The current issue and full text archive of this journal is available on Emerald Insight at:
https://www.emerald.com/insight/1517-7580.htm

Smoking effects on labor income: new evidence for Brazil

Leandro Pinheiro Vieira and Rafael Mesquita Pereira
Postgraduate Program in Applied Economics, Federal University of Rio Grande, Rio Grande, Brazil

Received 16 May 2023
Revised 3 November 2023
9 December 2023
Accepted 9 January 2024

Abstract

Purpose – This study aims to investigate the effect of smoking on the income of workers in the Brazilian labor market.
Design/methodology/approach – Using data from the 2019 National Health Survey (PNS), we initially address the sample selection bias concerning labor market participation by using the Heckman (1979) method. Subsequently, the decomposition of income between smokers and nonsmokers is analyzed, both on average and across the earnings distribution by employing the procedure of Firpo, Fortin, and Lemieux (2009) - FFL decomposition. Ñopo (2008) technique is also used to obtain more robust estimates.
Findings – Overall, the findings indicate an income penalty for smokers in the Brazilian labor market across both the average and all quantiles of the income distribution. Notably, the most significant differentials and income penalties against smokers are observed in the lower quantiles of the distribution. Conversely, in the higher quantiles, there is a tendency toward a smaller magnitude of this gap, with limited evidence of an income penalty associated with this habit.
Research limitations/implications – This study presents an important limitation, which refers to a restriction of the PNS (2019), which does not provide information about some subjective factors that also tend to influence the levels of labor income, such as the level of effort and specific ability of each worker, whether smokers or not, something that could also, in some way, be related to some latent individual predisposition that would influence the choice of smoking.
Originality/value – The relevance of the present study is clear in identifying the heterogeneity of the income gap in favor of nonsmokers, as in the lower quantiles there was a greater magnitude of differentials against smokers and a greater incidence of unexplained penalties in the income of these workers, while in the higher quantiles, there was low magnitude of the differentials and little evidence that there is a penalty in earnings since the worker is a smoker.

Keywords Smoking, Labor market, Brazil, Heckman model, FFL decomposition, Ñopo decomposition
Paper type Research paper

1. Introduction

While there has been a decline in the prevalence of smoking among the Brazilian population in recent decades, data from the 2019 National Health Survey (PNS) reveals that the smoking rate stands at 12.6%, with 11.4% being daily tobacco smokers. The prevalence of male smokers is reported at 15.9% (14.3% of whom are daily smokers), while for females, the rates are 9.6% (8.8% being daily smokers) (IBGE, 2020).

Grossman (1972) in the framework of his demand model for good health, based on the theory of human capital, proposed that health can be understood as a durable capital stock that generates healthy labor time (market activities) and entertainment (non-market activities). Grossman's model assumes that individuals inherit an initial health stock, subject to depreciation over time, but that this stock can be augmented through investments in

JEL Classification — C21, I12, J24, J31

EconomiA
Emerald Publishing Limited
e-ISSN: 2358-2820
p-ISSN: 1517-7580
DOI 10.1108/ECON-05-2023-0077

human capital, such as adopting healthy habits and seeking medical care. Conversely, the adoption of unhealthy behaviors, such as smoking, can lead to the depreciation of this health stock over the course of one's life, impacting productivity and labor income.

Several studies, such as Levine, Gustafson, and Velenchik (1997), Heineck and Schwarze (2003), Van Ours (2004), Auld (2005), Munasinghe and Sicherman (2006), Lokshin (2006), Grafova and Stafford (2009), Anger and Kvasnicka (2010), Hotchkiss and Pitts (2013), Bhai (2020), Darden, Hotchkiss, and Pitts (2021), examined the consequences of smoking on indicators of the labor market through the microeconomic perspective of an individual return considering it as a harmful habit to health. These studies support the hypothesis that there is an indirect monetary cost of smoking on individual earnings. Levine, Gustafson, and Velenchik (1997), for example, calculated wage losses ranging from 4% to 8% for North American workers.

While the subject has been extensively explored in existing literature, there is a limited number of contributions within the Brazilian context, notably by Almeida and Araújo Júnior (2017), Justus, Sant'Anna, Davanzo, and Moreira (2019), and Uhr, Parfitt, Uhr, and Ely (2021). These studies consistently suggest that smokers face wage penalties in the Brazilian labor market, albeit to varying degrees.

Considering this context, this paper seeks to examine the impact of smoking on income in the Brazilian labor market. To achieve this, we initially assess the influence of smoking on income by employing a modified version of the equation proposed by Mincer (1974), incorporating a correction for self-selection bias using Heckman's model (1979). Then we counterfactually analyze the income differentials between smokers and nonsmokers, using the decomposition methods of Firpo, Fortin, and Lemieux (2009) – FFL decomposition - and Ñopo (2008). The database used is the National Health Survey (PNS) 2019, a household-based survey conducted by the Brazilian Institute of Geography and Statistics (IBGE, 2021).

The findings indicate that smokers experience income penalties as a result of their smoking habit, with this disparity being more pronounced in the lower quantiles of the income distribution. Consequently, this study distinguishes itself from others in the literature by not only examining the mean income gap between smokers and nonsmokers but also by assessing this difference across various quantiles of the income distribution. This approach allows for an evaluation of the trends and nuances of this disparity within different income strata among Brazilian workers, facilitated by both the FFL procedure and the technique proposed by Ñopo.

In addition to this introduction, the theoretical framework is presented in section 2, and the methodological procedures in section 3. Then, we present and analyze the study's results and, finally, the final considerations.

2. Theoretical and empirical review

Works such as Strauss and Thomas (1998), Zarkin, French, Mroz, and Bray (1998), and Munasinghe and Sicherman (2006) agree that health and behaviors associated with it, like consuming alcoholic beverages or smoking, as well as schooling and experience, are forms of human capital. These choices are expected to be related to success in the labor market, as they would affect income or the results of work. In this context, these authors, among others, proposed an adaptation of the classic Mincerian equation (Mincer, 1974) by inserting the element of health as part of human capital.

According to Almeida and Araújo Júnior (2017), the adapted Mincer expression was represented, in a generic way, by:

$$W_i = w\left(K_i^H, K_i, X_i\right) + \xi_i \tag{1}$$

in which: W_i is the logarithm of the wage; K_i^H is the vector of measures of the element health of the capital human; K_i is the vector of measures not associated with human capital health (such as schooling and experience); X_i is the vector of other covariates (age, race, gender, etc.); ξ_i is the error term.

Anger and Kvasnicka (2010) reported that empirical studies investigating the impact of tobacco on labor income consistently identified significant wage penalties associated with smoking, ranging from 2% to 24%. According to Almeida and Araújo Júnior (2017), the literature highlights several factors that may elucidate the mechanisms through which smoking can adversely impact income or salary in the labor market. These factors include the disruption of manual execution tasks due to cigarette consumption, an increase in employer-relative costs, workplace discrimination, the inclination of smokers towards jobs with health insurance benefits rather than higher remuneration, and a high intertemporal discount rate leading to low investments in human capital. Justus *et al.* (2019) also corroborate that the negative association between smoking and income from work can occur through increased absenteeism, reduced productivity, and discrimination in the labor market.

The international literature [1] is rich in empirical studies on the effects of smoking on the labor market. One of the pioneering studies to verify the association between wages and smoking in the labor market was made by Leigh and Berger (1989). They adopted the classical regression model and a set of control variables to search for statistical associations between smoking and being overweight and current wage, using a national probability sample of American workers with weekly work hours of 20 hours or more (Quality of Employment Survey (QES) data from 1973 for the United States of America). As a result, no strong statistical associations were found between smoking or being overweight and annual earnings.

Levine *et al.* (1997) examined the effect of smoking on wages and employment in the US. They used data from the National Longitudinal Survey of Youth (NLSY) in 1984 and 1991 and implemented methods to account for differences in observed and unobserved individual characteristics across siblings, which may be correlated with both smoking and wages, to address the potential endogeneity problem. Although no robust and statistically significant effect on employment was observed, all alternative specification estimates indicated that smoking reduces wages by about 4 to 8%.

Auld (2005) presented estimates of likelihood in a system of limited dependent variables to investigate the relationship between Canadian wage patterns and smoking and drinking alcohol habits. Using data from the General Social Survey (GSS) from 1985 and 1991, the main findings were that smoking is associated with greater effects on income than drinking: estimates showed that smokers would receive 8% less than nonsmokers, and the wage penalty for smokers would be 24% with control for endogeneity.

