimativas agregadas para o período de 1995 a 2023. Em seguida, na subseção 4.2, apresentamos as estimativas dos retornos por níveis educacionais. Finalmente, na subseção 4.3, discutimos como a qualidade da educação e a expansão da mão de obra qualificada podem ter influenciado os retornos da educação no Brasil.<sup>16</sup>

<sup>14</sup> Brotherhood *et al.* (2019) utilizam uma metodologia similar para calcular medidas de qualidade da educação no Brasil.

<sup>15</sup> Os dados do SAEB e do IDEB podem ser acessados no site do INEP: <https://www.gov.br/inep>.

<sup>16</sup> As regressões estimadas podem ser consultados no Apêndice Online desse artigo: [https://mjribeiro.github.io/ap\\_online.pdf](https://mjribeiro.github.io/ap_online.pdf). Os retornos da escolaridade que estimamos nos níveis agregado, por grupos de escolaridade e estadual estão disponíveis no formato **.rds** no Zenodo: <https://zenodo.org>.

#### 4.1. Retornos da Escolaridade em Nível Agregado

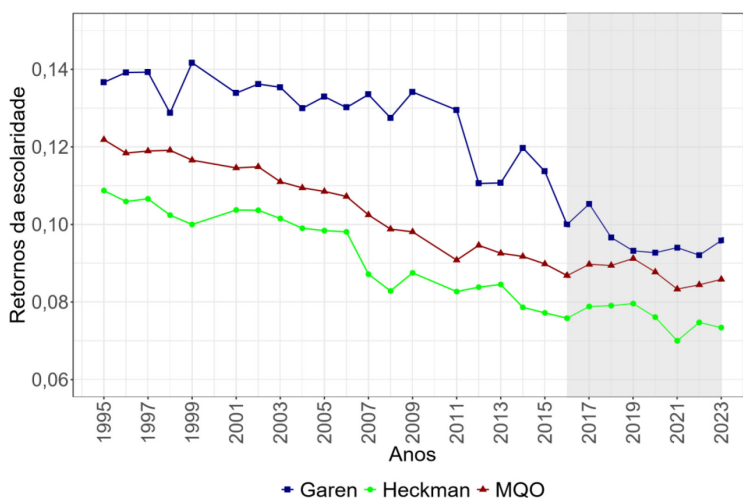
Nesta subseção, apresentamos as estimativas dos retornos da escolaridade em nível agregado. Essa medida foi calculada a partir das médias amostrais das variáveis *Exp*, *Raça*, *Urb*, *Agr*, *Serv*, e *Sexo*, conforme descrito na Equação (2).<sup>17</sup> No método de Garen, também foi utilizada a média de  $\hat{\eta}_i$ . Um retorno da escolaridade de 0,1, por exemplo, significa que cada ano adicional de estudo aumenta o salário-hora em 10%, em média.

A Figura 1 ilustra os resultados no período de 1995 a 2023, utilizando os três métodos distintos: Mínimos Quadrados Ordinários (MQO); o método de Heckman; e o método de Garen. A área clara da figura, que cobre o período de 1995 a 2015, mostra as estimativas dos retornos da escolaridade com base na PNAD. A partir de 2016, a área sombreada indica que essas estimativas passaram a ser feitas com a PNADC. Notamos que os três métodos apontam queda nos retornos da escolaridade. Essa descoberta sugere que um ano adicional de estudo, em média, está resultando em acréscimos cada vez menores ao salário dos trabalhadores brasileiros ao longo do tempo.<sup>18</sup>

As retrações médias em todo o período foram de 29,5%, 32,5% e 29,8%, considerando os métodos de MQO, Heckman e Garen, respectivamente. Essas quedas são equivalentes a retrações anuais médias de 0,93%, 1,01% e 0,94%, respectivamente. Durante o período em que a análise foi realizada com a PNAD, ou seja, entre 1995 e 2015, as retrações foram de 26,3%, 29,0% e 16,8%, o que equivale a quedas anuais médias de 1,17%, 1,28% e 0,78%, respectivamente. Com a PNADC, ou seja, entre 2016 e 2023, as quedas nos retornos da escolaridade foram de 1,1%, 3,2% e 4,2%, correspondendo a quedas anuais médias de 0,16%, 0,45% e 0,59%, respectivamente.

<sup>17</sup> **Exp:** tempo de experiência no mercado de trabalho; **Raça:** *dummy* com valor um caso o indivíduo seja branco e zero caso contrário; **Urb:** *dummy* com valor um caso more em áreas urbanas; **Agr:** *dummy* com valor um caso trabalhe no setor agrícola; **Serv:** *dummy* com valor um caso trabalhe no setor de serviços; **Sexo:** *dummy* com valor um caso seja do sexo masculino.

<sup>18</sup> No Apêndice A, comparamos os retornos da escolaridade estimados com base na PNAD e PNADC entre 2012 e 2015, período em que ambos os conjuntos de dados coexistem.



**Figura 1 - Evolução dos retornos da escolaridade no Brasil - 1995:2023**

*Fonte:* Elaborado pelos autores.

*Nota:* A área clara da figura, que cobre o período de 1995 a 2015, indica que os retornos da escolaridade foram estimados com base na PNAD, enquanto que a área sombreada, a partir de 2016, indica que as estimativas passaram a ser feitas com a PNADC.

Também observamos que, entre 2019 e 2021, houve uma queda mais acentuada nos retornos, especialmente pelos métodos MQO e Heckman. Essa redução está associada ao impacto da pandemia de COVID-19, que provocou uma diminuição no salário médio por hora durante o período (ver Tabela C2), pressionando os retornos para baixo.<sup>19</sup> Embora existam limitações devido às diferenças entre as bases de dados, como discutido na Seção 3, esses resultados sugerem que a tendência de queda nos retornos da escolaridade tem enfraquecido pelos três métodos de estimação, especialmente por MQO e Heckman.

Considerando os três métodos, o MQO superestima os retornos da escolaridade se houver viés de seleção amostral. Esse viés ocorre quando a análise se restringe aos indivíduos que participam do mercado de trabalho e, ao corrigi-lo, o método de Heckman tende a estimar retornos menores.

<sup>19</sup> Entre 2015 e 2016 houve crise econômica no Brasil devido a erros na condução da política econômica por parte do governo Dilma Rousseff, o que potencialmente afeta os retornos da escolaridade. No entanto, não é possível fazer uma análise conclusiva sobre os efeitos da crise nos retornos, uma vez que, esse período coincide com a transição da PNAD para a PNADC, dificultando a distinção entre os efeitos da mudança metodológica e os impactos da crise econômica.

Por outro lado, o MQO subestima os retornos se houver erro de medida ou endogeneidade. O método de Garen corrige esses vieses ao separar os efeitos da educação e das habilidades dos indivíduos sobre os salários. As estimativas por esse método indicam que os retornos da escolaridade nos salários são maiores do que o estimado pelo método MQO. Dessa forma, considerando todos os vieses, é possível que o retorno da escolaridade fique próximo ao estimado pelo método MQO visto que eles tendem a se compensar.

Em relação a literatura, diversos estudos têm apontado para um declínio nos retornos da escolaridade. Por exemplo, Moura (2008) estimou os retornos da educação no Brasil entre 1992 e 2004, identificando uma tendência de queda, exceto para trabalhadores com nível superior. De forma semelhante, Nakabashi e Assahide (2017) observaram queda significativa nos retornos da escolaridade nos salários de trabalhadores brasileiros, tanto entre jovens (15 a 29 anos) quanto adultos (25 a 64 anos), no período de 1997 a 2012. Os autores relataram uma redução no retorno de 21,2% para jovens e 15% para adultos.

Adicionalmente, Suliano e Siqueira (2012), utilizando dados de trabalhadores entre 25 e 64 anos, também constataram queda nos retornos salariais da educação entre 2001 e 2006. De maneira mais ampla, Montenegro e Patrinos (2014) estimaram os retornos da educação em 139 economias e apontaram tendência similar de declínio. Para o Brasil, eles observaram que o retorno da escolaridade caiu de 17,3%, em 1982, para 10,5%, em 2012, representando uma redução de, aproximadamente, 39%.

No entanto, esse fato não é exclusivo do Brasil. Em países com níveis de renda *per capita* semelhantes, como a China e algumas economias emergentes da Europa, também foi registrada diminuição nos retornos da escolaridade, conforme identificado por Asadullah e Xiao (2020) e Horie e Iwasaki (2022), respectivamente.

Dessa forma, os resultados apresentados anteriormente estão em linha com a literatura. O ponto de alento é que mesmo com a manutenção do processo de queda no retorno da escolaridade após 2015, ele tem se reduzido, com manutenção ou até leve elevação dos retornos entre 2021 e 2023 pelos três métodos utilizados, de acordo com os resultados apresentados na Figura 1.



