TEXTO PARA DISCUSSÃO

ISSN 0103-9466

334

The wage gap between smokers and non-smokers in Brazil

Marcelo Justus, Elder G. Sant'Anna, Eloá S. Davanzo, Gustavo C. Moreira

Março 2018

instituto de economia
Desde 1968

UNICAMP

The wage gap between smokers and non-smokers in Brazil

Marcelo Justus[1], Elder G. Sant'Anna[2], Eloá S. Davanzo[3], Gustavo C. Moreira[4]

Abstract

The aim of this study is to investigate the hypothesis that smoking reduces earnings. Data from the Special Survey on Tobacco Addiction carried in Brazil were used. We contribute to the previous literature by addressing, at the same time, the smoking endogeneity and sample selection bias resulting from the decision to participate in the labor market. Furthermore, we also investigate the role of observable and non-observable characteristics on wages differential between smokers and non-smokers. Our results support the hypothesis that smoking reduces wage. We also applying the Oaxaca-Blinder decomposition to measure the earnings gap between smokers and non-smokers, separating contributions of group differences into observable and non-observable characteristics.

Keywords: smoking, labor market, discrimination, human capital
JEL classification: I12, J24

1. Introduction

Smoking is a toxicomania characterized by physical and psychological addiction to nicotine, one of the 4,720 toxic substances contained in tobacco. Research associates smoking with at least fifty types of diseases, most of which are chronic and severe ones. It is estimated that the life expectancy of a smoking individual is at least 10 years shorter than that of a non-smoking one.

Cigarettes are consumed by almost 1 billion adults in the world. Without a doubt smoking is associated with a higher risk of developing serious diseases such as cancer, emphysema and cardiovascular diseases. As a consequence, it causes the death of more than six million smokers every year. Given this high number of deaths, the World Health Organization considers smoking the main cause of preventable death in the world.

In Brazil, tobacco use often takes the form of consumption of manufactured cigarettes. In 2013, there were 21.4 million smokers in Brazil, 18.5 million of whom were daily smokers.

Email addresses: mjustus@g.unicamp.br (Marcelo Justus), eldergenerozo@gmail.com (Elder G. Sant'Anna), eloadavanzo@gmail.com (Eloá S. Davanzo), gum@kth.se (Gustavo C. Moreira)

[1]Institute of Economics at the University of Campinas, São Paulo, Brazil.

[2]PhD student at the University of São Paulo, Brazil.

[3]PhD student in the Institute of Economics at the University of Campinas, São Paulo, Brazil

[4]Visiting scholar at the Department of Urban Planning and Environment, School of Architecture and the Built Environment, KTH Royal Institute of Technology, Stockholm, Sweden.

According to the Ministry of Health, about 200,000 deaths per year are related to tobacco consumption.

In economic terms, smoking increases expenditures with health care and decreases productivity due to morbidity and premature death, substantially reducing the stock of human capital in society (World Bank, 1999). According to estimates of World Bank (2015), when indirect costs are also taken into account, smoking accounts for losses estimated at US$ 1 trillion worldwide every year. The World Health Organization (WHO) considers smoking as the main cause of preventable death in the world.

Economic losses related to tobacco use are significant and affect mainly underdeveloped and developing countries, which account for 80 percent of the worldwide smokers' population (World Bank, 2015). The main tangible costs of tobacco use may be related to health care expenditures, loss of production due to death and illness and reduction of productivity. According to Eriksen et al. (2013), these tangible costs generate an annual expenditure of US$ 1 trillion globally. Moreover, there are also intangible costs, i.e. the suffering of those who get sick because they are smokers or passive smokers and deaths caused by smoking diseases.

Regarding the private economic costs of smoking, a clear reduction in smoker's disposable income can be observed, as smokers not only buy cigarettes but may also be forced spend money on medical treatment due to smoking. However, the costs associated with the effects of smoking on the labor market are not so evident.