Grafova and Stafford (2009) examined the existence of a wage differential between smokers and nonsmokers. Utilizing data from the Panel Study of Income Dynamics (PSID) spanning 1986, 1999, and 2001, and categorizing the sample into groups based on smoking history, they discerned a wage difference ranging from 8% to 12% between continuing smokers and three other groups: individuals who eventually quit smoking, those who had already quit, and those who had never smoked.

Hotchkiss and Pitts (2013) found that smokers receive lower wages by about 24%, two-thirds of this differential was explained by differences in observable characteristics between groups. Their analysis included a decomposition of the wage gap between smokers and nonsmokers in the United States of America with data from 1992 to 2011 using the correction of the Heckman selection model.

Bhai (2020), using data from the National Survey of Midlife Development or MIDUS, which contains information from a nationally representative sample of Americans, as well as a subsample of twins and singletons, noted that estimates from intrafamily models show that smokers earn approximately 15% to 16% less than nonsmokers. Furthermore, Darden, Hotchkiss, and Pitts (2021), based on data from the National Longitudinal Survey of Youth - 1997 Cohort (NLSY97), a nationally representative sample from the USA, showed that maintaining heavy smoking in adulthood results in a wage penalty at age 30 of 15.9% for women and 15.2% for men, so that the contemporary effect of heavy smoking, without any effects throughout the life cycle, explains 62.9% of the disparity wage against female smokers, however, explains only 20.4% of this gap against male smokers.

On the other hand, in the Brazilian literature, the effects of cigarette consumption on labor productivity are a theme that is still incipient. Almeida and Araújo Júnior (2017) stand out because they investigated the heterogeneity of the repercussions of unhealthy personal habits, expressed by smoking cigarettes on labor productivity. The authors developed empirical models applying the conditional quantile regression with instrumental variables (QRIV) using data from the health supplement of the National Household Sample Survey (PNAD) 2008. The results showed that smokers, regardless of the models conditioned for the mean or by quantile with and without instrumental variables, presented lower work performance, with a wage penalty for smokers with control for endogeneity ranging from 15.2% to 36.5% over the conditional distribution of individual incomes.

Justus *et al.* (2019) demonstrate that, in comparison to nonsmokers, men and women who smoke experience income penalties of 29.7% and 24.2%, respectively, with a significant portion of this disparity attributed to observable characteristics. The study utilized data from PNAD 2008, employing the Heckman procedure to address sample selectivity bias and the Oaxaca-Blinder method for decomposing the wage gap between smokers and nonsmokers.

Finally, Uhr *et al.* (2021), using data from the 2013 National Health Survey (PNS), estimated the smoking effect on the productivity of Brazilian workers and sports practice by using propensity score methods. In general, the main results showed that smoking negatively affects the productivity of Brazilian workers, being the individuals most affected are those of middle and older age. Still, according to the authors, tobacco use causes a decrease in physical activity among all age groups.

3. Methodological procedures

We employ the Heckman (1979) model to estimate the labor income equation, addressing potential issues of sample selectivity inherent in this type of estimation. According to the author, the said sampling selection bias may result from two reasons:

(1) There might be self-selection among the individuals or data units under investigation;

(2) The sample selection decisions made by analysts or data processors operate similarly to self-selection.

In econometric terms, the labor income equation presented below serves as the initial step in the Heckman (1979) model for correcting sample selection bias related to labor market participation:

$$W_i = x_i'\beta + \varepsilon_i \tag{2}$$

in which W_i represents a labor payment (wage or income, for example), x_i' corresponds to variables observed associated with productivity of the *i-th* individual, and ε_i is the error term. W is observed only for workers, that is, only those who are paid for their labor.

Thus, according to Cameron and Trivedi (2005), the participation in the sample is represented as follows:

$$W_{1i} = \begin{cases} 1\, if\ W_{1i}^* > 0 \\ 0\, if\ W_{1i}^* \leq 0 \end{cases}$$

resulting in the following truncated equation:

$$W_{2i} = \begin{cases} W_{2i}^* \, if \, W_{1i}^* > 0 \\ . \, if \, W_{1i}^* \leq 0 \end{cases}$$

Cameron and Trivedi (2005) argue that this model specifies that W_2 is observed when $W_1^* > 0$, implying that W_2 does not require any meaningful value when $W_1^* \leq 0$. To have W_2 positive, the density observed is $f^*(W_2^* | W_1^* > 0) \times \Pr[W_1^* > 0]$. Therefore, the selection equation presents the likelihood function:

$$L = \prod_{i=1}^{n} \{\Pr[W_{1i}^* \leq 0]\}^{1-W_{1i}} \{f(W_2^* | W_{1i}^* > 0) \times \Pr[W_{1i}^* > 0]\}^{W_{1i}} \quad (3)$$

in which the first term corresponds to the discrete contribution when $W_{1i}^* \leq 0$, since $W_{1i} = 0$, and the second term represents the continuous contribution when $W_{1i}^* > 0$.

Once the first step (sample selection equation) is completed, we move on to the second stage by estimating the labor income equation using the OLS [2] method. Cirino and Lima (2012) said that the income expression was based on the human capital theory, including more variables to control income discrepancies arising from agents' personal and productive characteristics, in addition to the labor market inclusion (information factor generated in the first stage of the model, i.e. in the sample selection equation).

According to Cameron and Trivedi (2005), the two-stage Heckman model expands OLS regression by including the estimated omitted regressor $\lambda_i(z_i'\gamma)$. The subsequent estimation model is obtained by OLS, using positive values of W_2

$$W_{2i} = \boldsymbol{x}_{2i}'\boldsymbol{\beta}_2 + \boldsymbol{\lambda}(\boldsymbol{x}_{1i}'\widehat{\boldsymbol{\beta}}_1)\sigma_{12} + \varepsilon_i \quad (4)$$

in which ε corresponds to the error term, $\widehat{\boldsymbol{\beta}}_1$ represents the coefficient obtained from the first stage of the *Probit* of W_1 in x_1, since $\Pr[W_{1i}^* > 0] = \varphi(\boldsymbol{x}_1'\boldsymbol{\beta}_1)$, and $\boldsymbol{\lambda}(\boldsymbol{x}_1'\widehat{\boldsymbol{\beta}}_1) = \varphi(\boldsymbol{x}_1'\widehat{\boldsymbol{\beta}}_1)/\Phi(\boldsymbol{x}_1'\widehat{\boldsymbol{\beta}}_1)$ is the inverse of the estimated Mills ratio. It is noteworthy that this regression does not directly give an estimate of σ_2^2. Therefore, because it is a truncated variance it is estimated that $\widehat{\sigma}_2^2 = N^{-1}\sum_i[\widehat{\varepsilon}_i^2 + \widehat{\sigma}_{12}^2\widehat{\lambda}_i(\boldsymbol{x}_1'\widehat{\boldsymbol{\beta}}_1 + \widehat{\lambda}_i)]$, which $\widehat{\varepsilon}_i$ corresponds to the OLS estimation residue of the equation $W_{2i} = \boldsymbol{x}_{2i}'\boldsymbol{\beta}_2 + \boldsymbol{\lambda}(\boldsymbol{x}_{1i}'\widehat{\boldsymbol{\beta}}_1)\boldsymbol{\sigma}_{\mathbf{12}} + \varepsilon_i$ and $\widehat{\lambda}_i = \boldsymbol{\lambda}(\boldsymbol{x}_{1i}'\widehat{\boldsymbol{\beta}}_1)$. Noting that, through $\widehat{\rho} = \widehat{\sigma}_{12}/\widehat{\sigma}_2$, you can then estimate the correlation between the two errors.

To analyze the income differentials between smokers and nonsmokers in the Brazilian labor market along the quantiles of the distribution, we will use the method proposed by Firpo, Fortin, and Lemieux (2009) - FFL, which emerged as an extension of the Oaxaca (1973) and Blinder (1973) – OB decomposition [3]. This technique is a two-stage estimation method applicable to any statistical distribution of interest, providing us with increased flexibility in analyzing the labor income determination model (Wang, Cheng, & Smyth, 2013).

The FFL procedure makes it possible to decompose the wage income differentials between smokers and nonsmokers by analyzing them at sample income quantiles, as opposed to the OB decomposition, in which these differentials are evaluated only at the sample mean. It is also noteworthy that the FFL method corresponds to a simple regression in which the dependent variable is rearranged from a transformed version of its: the recentered influence function (RIF), which may be applied to any statistical distribution of interest, becomes possible to compute an influence function, whose denotation is Q_θ, of the marginal unconditional distribution f_y (Salardi, 2013).