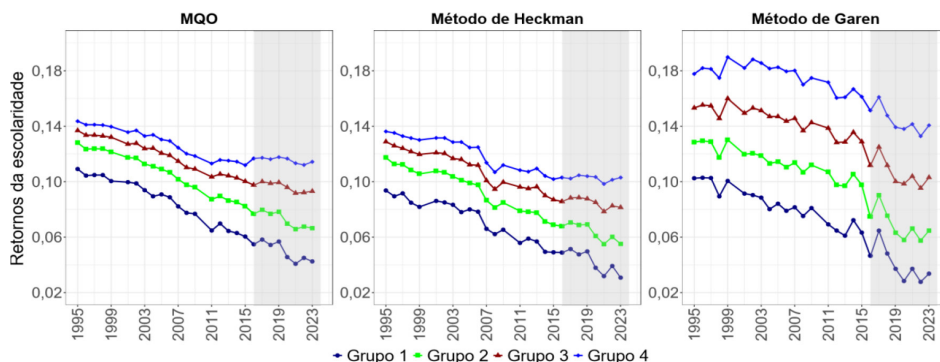
#### 4.2. Retornos da Escolaridade por Níveis Educacionais

Nesta subseção, apresentamos as estimativas dos retornos da escolaridade em que os trabalhadores foram agrupados de acordo com o nível de escolaridade. Em nossa abordagem, seguimos Garen (1984) e Sachsida *et al.* (2004), e reportamos o resultado da Equação (2) para cada um dos quatro grupos de indivíduos classificados de acordo com os anos de estudo. O procedimento é simples: basta substituir na Equação (2) as médias amostrais, condicionada ao grupo, das variáveis Exp, Raça, Urb, Agr, Serv e Sexo. No caso do método de Garen, utilizamos também a média condicional da variável  $\hat{\eta}_i$ .

Enfatizamos que os níveis de escolaridade foram estabelecidos da seguinte maneira: para a PNAD, Grupo 1 corresponde a 0 a 4 anos de estudo, Grupo 2 a 5 a 8 anos, Grupo 3 a 9 a 11 anos, e Grupo 4 a 12 anos ou mais. Já para a PNADC, Grupo 1 abrange 0 a 5 anos, Grupo 2 inclui 6 a 9 anos, Grupo 3 vai de 10 a 12 anos, e Grupo 4 representa 13 anos ou mais de estudo. Para facilitar a leitura utilizamos as denominações Grupo 1, 2, 3 e 4 ao longo da análise, para ambas as bases de dados.

Os resultados da Figura 2 mostram que os retornos da escolaridade são bastante heterogêneos.<sup>20</sup> Observamos que um ano adicional de estudo tem efeito positivo maior nos salários dos indivíduos mais escolarizados. Adicionalmente, na Figura 2, notamos que os retornos estão em queda em todas os grupos de escolaridade. As estimativas dos retornos via MQO e Heckman, com base na PNADC, mostram que a tendência de queda está enfraquecendo, sobretudo para os indivíduos do terceiro e quarto grupos. Segundo Colclough *et al.* (2010), houve uma mudança no padrão mundial dos retornos da escolaridade, que antes era côncavo e passou a ser convexo, indicando um impacto crescente da educação nos salários dos trabalhadores mais qualificados, o que está em linha com os resultados do presente estudo.

<sup>20</sup> No Apêndice C fornecemos as médias dessas estimativas.



**Figura 2- Evolução dos retornos da escolaridade no Brasil por níveis de escolaridade 1995:2023**

Fonte: Elaborado pelos autores.

Notas: (1) A área clara da figura, que cobre o período de 1995 a 2015, indica que os retornos da escolaridade foram estimados com base na PNAD, enquanto que a área sombreada, a partir de 2016, indica que as estimativas passaram a ser feitas com a PNADC. (2) Os grupos de escolaridade foram definidos conforme as seguintes categorias: PNAD - Grupo 1 (0 a 4 anos de estudo), Grupo 2 (5 a 8), Grupo 3 (9 a 11) e Grupo 4 (12 ou mais); PNADC - Grupo 1 (0 a 5 anos), Grupo 2 (6 a 9), Grupo 3 (10 a 12) e Grupo 4 (13 ou mais).

Pelo método de Garen, o enfraquecimento da tendência de queda não é tão evidente para os dois grupos de maior escolaridade em relação aos outros dois métodos, considerando o período 2016-2023. No entanto, nas estimativas pelo método de Garen, isso ocorre pela acentuada queda do retorno da escolaridade entre 2017 e 2020 para todos os grupos. A partir de 2020, não há tendência de queda nos retornos nas estimativas pelo método de Garen, indicando perda de força da tendência de queda nos quatro grupos de escolaridade nos últimos anos.

Na Tabela 1, apresentamos as médias das variáveis utilizadas para calcular os retornos da escolaridade. Ao comparar os cenários da PNAD e da PNADC, observamos melhoria na escolaridade da população brasileira, com redução na proporção de pessoas no primeiro grupo de escolaridade e aumento significativo no terceiro. A experiência média no mercado de trabalho também cresceu em todos os níveis educacionais, especialmente entre os indivíduos do segundo grupo.

Tabela 1 - Características dos indivíduos da amostra condicionado aos níveis de escolaridade - 1995:2023

PNAD - 1995:2015							
Escolaridade	Exp	Raça	Urb	Agr	Serv	Sexo	Proporção
Grupo 1	35,4	39,7	75,4	24,8	47,6	70,8	32,7
Grupo 2	26,3	48,4	90,2	7,6	62,7	64,4	26,6
Grupo 3	20,1	53,6	95,5	2,6	74,5	56,8	28,4
Grupo 4	18,9	75,6	97,8	1,7	81,9	52,6	12,4
PNADC - 2016:2023							
Escolaridade	Exp	Raça	Urb	Agr	Serv	Sexo	Proporção
Grupo 1	37,8	33,6	53,7	40,2	33,9	73,9	23,5
Grupo 2	29,7	36,6	70,3	22,8	46,1	67,6	13,4
Grupo 3	22,2	43,1	83,9	10,0	61,7	58,5	46,3
Grupo 4	18,0	65,3	94,9	2,8	80,0	49,1	16,7

Fonte: Elaborado pelos autores.

Notas: **(1) Exp:** tempo de experiência no mercado de trabalho; **Raça:** % de indivíduos brancos, **Urb:** % de indivíduos que residem em áreas urbanas; **Agr:** % de indivíduos que trabalham na agricultura; **Serv:** % de indivíduos que trabalham no setor de serviços; **Sexo:** % de indivíduos do sexo masculino; **Proporção:** % de pessoas em cada respectivo grupo de estudo. **(2)** Os grupos de escolaridade foram definidos conforme as seguintes categorias: PNAD: Grupo 1 (0 a 4 anos de estudo), Grupo 2 (5 a 8), Grupo 3 (9 a 11) e Grupo 4 (12 ou mais); PNADC: Grupo 1 (0 a 5 anos), Grupo 2 (6 a 9), Grupo 3 (10 a 12) e Grupo 4 (13 ou mais).

Além disso, observamos que a proporção de trabalhadores brancos diminuiu em todos os grupos, enquanto que a de urbanizados apresentou queda mais acentuada entre os menos escolarizados. Também notamos um aumento da participação na agricultura entre os trabalhadores com menor escolaridade, acompanhado por redução no setor de serviços para esse grupo. Esses resultados podem refletir mudanças estruturais na força de trabalho e nos níveis educacionais no Brasil, embora parte dessas variações ocorra pelas diferenças metodológicas entre a PNAD e a PNADC.

De modo geral, de acordo com os resultados apresentados na Tabela 1, os trabalhadores com mais anos de estudo e maiores retornos da escolaridade tendem a ter menos experiência no mercado de trabalho, são predominantemente brancos, vivem em áreas urbanas e atuam no setor de serviços. Em contrapartida, aqueles com menos anos de estudo e menores retornos da escolaridade possuem mais experiência, são majoritariamente homens não brancos (os dados apresentados em sexo aparecem resultados apenas de homens), com uma parcela significativa residindo em áreas rurais e trabalhando no setor agrícola.

Os resultados acima mostram que os trabalhadores com escolaridade mais elevada, sobretudo aqueles que estão no Grupo 4 (pelo menos um ano completo de ensino superior), possuem retornos consistentemente maiores em relação aos trabalhadores nos outros níveis de escolaridade. O enfraquecimento do processo de queda dos retornos nos últimos anos tem ocorrido com maior intensidade para os trabalhadores nos dois grupos de escolaridade mais elevada, mesmo com o aumento da oferta desses trabalhadores, como discutido anteriormente. Dessa forma, os trabalhadores com menores níveis de escolaridade tendem a ficar ainda mais para trás em termos de rendimento tanto pelos menores níveis de escolaridade quanto pelos retornos cada vez mais defasados.