There are few empirical studies published supporting the hypothesis that smoking can severely affect labor market outcomes through multiple channels, such as, wage (Van Ours, 2004; Lye and Hirschberg, 2004; Levine et al., 1997), absence (Leigh, 1995; Ault et al., 1991), accidents (Leistikow et al., 2000) and less chance of participation (Lee et al., 1991). Indirect effects can also be caused by non-observed preferences and by the behavior of persistent smokers (Grafova and Stafford, 2009). With regard to wage, these studies conclude from earnings equations where smoking is controlled. Ideally, these studies would have to solve two serious problems that arise in the attempt to identify the causal effect of smoking in wages, namely: sample selection – resulting from the decision to participate in the labor market – and smoking endogeneity. Undoubtedly, the endogeneity imposes major difficulty in order to identifying the effect of smoking on earnings.

On the one hand, some few empirical studies[5] have found that smoking reduces earnings (Levine et al., 1997; Auld, 1998; Lee et al., 1999; Braakmann, 2008; Anger and Kvasnicka, 2010). On the other hand, other few studies did not reject that smoking does not affect earnings (Van Ours, 2004; Heineck and Schwarze, 2003; Braakmann, 2008). However, none of this studies solved at the same time both selection sample and smoking endogeneity problems.

Recently, de Almeida and de Araújo Júnior (2017) found, using a instrumental quantile regression approach, that brazilians workers who smoke receive from 15.2% to 36.5% less than others workers. Our study is based on the same data, but we go a little further. Even if our estimations don't take into account the quantile effects neither a control to ex-

[5]We consider here only studies published in journals, i.e. working papers not were considered.

smokers, as the previous study, we consider gender differences in the labor market and we exploit a different set of instruments. Another novel contribuition of our approach lies in the decomposition of earnings applying the Oaxaca-Blinder decomposition to measure the earnings gap between smokers and non-smokers, which has been poorly investigated in the literature.

First exercise was disaggregating the sample according to gender and to smokers and non-smokers categories. We found that the difference between smokers and non-smokers workers' earnings is statistically significant for both men and women. The group of non-smoking men earns 25 percent more than that of smoking men. In the case of women, this difference is lower, 16 percent. But, is smoking harmful to labor market outcomes? Thus, the aim of this study is to test the hypothesis that *smoking reduces earnings*. For this purpose, an empirical strategy to jointly deal with smoking endogeneity and sample selection was applied. This reduction in wages can occur through mechanisms such as increased absenteeism, reduced productivity, and discrimination in the labor market. An important discussion on the discrimination hypothesis is presented by (Levine et al., 1997). The authors analyzed the effects of smoking on income and raised the hypothesis that discrimination occurred over the years as public intolerance to smoking became gradually stronger. Many employers institutionalized their own policies to ban smoking from their premises and some institutions adopted employment policies to hire non-smokers only. Those authors argue that in such scenario discriminatory employment practices can be adopted and reduce the wages of smokers and their expectations of employment.

This paper is structured as follows. Section 2 presents the data and sample. Section 3 describes the methodological procedures. Sections 4 and 5 show the main results and conclusion, respectively.

2. Data and Sample

We use data from the Special Survey on Tobacco Addiction (PETab, in the Brazilian acronym), which was jointly carried out with the 2008 National Household Sampling Survey (2008 PNAD, in the Brazilian acronym). The survey was conducted through a partnership between the Brazilian Institute for Geography and Statistics, the Ministry of Health, the National Cancer Institute, the Health Surveillance Secretariat and the National Health Surveillance Agency.

PNAD is a multipurpose random household survey that investigates several socioeconomic characteristics of the population, some on a permanent basis and others with variable periodicity, such as health status and smoking habit. An interesting advantage of this survey lies in its national coverage and, additionally, because this survey collects data on many other variables related to household structure and socioeconomic aspects of household members (labor, wage, education, housing characteristics, age, etc.). In addition to their representativeness at national level, these data cover several aspects related to tobacco addiction, such as: tobacco use, attempts to quit smoking, exposure to tobacco, access to awareness-raising campaigns and perceptions about the risks of smoking, as well as aspects related to buying cigarettes and tobacco products.

The data was collected from a sub-sample of households surveyed through the 2008 PNAD, covering individuals aged 15 and above in about 51,000 Brazilian households. The individuals included in that sub-sample answered questions related to the use of tobacco products, to their attempts to quit smoking, to their exposure to smoke and to their access to awareness-raising campaigns and to information on the risks of smoking, among other issues related to the main topic. For other people interviewed through the survey, information is also available on the habit of smoking, type of tobacco product used and amount consumed.