In this way, the Recentered Influence Function (RIF) can be written as follows ((Firpo, Fortin, & Lemieux, 2009)):

$$\mathrm{RIF}(\mathrm{y}, \mathrm{Q}_\theta) = \mathrm{Q}_\theta + \frac{\theta - \mathrm{I}\{\mathrm{y} \leq \mathrm{Q}_\theta\}}{\mathrm{f_y}(\mathrm{Q}_\theta)} = \mathrm{c}_{1,\theta}\mathrm{I}\{\mathrm{y} > \mathrm{Q}_\theta\} + \mathrm{c}_{2,\theta} \tag{5}$$

in which $\mathrm{c}_{1,\theta} = 1/\mathrm{f_y}(\mathrm{Q}_\theta)$ and $\mathrm{c}_{2,\theta} = \mathrm{Q}_\theta - \mathrm{c}_{1,\theta}(1-\theta)$.

We use the individual characteristics and the condition of a smoker or nonsmoker [4] to evaluate whether this behavior harmful to health (smoking) contributes to the income differential of the work of individuals. To this end, we applied the regression of the RIF function to estimate the income differentials between smokers and nonsmokers (FFL decomposition), which can be described as follows:

$$\begin{aligned} \Delta_\theta lnW_{SNS} &= Q_\theta(lnW_S) - Q_\theta(lnW_{NS}) \\ &= [Q_\theta(lnW_S) - Q_\theta(lnW_{S-NS})] + [Q_\theta(lnW_{S-NS}) - Q_\theta(lnW_{NS})] \end{aligned} \tag{6}$$

in which $\Delta_\theta lnW_{SNS}$ represents the wage (or income) differential among smoking workers S and nonsmokers NS in θ-th quantile and $\mathrm{Q}_\theta(ln\, W_{S-NS})$ is a counterfactual distribution of wage/income, that is, corresponds to the distribution of conditional wage/income of workers who are smokers, if they have the same marginal return in the various skills as those who are nonsmokers. The first term on the right side is the component that has an explanation of the difference in wage/income attributable to differences in personal and productivity characteristics, and the second term is the component without an explanation of the wage/income differential attributable to differences in the returns of personal and productive characteristics. The unexplained component of the difference in wages/income is the income differential attributable to the penalty for being a smoker.

Aiming to obtain greater robustness for the estimates, the other method adopted in this study for the decomposition of labor income between smokers and nonsmokers is the nonparametric procedure proposed by Ñopo (2008). It corresponds to a pairing technique to identify the portion of the income differential observed specifically among individuals with common observable characteristics, because it allows a more precise decomposition of the income differential (isolates the effects inside and outside common support of observable characteristics), an advantage over the parametric decompositions (Oaxaca-Blinder and FFL, for example), which does not guarantee the equivalence of the individuals compared. In addition, the income difference is decomposed into four, instead of only two components as occurs in the Oaxaca-Blinder approach (Ñopo, 2008; Britto & Waltenberg, 2014; Vaz, 2018). Thus, a comparison is made between individuals who are within and outside the common support of characteristics.

Derived from the differential calculated using the expected value of gains conditional on observable characteristics (linked to these gains) and the cumulative distribution function of observable characteristics, this analysis considers whether the individual is a smoker or non-smoker [5] (Ñopo, 2008):

$$\Delta \equiv E[w|\ S] - E[w|\ NS] \tag{7}$$

It is worth noting that, in some cases, the support of the distribution of characteristics in the smoking group may differ from that of the non-smoking group. So, the best option is to subdivide each term (right side) of equation (7) into two distinct terms: one within the common support (intersection of supports) of characteristics and the other outside the common support (specific to the group under analysis). Therefore, it can decompose the term that corresponds to the common support of characteristics, in the same way that it is performed in the Oaxaca-Blinder method approach, to obtain two terms, which are only defined by the common support, although interpreted in the same traditional way. The income differential is decomposed, considering the differences in the expected wages of smokers and nonsmokers inside and outside the common support of characteristics (Ñopo, 2008):

$$\Delta = \Delta_S + \Delta_{NS} + \Delta_x + \Delta_0 \tag{8}$$

in which Δ_S represents the proportion of the differential attributed to distinctions between two groups of smokers, allowing for a comparison with the nonsmoker groups. One of these nonsmoker groups exhibits characteristics that can be matched with those of smokers, while the other does not allow for such pairing. Δ_{NS} is analogous to the former, however, for nonsmokers rather than smokers. The third term, Δ_x, captures the differential explained by characteristics between smokers and nonsmokers who are in common support, with interpretation equivalent to the term of Oaxaca-Blinder decomposition, whatever $(X_S - X_{NS})'\widehat{\beta_{NS}}$. Finally, the term Δ_0 represents the "unexplained" portion of the income differential, which is equivalent to the "unexplained" component, $\overline{X'_S}(\widehat{\beta_S} - \widehat{\beta_{NS}})$, of the Oaxaca-Blinder procedure. It is emphasized that the first two terms, Δ_S and Δ_{NS}, in a way, conduct a "clean-up" in the last two, i.e. Δ_x and Δ_0, which are the main components of estimation (Ñopo, 2008).

Furthermore, Ñopo (2008) points out that the first three terms of the decomposition, Δ_S, Δ_{NS} and Δ_x, refer to wage premiums based on observable characteristics, while the last term Δ_0 captures a combination of unobservable differences awarded by the labor market, as well as discrimination (in the context of racial or gender differentials) or choice (in the context of differentials by professional categories), for example. We note that using equation (9) can be an even clearer comparison with the decomposition of the Oaxaca-Blinder method:

$$\Delta = (\Delta_S + \Delta_{NS} + \Delta_x) + \Delta_0 \tag{9}$$

Then, the newly formed sample consists of four types of individuals: paired smokers, paired nonsmokers, unpaired smokers, and unpaired nonsmokers, thus fulfilling the objective of quadripartite differential decomposition. It is important to highlight that, in the Ñopo (2008) technique, it is not recommended to use continuous variables, since this would greatly increase the probability of non-matching. So continuous variables must be transformed into categorical variables or, if possible, into binary variables (dummies). Thus, besides the inclusion of new variables, which constitutes another difference between the estimated wage equations for use with, the Oaxaca-Blinder methodology and the characteristics used for pairing with the Ñopo technique is precisely the transformation of continuous variables into categorical or binary (Ñopo, 2008).

This study utilizes data from the 2019 National Health Survey (PNS), a household-based survey conducted by the Brazilian Institute of Geography and Statistics (IBGE). According to IBGE (2021), the primary aim of the survey is to generate national data on the health status and lifestyles of the Brazilian population. Additionally, it aims to analyze the utilization and access to healthcare services, along with preventive actions in this context, to provide insights for the development of public policies in this domain.

Table 1 presents descriptive statistics (means and standard deviations) of the variables used in econometric models and their definitions. According to Pereira and Oliveira (2016, 2017), one way to obtain the variable wage (income) hour is from the division of income obtained by the individual in the main work by the hours worked per week, multiplied by 4.2 (based on a month of 30 days divided by the seven days of the week). According to IBGE (2020), a daily smoker is a person who makes daily use of at least one of the tobacco products that emit smoke, regardless of how long he smokes. In this work, a smoker is anyone who smokes a tobacco product daily or less than daily.

Table 1.
Descriptive statistics of variables used in the estimates of econometric models and their definitions