### *4.3. Possíveis Explicações para a Redução nos Retornos da Educação*

#### *4.3.1 Qualidade do Sistema Educacional*

Uma variável importante que possivelmente afeta os retornos da escolaridade é a qualidade do sistema educacional. Há sistemas educacionais eficazes e capazes de desenvolver nos alunos várias habilidades, enquanto outros quase não contribuem para a formação de capital humano. Nesse sentido, trabalhadores educados em instituições de ensino de melhor qualidade podem enviar para o mercado um sinal de maior produtividade.

Brotherhood *et al.* (2019) argumentam que os retornos da escolaridade estão fortemente relacionados à qualidade da educação. Eles demonstram que parte das diferenças nos retornos entre estados brasileiros pode ser explicada pela qualidade dos sistemas educacionais locais, encontrando uma correlação positiva entre esses retornos e indicadores como o gasto por aluno e as notas do SAEB.

Outros estudos, como Hanushek e Woessmann (2008) e Cetra *et al.* (2023), reforçam a importância de incluir medidas de qualidade da educação, por exemplo o SAEB e o IDEB, nas análises dos retornos salariais. Pesquisas como a de Card e Krueger (1992) e Bedi e Edwards (2002) mostram que frequentar escolas de maior qualidade resulta em rendas mais elevadas. No contexto brasileiro, Curi e Menezes-Filho (2014) confirmam que uma melhora de 10% na proficiência em matemática está associada a um aumento de 3% nos salários futuros.

A estratégia de verificar o desempenho regional do SAEB e IDEB sobre o retorno salarial posterior foi utilizada por Cetra *et al.* (2023) para os municípios brasileiros. Os autores consideraram os efeitos no desempenho do SAEB e IDEB, em 2007, como medida de qualidade média da educação municipal do 9º ano do ensino fundamental, quando os alunos têm entre 14 e 15 anos, sobre a renda média municipal dos trabalhadores entre 24 e 25 anos, em 2017. Utilizar uma medida de qualidade do sistema educacional sobre os retornos salariais dez anos depois também ajuda a mitigar o problema da causalidade reversa entre qualidade do ensino e retorno da escolaridade (Cetra *et al.*, 2023).

Para analisar se há relação entre a qualidade da educação e os retornos da escolaridade, estimamos os retornos para os 26 estados brasileiros e o Distrito Federal e estimamos a seguinte regressão:

$$Ret_{jt} = \theta_0 + \theta_1 Qe_{jt} + \sum_j \kappa_j C_j + \omega_{jt}, \quad (11)$$

Onde  $Ret_{jt}$  é o retorno da escolaridade do estado  $j$  no período  $t$ ,  $Qe_{jt}$  é uma medida de qualidade da educação;  $C_j$  são efeitos fixos de estado; e  $\omega_{jt}$  é um termo de erro. Entre as medidas de qualidade da educação estão as notas do SAEB nos exames de matemática e português e a pontuação do Índice de Desenvolvimento da Educação Básica (IDEB), ambas referentes aos alunos do terceiro ano do ensino médio.<sup>21</sup>

Para garantir uma comparação temporal adequada, utilizamos as médias das notas do SAEB e IDEB estaduais da educação recebida pelos jovens no final do ensino médio, quando estão com idades entre 16 e 20 anos, para os anos ímpares entre 2005 a 2013, e estimamos os efeitos dessas medidas de qualidade educacional nos retornos da escolaridade dez anos depois, ou seja, para os anos ímpares entre 2015 a 2023, considerando os trabalhadores entre 26 e 30 anos que possuíam pelo menos o ensino médio completo.<sup>22</sup>

<sup>21</sup> Há grande heterogeneidade na qualidade da educação no Brasil, seja entre as áreas rural e urbana, capital e interior, ou entre escolas pública e privada. No entanto, nosso interesse recai apenas sobre os efeitos médios da qualidade da educação nos retornos à escolaridade, motivo pelo qual não fazemos essas distinções em nossa análise.

<sup>22</sup> Os cortes poderiam ter sido realizados para os trabalhadores entre 26 e 28 anos, visto que a maior parte dos estudantes no terceiro ano do ensino médio possuem entre 16 e 18 anos. No entanto, os resultados são semelhantes quando realizados para este corte de idade.

Nossa hipótese é de que o desempenho dos jovens no IDEB e SAEB no terceiro ano do ensino médio, em 2005, deve afetar o retorno salarial dez anos depois, ou seja, em 2015, quando eles teriam idades entre 26 e 30 anos. O mesmo foi feito para os anos de 2007 e 2017, 2009 e 2019, 2011 e 2021, além de 2013 e 2023. Portanto, os dados são compostos de um painel dos 26 estados brasileiros mais o DF nos 5 períodos, totalizando 135 observações. Os anos ímpares são decorrentes da disponibilidade dos dados do SAEB e IDEB que ocorrem a cada dois anos.

A Tabela 2 ilustra os resultados. Notamos que os coeficientes para as três variáveis de qualidade da educação (SAEB Matemática, SAEB Português, e IDEB) são positivos, sugerindo que maior qualidade educacional está associada a maiores retornos da escolaridade nos estados brasileiros, o que está em linha com os resultados encontrados por Brotherhood *et al.* (2019) e Cetra *et al.* (2023). Esses resultados sugerem que investimentos em melhorias na qualidade da educação levam a aumentos salariais e, portanto, políticas econômicas para melhorar a qualidade dos sistemas educacionais públicos são importantes para reverter o processo de queda nos retornos da educação.<sup>23</sup>

**Tabela 2 - Modelos de regressão em painel - Retornos da escolaridade e medidas de qualidade da educação**

	Variável Dependente: Retornos da Escolaridade (Heckman)		
	(1)	(2)	(3)
SAEB Matemática	0,0002**	-	-
SAEB Português	-	0,0005***	-
IDEB	-	-	0,0117***
Constante	0,0421	-0,0180	0,0597***
EF de Estado	✓	✓	✓
Observações	135	135	135
R <sup>2</sup>	0,71	0,74	0,73
R <sup>2</sup> Ajustado	0,64	0,68	0,66
EP resíduos (df = 107)	0,01	0,01	0,01
Estatística F (df = 27; 107)	9,86***	11,40***	10,84***

Fonte: Elaborado pelos autores.

Notas: (1) Asteriscos simples (\*), duplo (\*\*) e triplo (\*\*\*) denotam significância a 10%, 5% e 1%, respectivamente. (2) EF: Efeito fixo; EP: Erros padrão. (3) Os retornos da escolaridade utilizados nessas regressões foram estimados com base na PNADC. (4) O painel de dados utilizado para essas estimações possui 26 estados mais o Distrito Federal e 5 períodos, totalizando 135 observações.

<sup>23</sup> Na Tabela 2, reportamos os resultados em que a variável dependente é o retorno da escolaridade estimado pelo método de Heckman. Nas estimativas dos retornos realizadas com os demais métodos, ou seja, via MQO e Garen, os resultados são similares. Não apresentamos esses resultados aqui para manter a objetividade da análise.

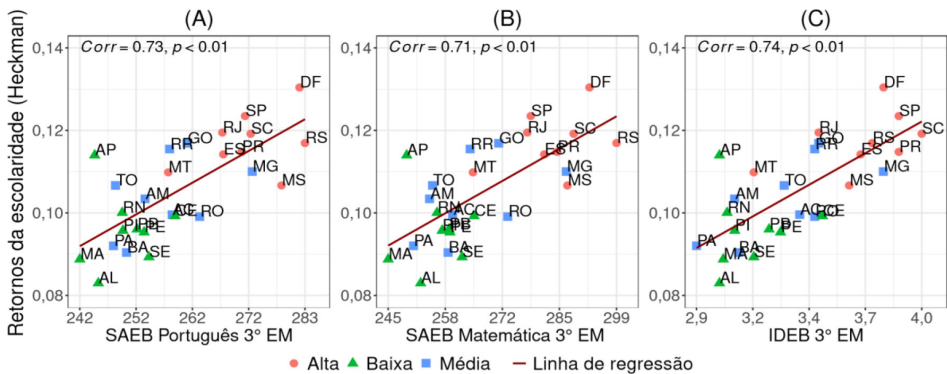
A Figura 3 apresenta a média dos dados utilizados na regressão da Tabela 2. No eixo  $y$ , temos a média dos retornos da escolaridade e, no eixo  $x$ , as médias do SAEB português, SAEB matemática e IDEB, respectivamente. Adicionalmente, classificamos os 26 estados brasileiros mais o Distrito Federal usando dados do PIB *per capita* de 2021.<sup>24</sup> Os nove primeiros são considerados renda alta, os nove do meio são renda média e os nove últimos são renda baixa.<sup>25</sup> Destacamos dois fatos importantes.