It should be noted that the PETab survey is carried out in Brazil as part of an initiative of the World Health Organization and of the Centers for Disease Control and Prevention. This partnership was established with the aim of promoting part of a survey conducted in 14 countries, including Brazil, entitled Global Adults Tobacco Survey.

To make the sample suitable for empirical modeling, we excluded all individuals under 18 years old or over 60 years old to reduce the labor market participation selection bias problem. Thus, our sample is restricted to individuals in the 18-60 age bracket. We also excluded individuals with ill-defined occupations; individuals who were still studying, observations for which the difference between the actual age of the person and that at which he or she began to work is negative due to an error in the database; individuals who worked but had no earnings; individuals who did not state their income; and individuals with wages in excess of R$ 100,000 (Brazilian currency).

After applying the above-mentioned filters to the sample, and with missing values, our empirical exercises began with two subsamples: women and men making up 54,772 and 81,974 observations, respectively. The sample factor associated to each observation was used for weighted.

3. Methodology

Earnings equation

The modelling exercise began with the standard linear regression model estimated by OLS,

$$y_i = \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i,$$

where y_i is the logarithm of hourly earnings from the main job for individual i; $\mathbf{x}_i$ is a row vector containing a dummy variable labeled by `smoke` – which assumes value 1 if the individual is a smoker and 0 otherwise, other control variables (which will be described later) and a constant; $\boldsymbol{\beta}$ is a column vector of coefficients and ε_i is the random disturbance with $\varepsilon_i \sim N(0, \sigma_\varepsilon)$.

This earnings equation is estimated separately by gender. Table 1 shows the variables used in the earnings equations.

As usual in earnings equations, education was proxied by years of schooling. We also considered the existence of a threshold effect, besides the years of schooling variable based on previous evidence from Brazil found by Hoffmann and Simão (2009) and Justus et al. (2015).

The returns on education are positive, suggesting that increases in earnings are substantially higher from 10 years of schooling. It should be noted that the first year of schooling

yielding the highest return is the 11th grade, the last grade of high school. Therefore, we considered the existence of a threshold effect, besides the years of schooling variable, and included the variable $S^{\lambda} = Z(S - \lambda)$ in the specification, where $\lambda = 10$ is the threshold, i.e. the value of schooling from which the return on education increases, and Z is a dummy variable that assumes value 0 for $S \neq \lambda$ and value 1 for $S > \lambda$.

Other control variables are experience in the labor market, as usually measured using a typical mincer model by *proxy* defined by the difference between the actual age of the person and that at which he or she began to work, and the square of this variable; a dummy variable to distinguish between white (Caucasian, Asian people) and non-white (black, mulatto, indigenous people); a dummy variable to distinguish between residence in an urban or rural area; a dummy variable for labor union membership; two dummy variables to distinguish between three activity sectors: agriculture (base group), industry and services; and twenty-six dummy variables to control for differences between states; three dummy variables to distinguish between two positions: employer, employee and self-employed (base group). Controls were included in all models for Brazil's 27 federated units (26 dummy variables).

Tabela 1: Variables used in the earnings equations

Variable	Definition
`log hourly earnings`	Logarithm of hourly earnings from the main job
`smoker`	1 if is a smoker and 0 otherwise
`non-labor income`	1 if earns non-labor income and 0 otherwise
`schooling`	Years of schooling
`threshold`	Threshold for schooling
`experience`	Years of experience
`experience squared`	Years of experience squared
`labor union`	1 if is a labor union membership and 0 otherwise
`urban`	1 if lives in an urban area and 0 otherwise
`agriculture`	1 if works in the agricultural sector and 0 otherwise
`industry`	1 if works in the industry sector and 0 otherwise
`service`	1 if works in the service sector and 0 otherwise
`self-employment`	1 if is self-employed and 0 otherwise
`employee`	1 if is an employee and 0 otherwise
`employer`	1 if is an employer and 0 otherwise
`white`	1 if is white and 0 otherwise

Source: prepared by the authors.

We know that OLS estimates for earnings equation may be biased due to the individual's decision to participate in the labor market. Thus, the second step in our empirical strategy in order to identify the effect of smoking was to apply Heckman's procedure to correct the sample selectivity bias (see Heckman, 1979).