Variables	Definition	Full sample	Smokers	Nonsmokers
Wage	Hourly wage (income)	13.9637 (33.6151)	11.5117 (16.7561)	14.1130 (34.3702)
Schooling level	The highest level of education achieved	8.7660 (2.5565)	7.9602 (2.5184)	8.8129 (2.5508)
Age	Age in years	35.2677 (21.6760)	47.1490 (15.7661)	34.7629 (21.7485)
Sex	1 = Man; 0 = Woman	0.4812 (0.4996)	0.6041 (0.4891)	0.4760 (0.4994)
Race	1 = White; 0 = Other	0.3605 (0.4801)	0.3353 (0.4721)	0.3616 (0.4805)
Urban zone	Residence zone: 1 = Urban; 0 = Country	0.7632 (0.4251)	0.7269 (0.4455)	0.7646 (0.4242)
Statutory status	Municipal, state, or federal public servant	0.0383 (0.1919)	0.0336 (0.1803)	0.0385 (0.1923)
Employment with a formal contract	Employment contract governed by the CLT[a]	0.1373 (0.3442)	0.1574 (0.3642)	0.1365 (0.3433)
Exclusion variables (First stage of Heckman model)				
Be working	1 = yes; 0 = no	0.4115 (0.4921)	0.6093 (0.4879)	0.4035 (0.4906)
Receives retirement	1 = yes; 0 = no	0.1500 (0.3571)	0.2295 (0.4205)	0.1466 (0.3537)
Receives pension	1 = yes; 0 = no	0.0305 (0.1720)	0.0321 (0.1762)	0.0305 (0.1719)
Receives alimony, donation, or allowance	1 = yes; 0 = no	0.0290 (0.1679)	0.0321 (0.1762)	0.0289 (0.1676)
Other sources of income	1 = yes; 0 = no	0.0930 (0.2904)	0.1491 (0.3562)	0.0907 (0.2872)
Responsible for the household	1 = yes; 0 = no	0.3204 (0.4666)	0.7353 (0.4412)	0.3037 (0.4598)
Health-related variables				
Smoker	1 = yes; 0 = no	0.0388 (0.1930)	–	–
Ex-smoker	Used to be smoker: 1 = yes; 0 = no	0.0825 (0.2751)	–	0.0858 (0.2801)
Use/used electronic cigarettes	1 = yes; 0 = no	0.0012 (0.0340)	0.0090 (0.0942)	0.0008 (0.0290)
There is any smoker in the family	1 = yes; 0 = no	0.0479 (0.2135)	0.5938 (0.4911)	0.0258 (0.1587)
Self-perception of your health	1 = good; 0 = not good	0.6764 (0.4678)	0.5925 (0.4914)	0.6798 (0.4666)
Consumes alcoholic beverages	1 = yes; 0 = no	0.0828 (0.2756)	0.4578 (0.4982)	0.0677 (0.2512)
Sedentary[b]	1 = yes; 0 = no	0.8053 (0.3959)	0.3990 (0.4897)	0.8217 (0.3828)
Low fruit consumption	1 = yes; 0 = no	0.1717 (0.3771)	0.6861 (0.4641)	0.1510 (0.3580)
Low vegetable consumption	1 = yes; 0 = no	0.1501 (0.3572)	0.5454 (0.4980)	0.1342 (0.3408)
BMI (Body Mass Index)	Weight in kilograms divided by the square of height in meters	26.3782 (4.8199)	25.5183 (4.7467)	26.5020 (4.8177)
Overweight	BMI >25: 1 = yes; 0 = no	0.8685 (0.3379)	0.4895 (0.4999)	0.8838 (0.3204)

(*continued*)

Variables	Definition	Full sample	Smokers	Nonsmokers
Had/has depression	1 = yes; 0 = no	0.0284 (0.1660)	0.1089 (0.3115)	0.0251 (0.1565)
He had/has a mental illness	1 = yes; 0 = no	0.0167 (0.1281)	0.0620 (0.2412)	0.0149 (0.1210)
He had/has schizophrenia	1 = yes; 0 = no	0.0010 (0.0320)	0.0053 (0.0724)	0.0009 (0.0292)

Note(s): Standard deviations are in parentheses
[a]Consolidation of labor laws. According to Brazil (2023), this Consolidation establishes the rules that regulate individual and collective labor relations, provided for therein
[b]Dummy variable that takes the value 1 if the individual did not practice any type of physical exercise or sport, if their work, did not do heavy cleaning, did not carry weight or did another heavy activity that requires intense physical effort, if to go to or from work, did not make any journey on foot or by bicycle, and if in their domestic activities, they did not do heavy cleaning, did not carry heavy weight or did not do any other heavy activity that requires intense physical effort
Source(s): Own elaboration from 2019 PNS data (IBGE, 2021)

Table 1.

In the context of the economic literature, Van Ours (2004) sought to address and mitigate potential endogeneity issues related to individuals' choice of smoking status. Traditionally, various factors such as health conditions, religious preferences, self-perception of health, family behavior, and cigarette prices have been used as instrumental variables. Anger and Kvasnicka (2010) emphasized the significance of family-specific characteristics as instruments to alleviate endogeneity in both current and past smoking status. For this purpose, they employed dummy variables indicating co-residence with at least one smoking family member or a nonsmoker.

From the perspective of the medical literature, studies on factors associated with smoking (lifestyle, health and/or psychological characteristics, psychiatric disorders, and eating habits) have demonstrated the existence of an association between smoking and some habits resulting from the lifestyle adopted by individuals, such as alcohol consumption, the use of electronic cigarettes and sedentary lifestyle. They also revealed an association between smoking and health and psychological characteristics, namely: being overweight and self-perception of their health status. Also, between smoking and psychiatric disorders, such as schizophrenia and depression, among others. Finally, the relationship between smoking and inadequate eating habits, such as low consumption of fruits and vegetables (Rondina, Gorayeb, & Botelho, 2007; Berto, Carvalhaes, & Moura, 2010; Bonnechère *et al.*, 2019; Lee & Lee, 2019).

In this regard, the *Ex-smoker* variable is included as a control, aiming to mitigate possible contamination of the nonsmoking group by individuals who have been smokers at some point in their lives. The variables *There is any smoker in the family, Self-perception of health, Alcoholic, e-cigarettes, Sedentary, Low fruit consumption, Low vegetable consumption, BMI, Overweight, Depression, Schizophrenia* and *Mental illness* were also included as controls. This addition aims to address a potential endogeneity issue related to the smoking variable, which may arise from unobserved factors such as lifestyle, health, psychological characteristics, psychiatric disorders, and eating habits, and which could be associated with the decision to smoke.

To control the effects related to jobs with different employment contract regimes, the variables *Statutory status*, referring to city, state, or federal public servants, and *Formal employment*, which identifies workers who have an employment contract governed by the CLT, are included in the estimations. This is significant because, as noted by Almeida and Araújo Júnior (2017), in the context of public servants, labor productivity may not necessarily impact wages. This is due to the additional factor of job stability, making it challenging to draw direct comparisons with employees in the private sector.

In general, descriptive statistics in Table 1 show that nonsmokers receive, on average, 22.6% more than smokers in the Brazilian labor market. Furthermore, nonsmokers achieve a higher level of education than smokers on average. It is noteworthy that smokers, on average, are 12 years older than nonsmokers and are more prevalent among men.

When it comes to health-related variables, it is noteworthy that most smokers have smokers in the family, an aspect opposite to that observed in nonsmokers. With respect to healthy habits, statistics indicate that smokers have a significantly higher rate of alcohol consumption compared to nonsmokers. Additionally, a majority of smokers tend to consume lower quantities of fruits and vegetables.

On the other hand, smokers exhibit a lower degree of sedentary behavior compared to nonsmokers and are also less frequently found in the BMI range indicating overweight. Notably, among nonsmokers, a significant portion falls into the category of overweight individuals.

Finally, there are more smokers than nonsmokers in groups of people who have/had depression, mental illnesses, and schizophrenia. It indicates a constructive interaction of these behaviors with what has been observed in the literature. Therefore, for a more accurate analysis of these relationships, the next section presents the results of the estimations of the econometric models used to answer the questions proposed in this research.

4. Results

Table 2 shows the results of the Heckman model estimation. In the first stage of this model, the estimated coefficients of the exclusion variables [6] show that all of them related to receiving retirement, pension, and other sources of income showed a negative sign, evidencing that they negatively affect the decision of individuals to be inserted in the labor market. According to Pereira and Oliveira (2016), receiving such income tends to decrease the probability of offering work to those who receive them. Because they have high reserve wages, they would require higher wages than these to exchange their leisure for work. On the other hand, the coefficient of the exclusion variable *Responsible for the household* was positive, suggesting that it tends to represent a positive factor in influencing the option of working.

In the second stage of the Heckman model (third column of Table 2), the coefficient of the variable λ (lambda) is statistically significant, indicating that the utilization of this model proved indispensable for addressing potential issues of sample selectivity and bias in the estimated coefficients. It is crucial to emphasize that the variable λ (lambda) does not have a correct sign, only its level of significance is important. Thus, the coefficient sign indicates only the direction of the relationship between that variable and income: if positive the observed factors in the first stage, which induce the individual to work, are also directly related to their income; in negative case, these factors are inversely related to their income (Psacharopoulos & Tzannatos, 1992; Pereira & Oliveira, 2017). As the estimated coefficient is negative, the factors influencing the probability of individuals providing labor are inversely correlated with income.

The term related to positive schooling levels indicates that each additional level of study increases income at mean by 21.4%, confirming the relevance of education in determining incomes (Pereira & Oliveira, 2017). The age, used as a proxy variable to experience at work, has a positive coefficient. It shows that experience contributes positively to the income of individuals in the country. And, the variable age squared, with a negative sign, indicates decreasing rates of return to worker productivity as their age increases, corroborating the literature (Pereira & Oliveira, 2017).