Primeiro, há uma clara relação entre as notas médias no SAEB e IDEB, considerando os alunos do terceiro ano do ensino médio, e os retornos médios salariais de cada ano adicional de escolaridade para os trabalhadores entre 26 e 30 anos. Os estados com maiores destaques nos testes, como Rio Grande do Sul (RS), Santa Catarina (SC), São Paulo (SP) e Distrito Federal (DF) são os mesmos que possuem os maiores retornos da escolaridade. Por outro lado, os estados com menores desempenho nos testes, como Alagoas (AL), Maranhão (MA) e Pará (PA) estão entre aqueles com menores retornos da escolaridade.

Segundo, tanto as medidas de qualidade da educação quanto os retornos da escolaridade são bastante heterogêneos entre os estados. Os estados com maiores rendas *per capita* são aqueles que tendem a ter melhores desempenho no SAEB e IDEB, assim como os estados de renda *per capita* baixa possuem maior propensão a terem os piores desempenhos nos testes de aprendizado. Esses resultados são esperados, pois os estados com maiores rendas *per capita* possuem mais recursos para realização de investimentos na melhoria de seus sistemas educacionais.

<sup>24</sup> Acre (AC), Alagoas (AL), Amapá (AP), Amazonas (AM), Bahia (BA), Ceará (CE), Distrito Federal (DF), Espírito Santo (ES), Goiás (GO), Maranhão (MA), Mato Grosso (MT), Mato Grosso do Sul (MS), Minas Gerais (MG), Pará (PA), Paraíba (PB), Paraná (PR), Pernambuco (PE), Piauí (PI), Rio de Janeiro (RJ), Rio Grande do Norte (RN), Rio Grande do Sul (RS), Rondônia (RO), Roraima (RR), Santa Catarina (SC), São Paulo (SP), Sergipe (SE), Tocantins (TO).

<sup>25</sup> Os dados do PIB *per capita* estadual podem ser obtidos no site do IPEA: <http://www.ipeadata.gov.br>.



**Figura 3 - Comparação entre os retornos da escolaridade e medidas de qualidade da educação**

Fonte: Elaborado com dados da Pesquisa Nacional de Amostras por Domicílios (PNADC) e do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP).

Notas: (1) A figura ilustra a média dos retornos da educação, calculados com base na PNADC, nos anos ímpares entre 2015 e 2023, dos trabalhadores de 26 a 30 anos e que possuem pelo menos o ensino médio completo. As médias dos indicadores SAEB português, SAEB matemática e IDEB são as médias dos anos ímpares entre 2005 e 2013. (2) Corr: Correlação de Pearson; p: Valor p da correlação.

Os resultados apresentados nessa subseção sugerem que a qualidade do sistema educacional é um elemento relevante para se entender o diferencial de retorno salarial da escolaridade entre os estados brasileiros. Com a expansão do sistema educacional, sobretudo a partir da Constituição de 1988, é possível que tenha ocorrido uma piora no sistema de educação pública. Oliveira e Araujo (2005) e Curi e Menezes-Filho (2014) argumentam que apesar da melhora do acesso ao sistema educacional a partir da promulgação da Constituição de 1988, o sistema educacional brasileiro público ainda não tinha obtido sucesso em prover uma educação de qualidade. Conforme os alunos formados em sistemas de qualidade mais baixas foram entrando no mercado de trabalho, os retornos salariais da educação foram reduzindo.

#### 4.3.2. Expansão da Mão de Obra Qualificada

No Brasil, a proporção de trabalhadores com mais escolaridade vem aumentando, conforme visto anteriormente. Essa mudança estrutural pode ser uma das forças por trás da queda nos retornos da escolaridade. Para investigar essa relação, primeiro estimamos os retornos da escolaridade para cada estado mais o Distrito Federal, por grupos de escolaridade, entre 1995 e 2023, com base na seguinte especificação:



$$\ln(Ret_{jt}) = \alpha_0 + \beta_1 \ln(prop_{jt}) + \alpha_2 D_b + \alpha_3 \ln(prop_{jt}) \times D_b + \sum_j \gamma_j C_j + \epsilon_{jt} \quad (12)$$

onde  $\ln(Ret_{jt})$  é logaritmo do retorno da escolaridade do estado  $j$  no período  $t$ ;  $\ln(prop_{jt})$  é o logaritmo da proporção de trabalhadores em cada respectivo nível de escolaridade;  $D_b$  é uma *dummy* que indica se a base de dados utilizada é a PNAD ou a PNADC;  $C_j$  são efeitos fixos de estado, e  $\epsilon_{jt}$  é um termo de erro. A interação entre  $\ln(Ret_{jt})$  e a *dummy*  $D_b$  permite que o impacto da proporção de trabalhadores no retorno da escolaridade varie dependendo da base de dados utilizada.

Uma vez que estamos interessados nos efeitos da proporção de trabalhadores sobre os retornos da escolaridade, reportamos apenas os resultados do coeficiente angular para as duas bases de dados.

$$\frac{\partial \ln(Ret_{ij})}{\partial \ln(prop_{ij})} = \beta_1 + \beta_3 D_b, \quad (13)$$

Os resultados da Tabela 3 indicam que o impacto da proporção de trabalhadores sobre os retornos da educação varia conforme o nível de escolaridade e entre as bases de dados analisadas. Constatamos que o aumento da proporção de trabalhadores no primeiro grupo de escolaridade, e em menor grau no segundo grupo, está associado a um aumento nos retornos da educação, tanto na PNAD quanto na PNADC. No entanto, essa relação não é significativa para o Grupo 2 na PNADC. Isso sugere que o aumento na oferta relativa de trabalhadores menos qualificados tende a elevar seus retornos salariais. Alternativamente, uma queda na oferta relativa de trabalhadores menos qualificados está associada a uma redução em seus retornos salariais.

Em contrapartida, os resultados mostram que, à medida que cresce a proporção de trabalhadores nos dois grupos de maior escolaridade, os retornos tendem a diminuir, independentemente da base de dados utilizada. Assim, o aumento na oferta relativa de trabalhadores mais qualificados exerce um efeito negativo sobre seus retornos salariais.

Tabela 3 - Efeitos da proporção de trabalhadores nos retornos salariais por nível de escolaridade

Escolaridade	Base de dados	Ef. Marginal
Grupo 1	PNADC	0,20***
	PNAD	0,34***
Grupo 2	PNADC	0,04
	PNAD	0,09*
Grupo 3	PNADC	-0,21***
	PNAD	-0,36***
Grupo 4	PNADC	-0,17***
	PNAD	-0,22***

Fonte: Elaborado pelos autores.

Notas: (1) A variável dependente é o retorno da escolaridade estimado pelo método de Heckman. (2) Asteriscos simples (\*), duplo (\*\*) e triplo (\*\*\*) denotam significância a 10%, 5% e 1%, respectivamente. (3) Consideramos efeitos fixos de estado. (4) O painel de dados utilizado nessas estimações possui 26 estados mais o Distrito Federal, 27 anos, e quatro níveis de escolaridade, totalizando 2916 observações. (5) Os grupos de escolaridade foram definidos conforme as seguintes categorias: PNAD - Grupo 1 (0 a 4 anos de estudo), Grupo 2 (5 a 8), Grupo 3 (9 a 11) e Grupo 4 (12 ou mais); PNADC - Grupo 1 (0 a 5 anos), Grupo 2 (6 a 9), Grupo 3 (10 a 12) e Grupo 4 (13 ou mais).

Levando em conta que a proporção de trabalhadores nos dois grupos de menor escolaridade vem caindo, como visto na Tabela 1, e que os retornos salariais dos trabalhadores nesses dois níveis foram os que mais sofreram retração no período analisado, (ver Figura 2), pelo menos parte dessa correlação positiva é proveniente da queda na proporção dos trabalhadores nesses grupos de escolaridade acompanhada de retrações mais do que proporcionais em seus retornos salariais em relação aos trabalhadores nos dois grupos de escolaridade mais elevada. Portanto, esses resultados indicam que mesmo com a relevante queda na proporção de trabalhadores que não avançaram além do segundo ciclo do fundamental, no período analisado, a demanda por trabalhadores com esse nível que qualificação caiu mais do que proporcionalmente. Esse fato decorre, pelo menos em parte, da substituição de trabalhadores menos qualificados por trabalhadores com maiores níveis de escolaridade conforme foi aumentando a oferta de trabalho qualificado.