The participation equation contains the same regressors as the earnings equation, except for the dummy variables for labor union membership, position and activity sectors. We also included other personal and family characteristics: a dummy variable for the existence of non-labor income (e.g., conditional cash transfer programs), which is 1 if the person has such an income and 0 otherwise; a dummy variable for marital status, which is 1 if the man is married and 0 otherwise; a dummy variable for children living in the same household, which

is 1 if there are children in the household and 0 otherwise; a dummy variable for position in the family, which is 1 if the man is the head of the family and 0 otherwise.

Finally, as suggested in the literature, it is possible that the variable `smoke` is endogenous. This imposes more difficulty in order to identifying the hypothetical causal effect of smoking on earnings. Thus, in the third step was applied an empirical strategy to deal at the same time with the sample selection bias resulting from the decision to participate in the labor market and smoking endogeneity (see Wooldridge, 2010, 567-570).

Two robust instruments for variable `smoker`, which is now treated as an endogenous variable, were used: i) number of smokers living in the same household (`family`) and ii) a dummy variable to indicate whether the individual had been already diagnosed with asthma or bronchitis (`respiratory`). It is assumed that both variables are correlated with smoking but do not affect earnings.

In short, the procedure was performed in three steps. First, the selection equation (participation equation) was estimated using all observations in the probit model and taking into account the two instruments cited above, besides the previously mentioned regressors. Second, the estimated inverse Mills ratios for all observations were calculated based on this equation. Third, using the selected subsample for which wages and smoking were observed we estimated the earnings equation also taking into account the inverse Mills ratios besides the controls variables cited previously. The parameters of this overidentified model were estimated by the optimal GMM.

Oaxaca-Blinder decomposition

If our results, in Section 4, indicate that smoking negatively affects the earnings, then, we are able to decompose this differential in order to meansure the influence of observable and non-observable characteristics. Oaxaca-Blinder decomposition (Oaxaca, 1973; Blinder, 1973) for smoking and non-smoking individuals was applying for this purpose.

Decomposition is performed in two stages. In the first one, earning equations are estimated for each of the groups. Once this is done, the difference between the logarithm of average earnings between workers in the two groups is calculated as

$$\begin{aligned} \mathrm{D} &= E(y_s) - E(y_{ns}) \\ &= E(\boldsymbol{\beta}_s'\mathbf{x}_s + \varepsilon) - E(\boldsymbol{\beta}_{ns}'\mathbf{x}_{ns} + \varepsilon) \\ &= E(\mathbf{x}_s)'\boldsymbol{\beta}_s' - E(\mathbf{x}_{ns})'\boldsymbol{\beta}_{ns}' \end{aligned}$$

where $E(\varepsilon) = 0$ was used. According Jann (2008), this equation can be rearranged from a twofold decomposition as

$$D{=}[E(\mathbf{x}_s) - E(\mathbf{x}_{ns})]'\boldsymbol{\beta}^* + E(\mathbf{x}_s)'(\boldsymbol{\beta}_s - \boldsymbol{\beta}^*) + E(\mathbf{x}_{ns})'(\boldsymbol{\beta}^* - \boldsymbol{\beta}_{ns})$$

where the term $[E(\mathbf{x}_s) - (\mathbf{x}_{ns})]'\boldsymbol{\beta}^*$ represents the earning differential that is explained by the mean observable characteristics of smoking and non-smoking individuals. The other component on the right side of this equation refers to the portion not explained by these characteristics. According to Jann (2008), the coefficient $\boldsymbol{\beta}^*$ represents a vector of coefficients related to non-discrimination.

4. Results and Discussion

Table 2 shows the earnings equations estimated by OLS, Heckman's procedure and IV-GMM with correction for sample selection bias. The selection equation estimates for Heckman's procedure, and the results of the first-stage regression of the endogenous variable `smoker` are available upon request.

In this study, we are interested in the variable `smoker`. However, it should be noted that for all control variables (e.g., schooling and experience) the results are the ones usually observed in studies on earnings determinants in the Brazilian labor market and international literature.