Variables	1st stage Dependent variable The decision to offer labor [1 = job offer; 0 = no offer]	2nd stage Hourly income logarithm
Schooling level	0.167*** (0.0415)	0.214*** (0.0367)
Age	0.0956*** (0.00304)	0.0200*** (0.00310)
Age squared	−0.00114*** (3.14e−05)	−9.60e−05** (4.07e−05)
Gender	0.686*** (0.0220)	0.187*** (0.0163)
Race	0.0700*** (0.0214)	0.177*** (0.0180)
Urban zone	0.105*** (0.0370)	0.197*** (0.0176)
Statutory		0.126*** (0.0255)
Employment with a formal contract		0.0687** (0.0320)
Receive retirement	−1.053*** (0.0320)	
Receive pension	−0.282*** (0.0314)	
Receives other sources of income	−0.598*** (0.0222)	
Responsible for the household	0.280*** (0.0349)	
Smoker		−0.0496** (0.0183)
Ex-smoker		−0.0463*** (0.00869)
Use/used electronic cigarettes		0.129** (0.0482)
There is a smoker in the family		−0.0447*** (0.0158)
Self-perception/self-assessment of your health		0.164*** (0.0101)
Consumes alcoholic beverages		0.128*** (0.0208)
Sedentary		0.0410** (0.0153)
Low fruit consumption		−0.0635*** (0.00839)
Low vegetable consumption		−0.101*** (0.0134)
Overweight		0.0422*** (0.00873)
Had/has depression		0.0281 (0.0201)
Had/has mental illness		0.0575 (0.0419)

(*continued*)

Table 2. Heckman model: estimation of the sample selection equation (*Probit model* – 1st stage) and estimation of the income's equation (2nd stage) for Brazil in 2019

Table 2.

Variables	1st stage	2nd stage
	Dependent variable	
	The decision to offer labor [1 = job offer; 0 = no offer]	Hourly income logarithm
Had/has schizophrenia		−0.416***
		(0.141)
Lambda (λ)		−0.105**
		(0.0394)
Constant	−3.434***	−0.867
	(0.640)	(0.581)
Dummies for CNAE[a] sectors	No	Yes
Observations	76.881	46.051
R^2	–	0.392
Pseudo-likelihood log (coefficient)	−68877799	–
Pseudo R^2	0.2625	–

Note(s): Statistical significance of the estimates defined by: * significant at 10% ($p < 0.1$); ** significant at 5% ($p < 0.05$); *** significant at 1% ($p < 0.01$). Standard errors are robust to heteroscedasticity clustered by federative units in parentheses. Because the PNS is a complex sample, both first and second-stage estimates were weighted by the sampling weight of everyone
[a]National Classification of Economic Activities developed by IBGE
Source(s): Own elaboration from 2019 PNS data (IBGE, 2021)

Regarding the gender variable, its positive coefficient shows that being male increases incomes by 18.7%, possibly signaling the existence of discrimination in favor of men in the Brazilian labor market, a result like that found by Silveira and Siqueira (2021). The positive coefficient of the race variable shows that being white increases yields by 17.7%, also signaling a possible existence of discrimination in favor of white people in the Brazilian labor market, a result also aligned with that of the study by Silveira and Siqueira (2021). The variable area of residence (urban zone) presents a positive coefficient, showing that living in an urban area positively influences the income of individuals by 19.7%.

The smoking variable shows that being a smoker has a negative effect of 4.96% on the income on average. This result indicates that smokers are penalized in their work income for their habit in the Brazilian labor market. Regarding aspects related to the habit of smoking, it is observed that being an ex-smoker and having smokers in the family also negatively affects income, however, the habit of using electronic cigarettes has a positive relationship with this income.

The results show that some health-related characteristics in which there is a prevalence of smokers, such as drinking alcoholic beverages, sedentary lifestyle, and being overweight, have positive effects on income, while other characteristics, such as low consumption of fruits and vegetables and having schizophrenia affect negatively on these earnings. It shows the importance of considering these habits and characteristics to isolate the effect of smoking on the estimation of the labor income equation.

To analyze the income differentials between nonsmokers and smokers in Brazil, Table 3 shows the results of FFL decomposition in the mean and throughout income distribution. The results indicate that nonsmokers receive more than smokers on average and in all quantiles analyzed. The lowest differential to nonsmokers (10.6%) was observed in the median (Q50) and the highest (25.1%) in the 90th percentile.

It is noted that, on average and across all analyzed quantiles, the explained component of the decomposition is both positive and statistically significant. This indicates that the personal, productivity, employment, and locational characteristics controlled in the estimations contribute to explaining the income differential in favor of nonsmokers in the

Table 3. Estimation of FFL decomposition in the mean and throughout the distribution of income (nonsmokers and smokers) to Brazil in 2019

Groups	Mean	Q10	Q25	Q50	Q75	Q90	Q99
				Quantiles			
		Dependent variable: income/hour logarithm					
Predicted values of income/hour logarithm							
Nonsmokers	2.26172	1.17431	1.72183	2.07928	2.6527	3.3602	4.68090
Smokers	2.07646	0.94445	1.59193	1.97359	2.5017	3.1089	4.53932
Difference	0.18527***	0.22986***	0.12989***	0.10570***	0.15097***	0.25129***	0.14157**
	(0.01847)	(0.02720)	(0.012273)	(0.011335)	(0.01563)	(0.02924)	(0.06023)
Decomposition components							
Explained	0.11170***	0.17092***	0.06515***	0.06165***	0.14454***	0.16133***	0.13212***
	(0.01167)	(0.01340)	(0.00435)	(0.00601)	(0.01026)	(0.011596)	(0.016993)
Unexplained	0.07356***	0.05894**	0.06474***	0.04405***	0.00643	0.08996***	0.00945
	(0.01546)	(0.02594)	(0.01095)	(0.01011)	(0.0138)	(0.02542)	(0.05779)
Observations	99.081	99.081	99.081	99.081	99.081	99.081	99.081

Note(s): Statistical significance of the estimates defined by: * significant at 10% ($p < 0.1$); ** significant at 5% ($p < 0.05$); *** significant at 1% ($p < 0.01$). Standard errors in parentheses. Q10 (income = R$ 350), Q25 (income = R$ 900), Q50 (income = R$ 1,200), Q75 (income = R$ 2,000), Q90 (income = R$ 4,000) and Q99 (income = R$ 15,000). All estimates for the mean and quantiles were controlled by school level, age, age squared, gender, race, urban area, statutory status, employment with a formal contract, CNAE sectors, federative units, and lambda

Source(s): Own elaboration from 2019 PNS data (IBGE, 2021)

Brazilian labor market. Specifically, 60.3% of the differential in favor of nonsmokers is accounted for by the controlled characteristics in the model, as observed in the average scenario. It is also noteworthy that in the highest quantiles (Q75 and Q99 specifically) these observed characteristics explain more than 90% of the income gap against smokers.

Conversely, a portion of this income differential remains unexplained by the observed characteristics, potentially linked to the earnings penalty faced by smokers due to this habit. For instance, 41.7% of the income differential favoring nonsmokers at the median (Q50) is not accounted for by the controls employed in the decomposition estimation. This unexplained portion could be attributed to the income penalty imposed on workers who smoke. Nevertheless, it is noteworthy that the unexplained component is not statistically significant in the Q75 and Q99 quantiles. This suggests that within these income ranges, the income differential favoring nonsmokers is solely accounted for by the observed characteristics. In other words, there is no evidence of a subjective penalty affecting workers' income within these distribution strata.

Table 4 shows the results of the Ñopo decomposition for the income gap between smokers and nonsmokers at the mean and income quantiles in Brazil. As highlighted by Britto and Waltenberg (2014) and Vaz (2018), the non-parametric technique proposed by Ñopo (2008) allows a more accurate decomposition of the wage differential by isolating the effects inside and outside the common support of characteristics. It is an advantage concerning parametric decompositions, such as the FFL decomposition, which does not guarantee the equivalence of the individuals compared.

The results show that, for the general decomposition (on average), nonsmokers earn more than smokers in the Brazilian labor market, corroborating the result observed in the FFL decomposition. However, this differential is only 8.39% in favor of nonsmokers, a value 120.8% smaller than that observed in the FFL decomposition. Britto and Waltenberg (2014) emphasize that parametric decompositions (Oaxaca-Blinder and FFL decompositions, for example) tend to overestimate the differential attributed to unobservable differences. This implies that the inability to ensure complete equivalence between the individuals compared within each group can lead to combinations of individual characteristics in one group that are not present in another.