Considerando os trabalhadores mais qualificados, há duas hipóteses que podem explicar a relação negativa entre a oferta relativa desses trabalhadores e os retornos salariais. A primeira é que os empregadores podem ofertar salários menores aos trabalhadores na mesma atividade, dado o excesso de mão de obra. Garcia (2021), por exemplo, observa elevação na oferta relativa de trabalhadores qualificados no Brasil, entre 1981 e 2014,

e conclui que esse é um elemento chave para se entender o declínio no retorno da educação. Adicionalmente, Pauli *et al.* (2012), em uma análise dos dados da RAIS entre 1985 e 2007, mostraram que o aumento expressivo no número de trabalhadores com qualificação intermediária (com níveis de ensino fundamental e médio completos) foi acompanhado por uma queda em seus salários em todos os setores da economia. Isso ocorreu durante o período de maior expansão da oferta de trabalhadores com esses níveis de qualificação, sugerindo que o aumento da escolaridade média dos trabalhadores brasileiros resultou mais da ampliação da oferta de ensino no país do que de um crescimento da demanda por parte das empresas e outras instituições.

A segunda via está associada ao fato de que trabalhadores mais qualificados estão descendo a escada ocupacional e indo para profissões que exigem menos qualificação. Isso resulta em um descompasso educacional, visto que o trabalhador tem um nível educacional maior do que aquele requerido pela profissão que ele exerce, gerando má alocação de capital humano na economia. Marioni (2021) analisa os efeitos do descompasso educacional nos salários dos trabalhadores brasileiros, concluindo que 25% dos trabalhadores formais possuem qualificação maior que a necessária em sua profissão, ou seja, são trabalhadores sobre-educados que ganham menos que os trabalhadores com mesmo nível de qualificação em empregos compatíveis com suas habilidades, resultando em queda nos retornos da educação.<sup>26</sup>

Nossos resultados apontam uma queda na proporção de trabalhadores com menores níveis de qualificação associada a uma retração em seu retorno salarial, além de um crescimento na proporção de trabalhadores mais escolarizados com queda em seu retorno salarial, o que sugere que vem ocorrendo uma substituição de trabalhadores com menos escolaridade por trabalhadores com mais escolaridade, pelo menos parcialmente, em ocupações semelhantes.

Reis (2020) encontra que entre os trabalhadores com ensino médio completo, os formados em escolas públicas têm uma maior tendência a estarem sobre-educados em comparação àqueles que frequentaram escolas privadas. Situação inversa é observada para os indivíduos com curso superior, ou seja, os ocupados com maior sobre-educação são os provenientes de faculdades ou universidades privadas. O autor argumenta que as deficiências na

<sup>26</sup> Evidências adicionais sobre o descompasso educacional no Brasil podem ser vistos em Reis (2017), Annegues e Souza (2020) e Santos *et al.* (2021).

formação de capital humano através da educação formal fazem com que os trabalhadores necessitem de mais anos de estudo para atingir determinada qualificação, algo que seria alcançado com menos tempo se a qualidade do ensino fosse satisfatória.

Embora a tendência de baixa dos retornos da escolaridade dos trabalhadores mais qualificados esteja enfraquecendo, um ambiente de negócios mais dinâmico que absorvesse a mão de obra mais escolarizada poderia mitigar ainda mais a queda, estimulando os investimentos em capital humano.

No presente estudo não abordamos se a relação negativa entre oferta de trabalhadores mais qualificados e retorno salarial é decorrente da redução salarial para cada tipo de profissão ou qualificação, ou seja, dos trabalhadores mais qualificados estarem recebendo menos dentro da própria ocupação, ou se decorre de uma má alocação desses trabalhadores. Porém, é provável que uma combinação desses dois efeitos ajude a explicar a queda no retorno da escolaridade para os trabalhadores mais qualificados, assim como problemas na qualidade dos sistemas educacionais, sobretudo públicos. Avanços nessa direção são importante linha de pesquisa futura.

## 5. Considerações Finais

Neste artigo, utilizamos dados da Pesquisa Nacional por Amostra de Domicílios (PNAD) e da PNAD Contínua (PNADC) para estimar os retornos da escolaridade do Brasil entre 1995 e 2023. O método de Mínimos Quadrados Ordinários (MQO) é bastante empregado na literatura. No entanto, ele pode gerar vieses nas estimativas. Dessa forma, também empregamos o método de Garen (1984), que corrige os vieses de habilidade, retorno e erro de medida, e o de Heckman (1979), que corrige o viés de seleção amostral.

Os resultados do presente estudo mostram que, no Brasil, os retornos salariais da educação vêm apresentando tendência de queda desde 1995. Embora existam limitações devido às diferenças entre as bases de dados, os resultados indicam que essa tendência tem se enfraquecido nos últimos anos. Também mostramos que os vieses de retorno, habilidade e erro de medida tendem a reduzir as estimativas dos retornos da escolaridade, enquanto o viés de seleção amostral tende a elevar o retorno.

Para captar os efeitos do grau de escolaridade atingido sobre os retornos da educação, agrupamos os trabalhadores de acordo com seu nível escolaridade e características como experiência, raça, zona de residência, sexo e setor de atividade. Os resultados mostram que a maior parte dos trabalhadores com mais anos de estudo são brancos, residem em áreas urbanas, trabalham no setor de serviços, possuem menos experiência no mercado de trabalho e desfrutam de retornos mais elevados.

As evidências empíricas também mostram que a tendência de queda dos retornos tem ocorrido em todos os níveis de escolaridade, mas em menor medida a partir de 2016, sobretudo para os trabalhadores dos dois grupos de escolaridade mais elevadas, ou seja, daqueles que completaram pelo menos um ano de ensino médio ou pelo menos um ano de ensino superior. Dessa forma, apesar do aumento da escolaridade e consequente elevação na proporção de trabalhadores nos dois grupos de escolaridade mais elevada, a diferença nos retornos entre os mais e menos escolarizados tem se ampliado nos últimos anos.

Apresentamos duas possíveis explicações para a tendência de queda nos retornos salariais. A primeira está relacionada ao aumento do acesso aos níveis de ensino básico e superior no Brasil, com concomitante piora em sua qualidade média. Os resultados do presente estudo sugerem que em estados onde a qualidade educacional, medida pelo SAEB e IDEB, é mais alta, os retornos da escolaridade também são mais elevados. Dessa forma, existe uma relação positiva entre qualidade do sistema de ensino e retornos da escolaridade.

A segunda explicação está associada à expansão da mão de obra qualificada. Observamos que o aumento na proporção de trabalhadores com maior escolaridade está relacionado a uma diminuição dos retornos para esses grupos, indicando um possível descompasso entre a oferta e a demanda de trabalhadores qualificados. Esse fato sugere que a absorção dessa mão de obra pelo mercado de trabalho, especialmente em funções não compatíveis com suas qualificações, é um desafio a ser enfrentado para que a educação continue a ser uma via atrativa, do ponto de vista do retorno financeiro.

Em síntese, tanto o aumento da escolaridade da força de trabalho quanto a baixa qualidade do sistema educacional brasileiro parecem ter contribuído para as trajetórias decrescentes dos retornos da escolaridade. No entanto, embora essas explicações estejam em linha com as evidências apresentadas, mais estudos são necessários para compreender melhor esse fenômeno.