Since dependent variable is the natural logarithm of earnings, if c is the estimated value of the conditional marginal effect, the estimated percentage change in earnings due to change in a dummy variable is $[\exp(c) - 1]100$. For second model, which was estimated based on sample selection correction, we computed the conditional marginal effect. The effects are evaluated at the sample means of the observations use to fit the model.

Based on IV-GMM estimates with correcting for sample selection bias, we found that *smoking had a greater impact on earnings*. Smoking reduces wage about 30.6 and 26.1 percent for man and woman, respectively. This is an serious economic consequence of being a smoker.

The higher magnitude after controlling for smoking endogeneity was also verified in previous studies. In Auld (1998), for example, control for simultaneity between wages and smoking suggests that smokers earn from about 20 percent to 67 percent less than non-smokers. This incremental effect after using instrumental variables was also observed by Van Ours (2004) when analyzing men's earnings. It is worth remembering that none of this studies applied a correction for sample selection bias.

Our results for the reductions observed in the wages of smokers as compared to those of non-smokers are corroborated by the literature, which provides several examples of ways by which smoking influences labor income. Anger and Kvasnicka (2010) shows that the wages paid to smokers can decrease due to their reduced productivity resulting from high rates of absenteeism and health problems or due to potential discrimination of smokers by employers and co-workers. Damages to one's health, however, can be irreversible. Smoking can therefore have a negative impact on both an individual's current capacity and on his or her wages in the future.

In relation to absenteeism, Halpern et al. (2001) show that the rate of absenteeism measured for workers who smoke currently was higher than that calculated for those who never smoked. It should be mentioned that in the group of individuals who were smokers in previous periods, absenteeism declined as they stopped smoking. With similar results, Weng et al. (2013) found that current smokers face a 33-percent higher risk of absenteeism than non-smokers. Those in the former group were absent from work for 2.64 more days per year on average than those in the latter.

In terms of productivity, smoking can reduce the net productivity of workers due to its effects on their ability to perform manual tasks (Levine et al., 1997) and to the high absenteeism rates recorded for smoking workers and/or their lower physical and mental

Tabela 2: Earnings equations using OLS, Heckman's procedure and IV-GMM with correction for sample selection bias: Brazilian individuals aged from 18 to 60 years old, by gender

	Man			Woman		
Variables	OLS	Heckit	IV-GMM	OLS	Heckit	IV-GMM
constant	0.115***	0.231***	0.272***	−0.0845**	−0.132***	−0.0427***
	(0.0239)	(0.0248)	(0.0264)	(0.0360)	(0.0457)	(0.0544)
smoker	−0.0737***	−0.0634***	−0.182***	−0.0335***	−0.0329***	−0.340***
	(0.00622)	(0.00637)	(0.0386)	(0.00959)	(0.00961)	(0.0591)
schooling	0.0416***	0.0385***	0.0361***	0.0367***	0.0382***	0.0330***
	(0.00105)	(0.00107)	(0.00122)	(0.00146)	(0.00169)	(0.00215)
threshold	0.164***	0.166***	0.166***	0.181***	0.180***	0.181***
	(0.00327)	(0.00327)	(0.00328)	(0.00328)	(0.00328)	(0.00332)
experience	0.0335***	0.0263***	0.0263***	0.0234***	0.0242***	0.0255***
	(0.000780)	(0.000877)	(0.00119)	(0.000920)	(0.00102)	(0.00111)
exp. squared	−0.000440***	−0.000294***	−0.000290***	−0.000321***	−0.000338***	−0.000352
	(0.0000162)	(0.0000180)	(0.0000241)	(0.0000204)	(0.0000226)	(0.0000242)
white	0.123***	0.117***	0.115***	0.112***	0.111***	0.106***
	(0.00573)	(0.00586)	(0.00576)	(0.00691)	(0.00694)	(0.00712)
urban	0.131***	0.169***	0.173***	0.145***	0.148***	0.159***
	(0.00913)	(0.00946)	(0.00979)	(0.0140)	(0.0140)	(0.0142)
labor union	0.159***	0.155***	0.150***	0.207***	0.207***	0.206***
	(0.00704)	(0.00704)	(0.00714)	(0.00926)	(0.00924)	(0.00936)
industry	0.342***	0.337***	0.335***	0.155***	0.155***	0.135***
	(0.0108)	(0.0107)	(0.0109)	(0.0270)	(0.0270)	(0.0274)
service	0.297***	0.295***	0.296***	0.267***	0.267***	0.258***
	(0.00998)	(0.00992)	(0.0100)	(0.0253)	(0.0253)	(0.0257)
employer	0.611***	0.611***	0.604***	0.642***	0.642***	0.636***
	(0.0161)	(0.0161)	(0.0161)	(0.0260)	(0.0260)	(0.0263)
employee	0.0474***	0.0435***	0.0499***	0.0351***	0.0358***	0.0331***
	(0.00735)	(0.00731)	(0.00736)	(0.0109)	(0.0109)	(0.0110)
mills ratio		−0.294***	−0.315***		0.036*	0.00252
		(0.0152)	(0.0319)		(0.0209)	(0.0274)
Number of observations	70,982	82,516	70,982	46,801	82,035	46,801
Censored observations		11,534			35,234	
Uncensored observations		70,982			46,801	