Components of decomposition	General (average)	Quantiles					
		Q10	Q25	Q50	Q75	Q90	Q99
Total difference (D)	−0.0839	−0.1138	−0.0456	−0.0085	−0.0131	−0.0086	−0.0041
DX	−0.0097	0.0006	−0.0023	−0.0036	−0.0018	−0.0041	0.0023
DS	0.0029	0.0034	0.0013	−0.0001	0.0012	0.0009	0.0024
DNS	−0.0402	−0.0436	−0.0068	0.0008	−0.0043	−0.0105	−0.0082
D0	−0.0369	−0.0742	−0.0378	−0.0055	−0.0082	0.0053	−0.0006
PercS	0.9297	0.9369	0.9322	0.9193	0.9423	0.9299	0.9002
PercNS	0.5681	0.6559	0.6661	0.5429	0.5659	0.5339	0.4501

Note(s): Q10 (income = R$ 300), Q25 (income = R$ 900), Q50 (income = R$ 1,200), Q75 (income = R$ 2,000), Q90 (income = R$ 4,000) and Q99 (income = R$ 15,000). D = DX + DM + DF + D0. DX = difference of observable characteristics within the common support. DS = part of the difference that can be explained by differences in the characteristics of smokers who are inside and outside the common support. DNS = part of the difference that can be explained by differences in the characteristics of nonsmokers who are on top and outside the common support. D0 = part of the unexplained differential. PercS = percentage of smokers who are within the common support. PercNS = percentage of nonsmokers who are within the common support. The pairing of the decomposition was performed to go from the controls related to characteristics of residence, personal, productive, employment, and locational of individuals (schooling level, gender, race, zone of residence, statutory status, employment with a formal contract, CNAE sectors, federative units)
Source(s): Own elaboration from 2019 PNS data (IBGE, 2021)

Table 4.
Decomposition of Ñopo for the income differential (smokers and nonsmokers) in the mean and by income quantiles for Brazil in 2019

The lowest differential to nonsmokers (0.41%) was observed in the 99th percentile and the highest (11.38%) in the 10th percentile. Furthermore, it is important to highlight the decrease in the income differential in favor of nonsmokers as the highest quantiles of the distribution are analyzed. It can be said that in the median (Q50) and the Q90 and Q99 quantiles, there are practically no average differences in income between smokers and nonsmokers.

Considering the part of this income gap related to observable characteristics (DX), i.e. residence, personal, productive, employment, and locational of individuals, we observed that the negative sign of these characteristics explains the difference in favor of nonsmokers. Only in the 10th and 90th quantile is it possible to observe that the characteristics used in the matching do not explain the difference in favor of nonsmokers.

About the residual component (or unexplained part) of the decomposition (D0), we note that 3.7% of this income differential is not explained by the controls adopted within the common support, i.e. signaling again the existence of a possible penalty on the income of smokers compared to nonsmokers in the Brazilian labor market, as observed in the FFL decomposition for the mean. However, it is important to highlight that this subjective penalty in earnings attributed to the smoking habit is significantly smaller than that observed in the FFL decomposition. Therefore, this result reiterates the importance of distinguishing the analysis between individuals that are inside and outside the common support of characteristics, as, possibly, individuals that are not equivalently comparable (outside the common support of characteristics) are overestimating the differences observed in the decomposition FFL (Dobner, Gonçalves, & Pereira, 2022).

Nevertheless, it is crucial to emphasize that in the 10th quantile, 7.42% of this gap in favor of nonsmokers is not accounted for by the controlled characteristics in the model. This unexplained portion represents the income penalty on workers due to the smoking habit within this income range. Notably, this is the most significant magnitude of penalty observed in the Ñopo decomposition, surpassing even that observed in the FFL decomposition for the same quantile by 25.8%.

This robustness of the analysis provided by the Ñopo decomposition remains translated into the percentages of smokers (PercS) and nonsmokers (PercNS) who are within the common support. For example, when estimating the sample mean (general sample) and in all quantiles analyzed, more than 90% of smokers in the sample were observed to be within the common range of characteristics. The same pattern is observed for nonsmokers, as in most quantiles it is possible to verify that more than 50% of nonsmokers are within the common support.

Comparing the results obtained in the FFL decomposition, it becomes evident that, on average and across all quantiles analyzed, the estimates from both decompositions are synergistic, that is, they point in the same direction, indicating the presence of an income differential unfavorable to smokers. However, a crucial aspect to consider is the attenuation in the magnitude of the coefficients in the Ñopo decomposition compared to the FFL decomposition. The differential against smokers is notably lower in the Ñopo decomposition, both in the mean and across quantiles: only 8.39% on average, 0.85% at the 50th/median, 0.86% at the 90th, and 0.41% at the 99th quantile. In contrast, the FFL decomposition yielded higher differentials: 18.53% on average, 10.57% at the 50th quantile/median, 25.13% at the 90th, and 14.16% at the 99th quantile.

When compared to other studies conducted at both international and Brazilian levels, the results obtained in this study for the average, in both FFL and Ñopo decompositions, align with the findings of the majority of empirical studies in the field. They reveal a negative association between smoking and labor-related income, along with income differentials against smokers when contrasted with non-smokers. This supports the hypothesis of a potential income penalty for smokers in the labor market. In the case of FFL decomposition (Table 3), the value of 18.53% found for the differential in the mean, favorable to nonsmokers, is synergistic and even with a magnitude similar to those found by Van Ours (2004), for workers from the Netherlands, using data from 2001, by Lokshin (2006) for Albania with data from 2005, and by Hotchkiss and Pitts (2013) for USA workers, using data from 1992–2011. When it comes to the Brazilian labor market, this result is close to that observed by Almeida and Araújo Júnior (2017) and Justus *et al.* (2019), both using data from 2008.

More recently, the study by Hotchkiss and Pitts (2013), in which the analysis included a decomposition of the wage gap between smokers and nonsmokers in the United States with data from 1992 to 2011. The findings revealed that, on average, smokers receive lower wages than nonsmokers by approximately 24%. Notably, two-thirds of this wage differential were explained by differences in observable characteristics between the two groups. This result is in line with what was observed in this study in the FFL decomposition for the mean, where it was found that 60.3% of the differential in favor of nonsmokers is explained by the controlled characteristics in the model for the average.

In the case of Ñopo decomposition (Table 4), the value found for the general decomposition (on average) of the income differential unfavorable to smokers of 8.39% is in line with and similar to that found by Levine *et al.* (1997) for the USA workers, using data from 1984 and 1991, by Auld (2005) for Canada with data from 1995–1991, by Grafova and Stafford (2009) for USA with data from 1986–2001, and by Anger and Kvasnicka (2010) for Germany with data from 2002. In the Brazilian context, this result is close to that observed by Uhr *et al.* (2021) with data from the 2013 PNS which also uses a matching method to assess the income gap between smokers and nonsmokers. This suggests a possible maintenance of the income differential between smokers and nonsmokers in the Brazilian labor market between 2013 and 2019.

Nevertheless, the current study contributes to the literature by examining the heterogeneity of differentials between smokers and nonsmokers across the entire labor income distribution. Notably, the most significant magnitudes of this gap in favor of nonsmokers are identified in the 10th quantile in both decompositions. Within this income stratum, the differential in favor of nonsmokers is more pronounced, with a substantial portion of this gap explained by the observed characteristics of workers, particularly in the case of FFL decomposition. Additionally, in both decompositions, it is evident that in this quantile (Q10), workers face income penalties due to their smoking habit. Notably, in the Ñopo decomposition, this component is larger than that observed in the FFL decomposition.

In relation to the Q75 and Q99 quantiles, a smaller income difference is observed compared to the average in both decompositions. However, in these quantiles, this difference is almost entirely explained by the controlled characteristics. Therefore, it is not possible to infer that there is an income penalty solely due to the habit of smoking, as is more evident in the lower quantiles. This pattern is consistently observed in both the FFL and Ñopo decompositions, leading to the conclusion that in these higher income ranges of the distribution, smokers do not experience workplace prejudice specifically due to their smoking habit in Brazil.

5. Final considerations

The study aimed to investigate the impact of smoking on the income of workers in the Brazilian labor market. To achieve this goal, we utilized data from the 2019 National Health Survey (PNS), a household-based survey conducted by the Brazilian Institute of Geography and Statistics (IBGE, 2021). Our empirical approach involved employing the Heckman (1979) method initially to address potential sample selection bias related to employment status. Subsequently, we conducted the income gap decomposition between smoking and nonsmoking workers, both on average and across the income distribution, using the procedure outlined by Firpo, Fortin, and Lemieux (2009). To enhance the robustness of our estimates, we also employed the technique proposed by Ñopo (2008). Overall, the results demonstrated robustness, as the econometric strategies employed effectively mitigated significant issues in the estimations, both at the mean and across the income distribution.