## Referências

- Abiodun, O. I., Jantan, A., Omolara, A. E., Dada, K. V., Mohamed, N. A., e Arshad, H. (2018). State-of-the-art in artificial neural network applications: A survey. *Heliyon*, 4(11).
- Annegues, A. C. e Souza, W. P. S. d. F. (2020). Retorno salarial do overeducation: Viés de seleção ou penalização ao excesso de escolaridade? *Revista Brasileira de Economia*, 74:119–138.
- Asadullah, M. N. e Xiao, S. (2020). The changing pattern of wage returns to education in post-reform China. *Structural Change and Economic Dynamics*, 53:137–148.
- Bedi, A. S. e Edwards, J. H. (2002). The impact of school quality on earnings and educational returns—evidence from a low-income country. *Journal of Development Economics*, 68(1):157–185.
- Blundell, R., Dearden, L., e Sianesi, B. (2001). Estimating the returns to education: Models, methods and results. Working paper.
- Bosi, S., Lloyd-Braga, T., e Nishimura, K. (2021). Externalities of human capital. *Mathematical Social Sciences*, 112:145–158.
- Bound, J., Jaeger, D. A., e Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association*, 90(430):443–450.
- Brotherhood, L. M., Ferreira, P. C., e Santos, C. (2019). Education quality and returns to schooling: Evidence from migrants in Brazil. *Economic Development and Cultural Change*, 67(3):439–459.
- Card, D. (2001). Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica*, 69(5):1127–1160.
- Card, D. e Krueger, A. B. (1992). Does school quality matter? returns to education and the characteristics of public schools in the United States. *Journal of Political Economy*, 100(1):1–40.
- Cetra, G., Nakabashi, L., e Pazello, E. T. (2023). Qualidade da educação e salários nos municípios brasileiros. *Pesquisa e Planejamento Econômico*.
- Colclough, C., Kingdon, G., e Patrinos, H. (2010). The changing pattern of wage returns to education and its implications. *Development Policy Review*, 28(6):733–747.
- Curi, A. Z. e Menezes-Filho, N. (2014). The relationship between school performance and future wages in Brazil. *Economia*, 15(3):261–274.
- Deming, D. J. (2022). Four facts about human capital. *Journal of Economic Perspectives*, 36(3):75–102.
- Garcia, P. (2021). Skill prices and compositional effects on the declining wage inequality in Latin America: Evidence from Brazil. *Revista Brasileira de Economia*, 75(2).
- Garen, J. (1984). The returns to schooling: A selectivity bias approach with a continuous choice variable. *Econometrica: Journal of the Econometric Society*, pages 1199–1218.
- Gollin, D., Lagakos, D., e Waugh, M. E. (2014). The agricultural productivity gap. *The Quarterly Journal of Economics*, 129(2):939–993.
- Hanushek, E. A. e Woessmann, L. (2008). The role of cognitive skills in economic development. *Journal of economic literature*, 46(3):607–668.
- Heckman, J., Tobias, J. L., e Vytlačil, E. (2003). Simple estimators for treatment parameters in a latent-variable framework. *Review of Economics and Statistics*, 85(3):748–755.
- Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica: Journal of the Econometric Society*, pages 153–161.
- Horie, N. e Iwasaki, I. (2022). Returns to schooling in European emerging markets: A meta-analysis. *Education Economics*.

- IBGE (2015a). Pesquisa nacional por amostra de domicílios 2015: Notas metodológicas. Nota Técnica. Disponível em: <http://www.ibge.gov.br>.
- IBGE (2015b). Pesquisa nacional por amostra de domicílios contínua: Notas metodológicas. Nota Técnica. Disponível em: <http://www.ibge.gov.br>.
- Kassouf, A. L. (1997). Retornos à escolaridade e ao treinamento nos setores urbano e rural do Brasil. *Revista de Economia e Sociologia Rural*, 35(2):59–76.
- Lumley, T. (2004). Analysis of complex survey samples. *Journal of Statistical Software*, 9(1):1–19.
- Marcelo, R. e Wyllie, R. (2006). Retornos para educação no Brasil: Evidências empíricas adicionais. *Economia Aplicada*, 10:349–365.
- Marioni, L. d. S. (2021). Overeducation in the labour market: Evidence from Brazil. *Education Economics*, 29(1):53–72.
- Menezes Filho, N. A. (2003). A evolução recente da educação no Brasil. *Tese de Livre Docência*, Universidade de São Paulo.
- Mincer, J. (1958). Investment in human capital and personal income distribution. *Journal of Political Economy*, 66(4):281–302.
- Mincer, J. (1974). Schooling, experience, and earnings. *Human Behavior & Social Institutions*, (2).
- Monte, P. A. d., Ramalho, H. M. d. B., e Pereira, M. d. L. (2011). O salário de reserva e a oferta de trabalho: Evidências para o Brasil. *Economia Aplicada*, 15(4):613–639.
- Montenegro, C. E. e Patrinos, H. A. (2014). Comparable estimates of returns to schooling around the world. *World Bank Policy Research Working Paper*, (7020).
- Moura, R. L. d. (2008). Testando as hipóteses do modelo de mincer para o Brasil. *Revista Brasileira de Economia*, 62:407–449.
- Nakabashi, L. e Assahide, L. (2017). Estimando o retorno da escolaridade dos jovens por classe de renda: 1997–2012. *Pesquisa e Planejamento Econômico–PPE, Brasília*, 47(3):137–83.
- Oliveira, R. P. d. e Araujo, G. C. d. (2005). Qualidade do ensino: uma nova dimensão da luta pelo direito à educação. *Revista brasileira de educação*, pages 5–23.
- Pauli, R. C. d., Nakabashi, L., e Sampaio, A. V. (2012). Mudança estrutural e mercado de trabalho no Brasil. *Brazilian Journal of Political Economy*, 32:459–478.
- Pereira, V. d. F., Lima, J. E. d., Lima, J. R. F. d., Braga, M. J., e Mendonça, T. G. d. (2013). Avaliação dos retornos à escolaridade para trabalhadores do sexo masculino no Brasil. *Revista de Economia Contemporânea*, 17:153–176.
- Perez-Silva, R. e Partridge, M. D. (2020). Concentration of human capital, externalities and the wage gap in us metro areas. *Regional Studies*, 54(11):1564–1573.
- Psacharopoulos, G. e Patrinos, H. A. (2018). Returns to investment in education: A decennial review of the global literature. *Education Economics*, 26(5):445–458.
- Reis, M. C. (2017). Educational mismatch and labor earnings in Brazil. *International Journal of Manpower*.
- Reis, M. C. (2020). Os ensinos público e privado no Brasil e a incidência de sobre-educação no mercado de trabalho. *Economia Aplicada*, 24(3):367–392.
- Sachsida, A., Loureiro, P. R. A., e Mendonça, M. J. C. d. (2004). Um estudo sobre retorno em escolaridade no Brasil. *Revista Brasileira de Economia*, 58:249–265.
- Santiago, L., Botassio, D. C., e Teixeira, E. C. (2023). Uma nota sobre o retorno social da educação profissional no Brasil. *Revista Brasileira de Economia*, 77(4).
- Santos, M. M. d., Mariano, F. Z., Arraes, R. d. A., e Oliveira, C. S. d. (2021). A armadilha da sobre-educação no primeiro emprego: evidências para o Brasil. *Estudos Econômicos (São Paulo)*, 51:415–452.

- Silva, T. B. d., Bondezan, K. d. L., e Lucas, M. d. S. (2022). Retornos para educação no Brasil: enfoque nas regiões sul e nordeste para o ano de 2017. *Interações (Campo Grande)*, 23:67–83.
- Souza, W., Figueiredo, E., Annegues, A. C., e Stampe, M. Z. (2022). Viés da loglinearização: estimando o retorno da educação através de regressões quantílicas. *Economia Aplicada*, 26(1):101–116.
- Suliano, D. C. e Siqueira, M. L. (2012). Retornos da educação no Brasil em âmbito regional considerando um ambiente de menor desigualdade. *Economia Aplicada*, 16(1):137–165.
- Van Zaist, J. K., Nakabashi, L., e Salvato, M. A. (2010). Retornos privados da escolaridade no Paraná. *Revista Economia*.
- Vaz, B. O. E. (2017). Educação, setores de atividade e produtividade. In Bonelli, R., Veloso, F., e Pinheiro, A. C., editors, *Anatomia da produtividade*, chapter 15, pages 405–427. Elsevier.
- Vaz, B. O. E. e Barreira, T. C. (2021). Metodologia de retropolação da pesquisa nacional por amostra de domicílios contínua de 1992 a 2012. *Estudos Econômicos (São Paulo)*, 51(04):759–782.
- Wu, Y.-c. e Feng, J.-w. (2018). Development and application of artificial neural network. *Wireless Personal Communications*, 102:1645–1656.
- Zhang, J., Zhao, Y., Park, A., e Song, X. (2005). Economic returns to schooling in urban China, 1988 to 2001. *Journal of Comparative Economics*, 33(4):730–752.

## Apêndice A - Comparação dos Retornos da Escolaridade nos anos que a PNAD e PNADC Coexistem

A PNAD foi iniciada em 1967 e encerrada em 2015. A ideia era substituí-la pela PNAD Contínua (PNADC) que foi iniciada em 2012. vigora até a presente data, e foi elaborada com uma metodologia atualizada.<sup>27</sup> Sendo assim, de 2012 a 2015 a PNAD e PNADC coexistiram. A longo do texto, apresentamos as estimações desse período apenas com base na PNAD. Nesta seção, preenchemos essa lacuna e apresentamos os retornos da escolaridade para esse intervalo de tempo com base em ambos os conjuntos de dados.

Na Figura A1 apresentamos uma comparação entre os retornos da escolaridade, no nível agregado, estimado com os três métodos (MQO, Garen e Heckman) e ambas as PNADs. O modelo de MQO é o que apresentou menos divergência nas estimações. Por outro lado, o método de Garen apresentou maior discrepância. A Tabela A1 apresenta as estatísticas descritivas dessas estimações. Notamos que em média o método de Garen e de Heckman estimados pela PNAD e PNADC divergem em 1%. Essas divergências podem ser atribuídas as diferenças metodológicas entre a PNAD e PNADC que discutimos na Seção 3.