Notes: Robust standard errors are given in parentheses.
***, ** and * denote significance at 1% level, 5% level and 10% level, respectively.
Dummy variables for Brazilian states were used.

resistance (Grafova and Stafford, 2009). Considering subjective productivity (productivity as assessed by others and personal life satisfaction), Halpern et al. (2001) showed significant trends with higher figures for those who never smoked in their life, lower figures for current smokers, and intermediate figures for individuals who were smokers in previous periods.

In view of the evidence of effects of smoking on earnings in all the estimated models, we decomposed this differential in order to meansure the influence of observable and non-observable characteristics. Oaxaca-Blinder decomposition (Oaxaca, 1973; Blinder, 1973) for smoking and non-smoking individuals was applying for this purpose. Table 3 show the results.

Tabela 3: Oaxaca-Blinder decomposition for logarithm of hourly earnings

	Man			Woman		
	OLS	Heckit	IV-GMM	OLS	Heckit	IV-GMM
Smoker	1.216***	1.193***	1.275***	1.130***	1.233***	0.997***
Non-smoker	1.456***	1.549***	1.530***	1.283***	1.291***	1.228***
Difference	−0.240***	−0.356***	−0.255***	−0.152***	−0.059	−0.232***
Explained	−0.176***	−0.202***	−0.169***	−0.126***	−0.130***	−0.132***
Unexplained	−0.063***	−0.154***	−0.086***	−0.026***	0.071	−0.100**

Note: ***, ** and * denote significance at 1% level, 5% level and 10% level, respectively.

Considering the results from decomposing the model with instrumental variables, around two-thirds of the wage differential between smoking and non-smoking men are due to their observable characteristics, whereas for women this portion amounts to about 57 percent. As far as we know, the only study that sought to analyze the decomposition of wage differentials between smokers and non-smokers was one conducted by Hotchkiss and Pitts (2013). The authors found that the differential between the groups was of about 24 percent, two-thirds of which were explained by differences in observable characteristics.

Our result corroborates by Becker and Murphy (1988) theory of rational addiction, which suggests a higher intertemporal preference for the present for individuals with an addiction of some kind. Considering their higher preference for the present, smokers have lower incentives to invest in human capital, since they will not be able to enjoy its returns for the same period of time as non-smokers. There was virtually no change in the magnitude of wage differentials between smokers and non-smokers that can be explained by observable characteristics for both males and females and regardless of the model used for the decomposition.

Still regarding the decomposition, we could observe a significant effect of non-observable characteristics on the earnings of smokers compared non-smokers for both men and women. This effect may result, for example, from a certain productive heterogeneity not controlled for by the model's exogenous variables or even from discrimination toward smokers in the labor market. This significance can be justified by studies that confirm that smokers generate higher costs for companies. Smoking workers can be more expensive for their employers due to their increased absenteeism, higher health insurance premium, higher maintenance costs and negative effects on the company's image. Due to health problems associated with smoking, smokers themselves may prefer jobs that provide partial or full health insurance to higher-wage jobs (Levine et al., 1997). The greater magnitude of the non-observable effect

on women's earnings deserves special mention.

In addition, when we decomposed the wage differential between smokers and non-smokers, we saw that about two-thirds of this difference was due to observable characteristics for men, while for women the figure was 57 percent.