Regarding the results, firstly, the estimation of the income equation via the Heckman model shows that even correcting the sample selection problem and controlling factors related to personal, productive, health characteristics and factors that may influence the adoption of smoking habits, it has been observed that smokers face income penalties in the Brazilian job market. At this point, it is worth highlighting the statistical significance of a large part of the controls related to health and behavioral characteristics, confirming the need to use this information to "clean up" as much as possible the effect of being a smoker on labor income.

The results from the counterfactual estimations using the methods of Firpo, Fortin, and Lemieux (2009) and Ñopo (2008) for the income gap between smoking and nonsmoking workers, both at the mean and across almost all quantiles (strata) of the income distribution, align with the existing literature in the field. These findings provide evidence of an income penalty for smokers in the Brazilian labor market, particularly pronounced in the lower quantiles (Q10, Q25, and the median – Q50). In the highest quantiles, mainly Q75 and Q99, it was observed that the income differential against smokers is explained exclusively by the controlled characteristics of the workers, that is, at these income levels there were no subjective penalties on smokers' income due to their habit.

It is also important to highlight the gain in robustness observed with the use of the observed characteristics matching technique proposed by Ñopo (2008), as by guaranteeing the analysis of the income differential between smokers and nonsmokers within the common support of characteristics with a high in the number of matched workers, it was possible to confirm the overestimation of the earnings gap against smokers observed from the estimation of the parametric model (FFL decomposition) that does not have this validation. Furthermore, it is worth highlighting the reduction in this gap in income from the lowest to the highest quantile observed in the decomposition of Ñopo (2008), showing that most of the difference in favor of nonsmokers, as well as the penalty in income due to the habit of smoking, is characteristic of jobs that pay in the lowest quantiles of the distribution.

The significance of the current study becomes evident in its ability to discern the heterogeneity of the income gap favoring nonsmokers. In the lower quantiles, a more pronounced magnitude of differentials against smokers and a higher incidence of unexplained penalties in the income of these workers were observed. Conversely, in the higher quantiles, there was a lower magnitude of differentials, with little evidence indicating a penalty in earnings for workers who smoke.

Lastly, it is important to highlight a significant limitation in this study, namely the constraint imposed by the 2019 PNS dataset, which lacks information on certain subjective factors that can also impact levels of labor income. These factors include the individual's level of effort and specific abilities, applicable to both smokers and nonsmokers. Such factors could be linked to latent individual predispositions influencing the decision to smoke. However, it is anticipated that this research will inspire further investigations in future studies, addressing these limitations by leveraging alternative databases that address the aforementioned issues.

Notes

1. A summary of international and national work on this topic can be seen in Table A1 in the appendix.
2. Ordinary Least Squares.
3. The Oaxaca-Blinder procedure decomposes the income differential between observable and unobservable characteristics, and observable characteristics are the productive and personal characteristics of individuals and non-observable characteristics, the component that demonstrates the unexplained part. More information in Oaxaca (1973) and Blinder (1973).

4. Adapted for smokers and nonsmokers from Wang *et al.* (2013) who examined wage compensation and income differentials (risk premiums) for migrant workers taking risky and safe jobs (at risk and without risk).
5. Adapted for smokers and nonsmokers from Britto and Waltenberg (2014), which evaluated the attractiveness of high school teacher occupation, as expressed by salary differentials between this category of teachers and three comparison groups.
6. Variables that affect the probability of individuals offering or not to labor, but that do not influence (directly) their labor income, as can be seen in greater detail in section 3.

References

Almeida, A. T. C. de, & Araújo Júnior, I. T. de (2017). Tabagismo e penalização salarial no mercado de trabalho brasileiro. *Economia Aplicada*, *21*(2), 249–76. doi: 10.11606/1413-8050/ea146024.

Anger, S., & Kvasnicka, M. (2010). Does smoking really harm your earnings so much? Biases in current estimates of the smoking wage penalty. *Applied Economics Letters*, *17*(6), 561–564. doi: 10.1080/13504850802260846.

Auld, M. C. (2005). Smoking, drinking, and income. *Journal of Human Resources*, *40*(2). Available from: https://econpapers.repec.org/article/uwpjhriss/v_3a40_3ay_3a2005_3ai_3a2_3ap505-518.htm

Berto, S. J. P., Carvalhaes, M. A. B. L., & Moura, E. C. de (2010). Tabagismo associado a outros fatores comportamentais de risco de doenças e agravos crônicos não transmissíveis. *Cadernos de Saúde Pública*, *26*(agosto), 1573–1582. doi: 10.1590/S0102-311X2010000800011.

Bhai, M. (2020). The earning losses of smokers. *IZA Journal of Labor Economics*, *9*(1). doi: 10.2478/izajole-2020-0001.

Blinder, A. S. (1973). Wage discrimination: Reduced form and structural estimates. *The Journal of Human Resources*, *8*(4), 436–455. doi: 10.2307/144855.

Böckerman, P., Hyytinen, A., & Kaprio, J. (2015). Smoking and long-term labour market outcomes. *Tobacco Control*, *24*(4), 348–353. doi: 10.1136/tobaccocontrol-2013-051303.

Bonnechère, B., Cissé, K., Millogo, T., Ouédraogo, G. H., Garanet, F., Ouedraogo, M. A., . . . & Kirakoya-Samadoulougou, F. (2019). Tobacco use and associated risk factors in burkina faso: Results from a population-based cross-sectional survey. *BMC Public Health*, *19*(1), 1466. doi: 10.1186/s12889-019-7826-6.

Braakmann, N. (2008). The smoking wage penality in the United Kingdom: Regression and matching evidence from the British household survey Panel, Working Paper Series in Economics, Working Paper Series in Economics, Agosto. Available from: https://ideas.repec.org/p/lue/wpaper/96.html

Brazil (2023). Consolidation of labor laws Decree-law No. 5,452, of May 1, 1943. Available from: https://www.planalto.gov.br/ccivil_03/decreto-lei/del5452.htm (accessed 21 October 2023).

Brazilian Institute of Geography and Statistics (IBGE) (2020). Research board. Work and income coordination. National Health Survey 2019 - PNS. 'Percepção do estado de saúde, estilos de vida, doenças crônicas e saúde bucal'. Brasil e grandes regiões/IBGE, IBGE, Rio de Janeiro. 113. Available from: https://biblioteca.ibge.gov.br/visualizacao/livros/liv101764.pdf (accessed 30 Mai 2021).

Brazilian Institute of Geography and Statistics (IBGE) (2021). National health survey 2019 – PNS. Available from: https://www.ibge.gov.br/estatisticas/sociais/saude/9160-pesquisa-nacional-de-saude.html (accessed 30 Mai 2021).

Britto, A., & Waltenberg, F. (2014). É atrativo tornar-se professor do ensino médio no Brasil? Evidências com base em decomposições paramétricas e não paramétricas. *Estudos Econômicos (São Paulo)*, *44*(1), 5–44. doi: 10.1590/s0101-41612014000100001.

Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: Methods and applications*. New York: Cambridge University Press.

Cirino, J. F., & Lima, J. E. de (2012). Diferenças de rendimento entre as regiões metropolitanas de Belo Horizonte e Salvador: Uma discussão a partir da decomposição de Oaxaca-Blinder. *Revista Econômica Do Nordeste*, *43*(2), 371–390. doi: 10.61673/ren.2012.216.

Darden, M. E., Hotchkiss, J. L., & Pitts, M. M. (2021). The dynamics of the smoking wage penalty. *Journal of Health Economics*, *79*, 102485. doi: 10.1016/j.jhealeco.2021.102485.

Dobner, L. de O., da Rocha Gonçalves, R., & Pereira, R. M. (2022). Análise setorial dos diferenciais de rendimentos entre os gêneros no Rio Grande do Sul: Uma abordagem paramétrica e não paramétrica. *Revista Brasileira de Estudos Regionais e Urbanos*, *16*(2), 214–238. doi: 10.54766/rberu.v16i2.889.

Firpo, S., Fortin, N. M., & Lemieux, T. (2009). Unconditional quantile regressions. *Econometrica*, *77*(3), 953–973. doi: 10.3982/ECTA6822.

Grafova, I. B., & Stafford, F. P. (2009). The wage effects of personal smoking history. *ILR Review*, *62*(3), 381–393. doi: 10.1177/001979390906200307.

Grossman, M. (1972). On the concept of health capital and the demand for health. *Journal of Political Economy*, *80*(2), 223–255. doi: 10.1086/259880.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, *47*(1), 153–161. doi: 10.2307/1912352.