<sup>27</sup> Informações adicionais podem ser consultadas no site do IBGE: <https://www.ibge.gov.br>.



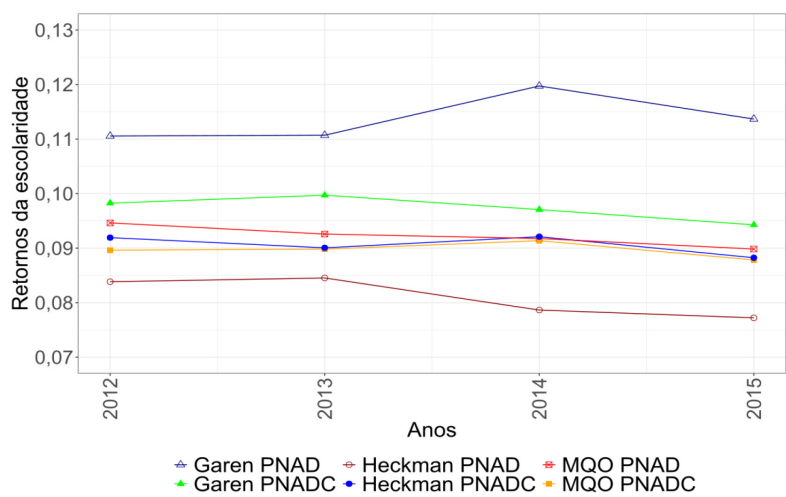


Figura A1 - Comparação entre retornos da escolaridade estimados com base na PNAD e PNADC - 2012:2015

Fonte: Elaborado pelos autores.

Tabela A1 - Estatísticas descritivas dos retornos da escolaridade - 2012:2015

Método/ Base de dados	Mínimo	Q25	Mediana	Q75	Máximo	Média	Desv. p.
Garen PNAD	0,11	0,11	0,11	0,12	0,12	0,11	0,004
Garen PNADC	0,09	0,10	0,10	0,10	0,10	0,10	0,002
Heckman PNAD	0,08	0,08	0,08	0,08	0,08	0,08	0,003
Heckman PNADC	0,09	0,09	0,09	0,09	0,09	0,09	0,001
MQO PNAD	0,09	0,09	0,09	0,09	0,09	0,09	0,001
MQO PNADC	0,09	0,09	0,09	0,09	0,09	0,09	0,001

Fonte: Elaborado pelos autores.

Nota: Q25: primeiro quartil; Q75: terceiro quartil.

## Apêndice B - Estimação das Variáveis Indisponíveis na PNADC

### B.1. Redes Neurais

Para estimar as variáveis presença da mãe viva e a existência de filhos, que não estão disponíveis na PNAD Contínua (PNADC), utilizamos redes neurais artificiais (RNAs).<sup>28</sup> Inspiradas no funcionamento do cérebro humano, as RNAs são compostas por múltiplos nós interconectados, conhecidos como neurônios. Cada um desses nós executamos uma função de saída específica, chamada de função de ativação. As conexões entre os nós possuem pesos associados, que indicam a força do sinal transmitido, funcionando de forma análoga à memória na rede. A saída de uma RNA é determinada pela configuração da rede, pelos valores dos pesos e pelas funções de ativação utilizadas (Wu e Feng 2018).

Considere que o vetor  $Z$  representa as  $N$  características do indivíduo e  $y \in \{0, 1\}$  contém os valores observados de uma variável binária. As camadas de entrada da rede transmitem, com peso  $W$ , as características para a camada de saída. A função linear  $XZ = \sum_{i=1}^N w_i z_i$  é utilizada para prever a variável dependente  $y$  que é computada da seguinte forma:<sup>29</sup>

$$\hat{y} = f(XZ) = f\left(\sum_{i=1}^N w_i z_i\right), \quad (14)$$

onde  $f(x) = 1/(1 + e^{-x})$  é a função sigmoide que exerce o papel de função de ativação e retorna um número entre 0 e 1.

O objetivo da rede neural é encontrar o vetor de pesos  $W$  que minimiza a distância entre os valores observados e previstos, ou seja:

$$L = \sum (y - \hat{y})^2 = \sum (y - f(WX))^2 \quad (15)$$

Esse tipo de função objetivo é conhecido como função de perda e se assemelha muito à função de perda do método de mínimos quadrados ordinários (MQO).

<sup>28</sup> As RNAs são uma das técnicas mais utilizadas em Aprendizado de Máquina, sendo aplicadas em diversas áreas, como grandes modelos de linguagem (LLMs), reconhecimento de padrões em imagens e previsão de variáveis como preços, clima, e, características categóricas que definem indivíduos ou grupos (Abiodun *et al.* 2018).

<sup>29</sup> Para simplificar a exposição consideramos uma RNA com apenas uma camada e viés de neurônio igual a zero.

## B.2. Configuração e Acurácia da RNA

Para estimar a *dummy Filhos<sub>i</sub>* utilizamos na camada de entrada da RNA as seguintes variáveis: *Raça<sub>i</sub>* assume o valor 1 para indivíduos brancos e 0 para os demais; *Urb<sub>i</sub>* toma o valor 1 para aqueles que residem em áreas urbanas e 0 para áreas rurais; *Idade<sub>i</sub>* que se refere a idade do indivíduo; também incluímos um conjunto de *dummies* que indicam a região de residência (Sudeste, Nordeste, Centro-Oeste e Sul). Para estimar a *dummy Mãe<sub>i</sub>* utilizamos o mesmo conjunto de variáveis acrescido da variável anos de estudo. Os dados foram extraídos das PNADs de 1995 a 2015, totalizando mais de 1,5 milhões de observações.

Em nossa estimação, utilizamos uma RNA do tipo *feedforward*, com uma camada oculta e função de ativação sigmoide, o que gera saídas entre 0 e 1, sendo essa configuração especialmente adequada para problemas de classificação. Para identificar o modelo com a melhor acurácia para cada variável, adotamos o método de validação cruzada *10-fold*.<sup>30</sup> Além disso, para determinar o número ideal de neurônios na camada oculta, empregamos o método de busca em grid, testando diferentes configurações, variando de 5 a 25 neurônios. A configuração que apresentou os melhores resultados, em ambos os casos, foi a de 15 neurônios. A acurácia média da RNA foi de 69% na estimativa da variável *Filhos<sub>i</sub>* e de 75% na variável *Mãe<sub>i</sub>*.

## Apêndice C - Estatísticas Descritivas

### C.1. Retornos da Escolaridade por Níveis de Educacionais

Na Tabela C1 apresentamos as estatísticas descritivas dos retornos da escolaridade estimados com base na PNAD e PNADC, por níveis de escolaridade. Como enfatizamos, os maiores níveis educacionais possuem retornos da escolaridade mais elevados. Por exemplo, pelo método de Garen, a média dos retornos para quem tem a partir de 12 anos de estudo é 2,3 vezes maior do que aquela dos trabalhadores com Grupo 1 anos de estudo, além disso, possuem maior desvio padrão. Também verificamos que o

<sup>30</sup> O método de validação cruzada *10-fold* consiste em dividir o conjunto de dados aleatoriamente em 10 partes (*folds*) iguais, treinar o modelo utilizando 9 dessas partes e avaliar sua acurácia na parte restante. Esse procedimento é repetido 10 vezes, de modo que cada *fold* seja utilizado uma vez como conjunto de teste e as outras 9 como conjunto de treino. Ao final dessas 10 iterações, a acurácia do modelo é calculada como a média dos resultados obtidos em cada *fold*.

retorno da escolaridade pelo método de Garen é maior do que aquele via MQO a partir do Grupo 3, indicando uma mudança no viés gerado pelo método MQO.

**Tabela C1 - Estatísticas descritivas dos retornos da escolaridade - 1995:2023**

PNAD - 1995:2015								
Escolaridade	Método	Mínimo	Q25	Mediana	Q75	Máximo	Média	Desv. p.
Grupo 1	Garen	0,06	0,07	0,08	0,09	0,10	0,08	0,01
Grupo 1	Heckman	0,05	0,06	0,08	0,08	0,09	0,07	0,01
Grupo 1	MQO	0,06	0,07	0,09	0,10	0,11	0,09	0,02
Grupo 2	Garen	0,10	0,11	0,11	0,12	0,13	0,11	0,01
Grupo 2	Heckman	0,07	0,08	0,10	0,11	0,12	0,09	0,02
Grupo 2	MQO	0,08	0,09	0,11	0,12	0,13	0,11	0,02
Grupo 3	Garen	0,13	0,14	0,15	0,15	0,16	0,14	0,01
Grupo 3	Heckman	0,09	0,10	0,11	0,12	0,13	0,11	0,01
Grupo 3	MQO	0,10	0,11	0,12	0,13	0,14	0,12	0,01
Grupo 4	Garen	0,16	0,17	0,18	0,18	0,19	0,18	0,01
Grupo 4	Heckman	0,10	0,11	0,12	0,13	0,14	0,12	0,01
Grupo 4	MQO	0,11	0,12	0,13	0,14	0,14	0,13	0,01
PNADC - 2016:2023								
Escolaridade	Método	Mínimo	Q25	Mediana	Q75	Máximo	Média	Desv. p.
Grupo 1	Garen	0,03	0,03	0,04	0,05	0,06	0,04	0,01
Grupo 1	Heckman	0,03	0,04	0,04	0,05	0,05	0,04	0,01
Grupo 1	MQO	0,04	0,04	0,05	0,06	0,06	0,05	0,01
Grupo 2	Garen	0,06	0,06	0,07	0,08	0,09	0,07	0,01
Grupo 2	Heckman	0,06	0,06	0,06	0,07	0,07	0,06	0,01
Grupo 2	MQO	0,07	0,07	0,07	0,08	0,08	0,07	0,01
Grupo 3	Garen	0,10	0,10	0,10	0,11	0,13	0,11	0,01
Grupo 3	Heckman	0,08	0,08	0,09	0,09	0,09	0,08	0,00
Grupo 3	MQO	0,09	0,09	0,10	0,10	0,10	0,10	0,00
Grupo 4	Garen	0,13	0,14	0,14	0,15	0,16	0,14	0,01
Grupo 4	Heckman	0,10	0,10	0,10	0,10	0,10	0,10	0,00
Grupo 4	MQO	0,11	0,11	0,12	0,12	0,12	0,12	0,00

Fonte: Elaborado pelos autores.

Notas: (1) Q25: primeiro quartil; Q75: terceiro quartil. (2) Os grupos de escolaridade foram definidos conforme as seguintes categorias: PNAD: Grupo 1 (0 a 4 anos de estudo), Grupo 2 (5 a 8), Grupo 3 (9 a 11) e Grupo 4 (12 ou mais); PNADC: Grupo 1 (0 a 5 anos), Grupo 2 (6 a 9), Grupo 3 (10 a 12) e Grupo 4 (13 ou mais).

*C.2. Médias das Variáveis Utilizadas na Pesquisa***Tabela C2 - Médias das variáveis utilizadas na pesquisa - 1995:2023**

PNAD - 1995:2023												
Anos	Y	S	Exp	Filhos	Raça	Sind	Urb	Sexo	Mãe	Agr	Ind	Serv
1995	7,1	6,0	26,9	59,7	55,7	19,6	86,1	65,9	69,0	13,1	25,5	61,4
1996	7,7	6,3	26,8	57,8	56,5	19,0	86,1	65,3	69,0	12,8	25,0	62,2
1997	7,6	6,3	26,7	57,1	55,7	18,3	85,9	65,6	69,4	12,8	25,2	62,1
1998	7,5	6,5	26,8	55,6	55,4	17,9	85,6	65,1	69,5	12,0	25,3	62,7
1999	6,9	6,6	26,8	54,6	55,5	17,5	85,6	64,9	69,5	12,4	24,6	63,0
2001	6,7	6,8	26,6	52,6	53,8	16,8	89,3	63,9	70,4	10,9	24,5	64,6
2002	6,3	7,0	26,6	51,3	53,4	17,5	89,5	63,6	70,7	10,8	26,3	62,9
2003	6,4	7,1	26,6	50,0	52,2	18,0	88,9	63,5	70,9	11,4	25,6	63,1
2004	6,3	7,2	26,5	49,3	51,3	18,0	87,4	63,2	70,9	12,2	25,6	62,2
2005	6,4	7,3	26,4	48,2	49,5	18,4	87,3	62,7	71,3	11,7	25,6	62,7
2006	6,8	7,5	26,3	47,6	49,1	18,4	87,6	62,1	71,5	11,1	25,7	63,3
2007	7,2	7,6	26,3	46,1	48,2	17,0	87,9	61,9	71,2	10,7	26,1	63,2
2008	7,0	7,8	26,3	44,1	47,3	17,5	87,9	61,5	71,1	10,4	26,6	63,0
2009	7,2	8,0	26,2	43,6	46,9	17,0	87,9	60,9	72,0	10,4	26,1	63,5
2011	9,0	8,1	26,2	42,1	46,6	15,9	89,1	61,2	72,0	10,2	25,3	64,5
2012	9,3	8,4	26,2	40,8	44,8	15,7	89,2	60,6	72,3	9,4	25,9	64,7
2013	10,4	8,5	26,3	39,8	44,7	14,8	89,1	60,5	72,1	9,1	26,2	64,7
2014	9,7	8,6	26,4	38,8	43,9	15,3	89,2	59,9	71,9	8,9	25,8	65,4
2015	10,5	8,8	26,5	38,0	43,2	17,7	88,6	60,0	72,1	9,2	24,8	66,0
PNADC - 2016:2023												
Anos	Y	S	Exp	Filhos	Raça	Sind	Urb	Sexo	Mãe	Agr	Ind	Serv
2016	7,2	9,2	26,2	43,9	43,6	13,7	76,5	62,0	80,6	17,9	26,2	55,9
2017	7,3	9,4	26,0	43,3	43,4	13,6	76,6	61,2	80,6	17,5	25,5	56,9
2018	7,5	9,7	25,9	42,0	43,2	11,8	77,2	60,7	80,1	16,8	25,1	58,1
2019	7,4	9,8	26,0	41,1	42,5	10,6	77,2	60,7	79,8	16,6	25,2	58,1
2020	7,3	10,1	26,0	38,7	43,8	18,6	77,5	61,3	79,3	17,3	24,6	58,0
2021	6,7	10,1	26,3	38,1	43,4	19,4	77,2	61,2	78,5	18,0	25,4	56,5
2022	7,5	10,1	26,3	38,0	42,7	9,0	77,2	60,6	78,2	17,0	25,2	57,7
2023	7,8	10,3	25,9	37,6	42,4	7,9	77,1	60,1	78,7	16,1	24,7	59,1

Fonte: Elaborado pelos autores.

Notas: (1) Y: Salário hora médio; S: Média dos anos de estudo; Exp: Experiência; Filhos: % de indivíduos com filhos menores de 14 anos; Raça: % de indivíduos brancos; Sind: % de trabalhadores sindicalizados; Urb: % de Moradores na área urbana; Sexo: % de Homens; Mãe: % de indivíduos que tem a mãe viva; Agr: % de trabalhadores da agricultura; Ind: % de trabalhadores da indústria; Serv: % de trabalhadores dos serviços. (2) Para o período de 2016 a 2023 estimamos as variáveis Mãe e Filhos utilizando redes neurais.

## DECLARAÇÃO DE DISPONIBILIDADE DE DADOS

Os dados utilizados neste estudo estão disponíveis mediante solicitação ao autor. Dados adicionais e informações complementares também poderão ser fornecidos para fins de verificação ou replicação. A disponibilização está condicionada à inexistência de restrições de acesso público.

## ♦ AGRADECIMENTOS

Agradecemos aos pareceristas anônimos e aos participantes do 50º Encontro Nacional de Economia pelos comentários e sugestões. Possíveis erros são nossos. Marcos Ribeiro agradece a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pelo suporte financeiro número 2024/06763-0. Luciano Nakabashi agradece ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo suporte financeiro número 305037/2023-4.

## CONTRIBUIÇÕES DE AUTORIA

**MR:** Conceitualização, Curadoria de dados, Análise formal, Aquisição de financiamento, Investigação, Metodologia, Programas, Supervisão, Validação, Visualização, Escrita - rascunho original e Escrita - revisão e edição.

**LN:** Conceitualização, Análise formal, Investigação, Metodologia, Supervisão, Validação, Visualização, Escrita - rascunho original e Escrita - revisão e edição.

**FJ:** Conceitualização, Análise formal, Investigação, Metodologia, Supervisão, Validação, Visualização.

## CONFLITO DE INTERESSE

Os autores declaram não terem quaisquer conflitos de interesse.

## EDITOR-CHEFE

Dante Mendes Aldrighi  <https://orcid.org/0000-0003-2285-5694>

Professor - Department of Economics University of São Paulo (USP)

## EDITOR ASSOCIADO

Fábio Waltenberg  <https://orcid.org/0000-0003-3404-7424>