Heineck, G., & Schwarze, J. (2003). Substance use and earnings: The case of smokers in Germany, IZA Discussion Paper 743, Institute of Labor Economics (IZA). Available from: https://econpapers.repec.org/paper/izaizadps/dp743.htm

Hotchkiss, J. L., & Pitts, M. M. (2013). Even one is too much: The economic consequences of being a smoker. *SSRN Electronic Journal*. doi:10.2139/ssrn.2359224.

Justus, M., Sant'Anna, E., Davanzo, E. S., & Moreira, G. C. (2019). The wage penalty of smoking in Brazil: Evidence from the special survey on tobacco addiction. *Economia Aplicada*, *23*(2), 127–144. doi: 10.11606/1980-5330/ea142568.

Lee, Y., & Lee, K.-S. (2019). Factors related to smoking status among young adults: An analysis of younger and older young adults in Korea. *Journal of Preventive Medicine and Public Health = Yebang Uihakhoe Chi*, *52*(2), 92–100. doi: 10.3961/jpmph.18.201.

Leigh, J. P., & Berger, M. C. (1989). Effects of smoking and being overweight on current earnings. *American Journal of Preventive Medicine*, *5*(1), 8–14. doi: 10.1016/s0749-3797(18)31122-x.

Levine, P., Gustafson, T. A., & Velenchik, A. D. (1997). More bad news for smokers? The effects of cigarette smoking on wages. *ILR Review*, *50*(3), 493–509. doi: 10.1177/001979399705000307.

Lokshin, M. (2006). *Forgone earnings from smoking: Evidence for a developing country* (p. 4018). World Bank Group. Available from: https://publications.hse.ru/en/preprints/66298596

Mincer, J. (1974). *Schooling, experience, and earnings*. NBER Books, National Bureau of Economic Research. Available from: https://econpapers.repec.org/bookchap/nbrnberbk/minc74-1.htm

Munasinghe, L., & Sicherman, N. (2006). Why do dancers smoke? Smoking, time preference, and wage dynamics. *Eastern Economic Journal*, *32*(4), 595–616.

Ñopo, H. (2008). Matching as a tool to decompose wage gaps. *The Review of Economics and Statistics*, *90*(2), 290–299. doi: 10.1162/rest.90.2.290.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review*, *14*(3), 693–709. doi: 10.2307/2525981.

Pereira, R. M., & Oliveira, C. A. de (2016). Os diferenciais de salário por gênero no Rio Grande do Sul: uma aplicação do Modelo de Heckman e da decomposição de Oaxaca-Blinder. *Redes*, *21*(1), 148–173. doi: 10.17058/redes.v21i1.3583.

Pereira, R. M., & Oliveira, C. A. de (2017). Discriminação por gênero no mercado de trabalho local: Um estudo para os municípios do Rio Grande do Sul. *Análise Econômica*, *35*(68). doi: 10.22456/2176-5456.41462.

Psacharopoulos, G., & Tzannatos, Z. (1992). *Latin American women's earnings and participation in the labor force*. Washington, DC: World Bank Publications.

Rondina, R. de C., Gorayeb, R., & Botelho, C. (2007). Características psicológicas associadas ao comportamento de fumar tabaco. *Jornal Brasileiro de Pneumologia*, *33*(outubro), 592–601. doi: 10.1590/S1806-37132007000500016.

Salardi, P. (2013). An analysis of pay and occupational differences by gender and race in Brazil - 1987 to 2006. Economics PhD Theses. Department of Economics, University of Sussex Business School. Available from: https://econpapers.repec.org/paper/sussusphd/0513.htm

Silveira, L. S., & Siqueira, N. L. (2021). Segregação ocupacional e diferenciais de renda por gênero e raça no Brasil: uma análise de grupos etários. *Revista Brasileira de Estudos de População*, *38*(julho), 1–22. doi: 10.20947/S0102-3098a0151.

Strauss, J., & Thomas, D. (1998). Health, nutrition, and economic development. *Journal of Economic Literature*, *36*(2), 766–817.

Uhr, D. de A. P., Parfitt, R., Uhr, J. Z., & Ely, R. (2021). O efeito do tabagismo sobre a produtividade no trabalho dos brasileiros. *Revista Brasileira de Economia de Empresas*, *21*(1). Available from: https://portalrevistas.ucb.br/index.php/rbee/article/view/11561

Van Ours, J. C. (2004). A pint a day raises a man's pay; but smoking blows that gain away. *Journal of Health Economics*, *23*(5), 863–886. doi: 10.1016/j.jhealeco.2003.12.005.

Vaz, D. V. (2018).Diferenças salariais por gênero no setor público brasileiro no período 2002-2015: Magnitude, evolução e determinantes. *Acta Scientiarum. Human and Social Sciences*, *40*(2), e41507–e41507. doi: 10.4025/actascihumansoc.v40i2.41507.

Wang, H., Cheng, Z., & Smyth, R. (2013). Compensating wage & income differentials for occupational risk: Evidence from migrant workers in China's Pearl river Delta. Monash Economics Working Papers, Monash Economics Working Papers, maio. Available from: https://ideas.repec.org/p/mos/moswps/2013-20.html

Zarkin, G. A., French, M. T., Mroz, T., & Bray, J. W. (1998). Alcohol use and wages: New results from the national household survey on drug abuse. *Journal of Health Economics*, *17*(1), 53–68. doi: 10.1016/s0167-6296(97)00023-4.

Corresponding author
Rafael Mesquita Pereira can be contacted at: rafaelmesquita@furg.br

Appendix

Authors	Location (country)	Methodology	Period (years)	Relation/differential against smokers
Leigh and Berger (1989)	USA	OLS	1973	−3.5[a]%
Levine *et al.* (1997)	USA	OLS	1984	−4.2%
		Difference between siblings	1991	−6.9%
			1984	−8.0%
			1991	−8.1%
			Pooling	−7.9%
Heineck and Schwarze (2003)	Germany	LS2S	1998	−2.5%
		Panel	1998–2001	−0.7[a]%
Van Ours (2004)	Netherlands	OLS	2001	−10.7%
		LS2S	2001	−16.6%
Auld (2005)	Canada	ML	1985 e 1991	−8.3%
		FIMSL-VI	1985 e 1991	−24.0%
Munasinghe and Sicherman (2006)	USA	Dynamic (1)	1979–1994	−0.5%
		Dynamic (2)	1979–1994	−0.2%
		Dynamic (3)	1979–1994	−0.2%
Lokshin (2006)	Albania	OLS	2005	−4.8%
		LS2S	2005	−25.6%
Braakmann (2008)	UK	OLS	1991–2005	−3.1%
		Panel-FE	1991–2005	−0.9[a]%
		Panel-IV	1991–2005	−0.2[a]%
Grafova and Stafford (2009)	USA	OLS	1986	−3.4%
			1999	−9.2%
			2001	−10.9%
		OLS-pooled	1986–2001	−7.4%
		OLS-FE	1986–2001	−0.9%
Anger and Kvasnicka (2010)	Germany	OLS	2002	−4.5%
		LS2S	2002	−9.9%
Hotchkiss and Pitts (2013)	USA	HM	1992–2011	−24%
Böckerman *et al.* (2015)*	Finland	OLS	1990–2004	−2.16%
		Twins	1990–2004	−1.41%
		Twins-DZ	1990–2004	−1.30%
		Twins-MZ	1990–2004	−1.85%
Almeida and Araújo Júnior (2017)	Brazil	QRIV	2008	−15.2% to −36.5%
Justus *et al.* (2019)	Brazil	OBD	2008	−29.7% (W) and −24.2% (M)
Uhr *et al.* (2021)	Brazil	PSM	2013	−7.89%
Bhai (2020)	USA	OLS-FE	1996 2006 2014	−15% to −16%
Darden *et al.* (2021)	USA	Dynamic	1997	−15.9% (W) and −15.2% (M)

Note(s): Legend (Methodology): OLS = Ordinary Least Squares; LS2S = Least Squares in two stages; ML = Maximum Likelihood; FIMSL = full information maximum simulated likelihood; Twins = model of differences between twins; DZ = dizygotic twins; MZ = monozygotic twins. In Munasinghe and Sicherman (2006), Dynamic (1), Dynamic (2), and Dynamic (3) represent uncontrolled regression, with limited controls and complete controls; QRIV = Conditional Quantile Regression with Instrumental Variables; HM = Heckman Model (Method); OBD = Oaxaca-Blinder decomposition; PSM – Propensity Score Matching. *a* = Not statistically significant; * Explanatory variable is the number of cigarette packets consumed
Source(s): Adapted/complemented by Almeida and Araújo Júnior (2017)

Table A1. Synthesis of the international and Brazilian literature on smoking in the labor market – wage/income ratio/differential