Based on these results and comparing gender differences, we saw that smoking women suffer a wage penalty that is higher than that suffered by smoking men by 12 percentage points. Furthermore, assuming that the non-explained portion of this wage differential is due to discrimination toward smokers, we found evidence that wage discrimination toward smoking women is greater than toward smoking men.

Robustness checks

Regarding to our instrument variables, as expected the coefficient for `family` is positive and highly significant statistically ($p < 0.001$) for both man ($\beta = 0.150$) and woman ($\beta = 0.122$). The `respiratory` variable is statistically significant at 1% level only in the first-stage equation estimated for women.

After the GMM estimation, we performed a robust test of endogeneity (orthogonality conditions). The GMM C statistic is χ^2 distributed, with one degree of freedom, under the null hypothesis that the regressor is exogenous. We apply the test to our model with one potentially endogenous regressor, `smoker`, intrumented by `family` and `respiratory`. The stastistic is $\chi^2 = 10.41$ ($p = 0.0013$) and $\chi^2 = 29.13$ ($p = 0.0000$) for man and woman, respectively. For both genders, the statistical test leads to strong rejection of the null hypothesis that `smoker` is an exogenous variable in the earnings equations. We conclude that is endogenous.

We also applied Hansen's J test to test the validity of the overidentified restrictions. The statistic is $\chi^2 = 0.78$ ($p = 0.3782$) and $\chi^2 = 0.36$ ($p = 0.5468$) for man and woman, respectively. Because $p > 0.05$, we do not reject the null hypothesis. The failure to reject H_0 is interpreted as indicating that at least one of the instruments is valid. We conclude that overidentifying is valid. For details about this endogeneity and overidentification tests see (Cameron and Trivedi, 2009).

Finally, we estimated the earning equations once again disregarding variables related to activity sector and position in the occupation, as these are potentially endogenous characteristics, for check the robustness of the estimates. For all variables the estimates are virtually the same as compared to those shown in the Table 2. These additional results are available upon request.

5. Concluding remarks

We do not reject the hypothesis that *smoking reduces earnings*, i.e., smoking really harms wage. Our results are in line with those recently presented by de Almeida and de Araújo Júnior (2017), but our estimates are higher, i.e., men (women) who smokes receive less 30.6% (26.1%) than others non-smokers workers. Furthermore, when we decomposed the wage differential between smokers and non-smokers, we saw that a part significant of this

difference is due to observable characteristics for both men and women. This final exercise provide an evidence that women smokers suffer more wage discrimination than men smokers.

It is well known how harmful smoking is. We provide evidence that the private costs of smoking are not limited to health-related aspects, but that they also affect wage. Productivity on the decline, high absenteeism rates, and the higher costs borne by employers are possible reasons referred to in the literature. This paper contributes to the literature by showing that the decrease in incomes is partly explained by discrimination in the labor market, especially against women. We provide more evidence of smoking harm in an unprecedented way by addressing the topic in the context of a developing country with approximately 30 million smokers and 200,000 deaths caused by tobacco use every year. There is no doubt that the best way to eliminate this negative influence is by quitting smoking. Quitting cigarette addiction is a decision that involves multiple factors related to individual characteristics and depends on public policies specifically designed to assist individuals in quitting smoking.

Acknowledgements

Marcelo Justus thanks the National Council for Technological and Scientific Development (CNPq) for financial support to conduct this research (process number 442483/2014-7), and for his Productivity in Research Grant.

References

Anger, S. and M. Kvasnicka (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.

Auld, M. C. (1998). Wages, alcohol use, and smoking: Simultaneous Estimates. pp. 1–29.

Ault, R. W., R. B. Ekelund Jr, J. D. Jackson, R. S. Saba, and D. S. Saurman (1991). Smoking and absenteeism. *Applied Economics 23*(4), 743–754.

Becker, G. S. and K. M. Murphy (1988). The theory of rational addiction. *Journal of Political Economy 96*(4), 675–700.

Blinder, A. S. (1973). Wage discrimination: reduced form and structural estimates. *Journal of Human Resources*, 436–455.

Braakmann, N. (2008). The smoking wage penalty in the United Kingdom: Regression and matching evidence from the british household panel survey. *Working Paper Series in Economics* (96).

Cameron, A. C. and P. K. Trivedi (2009). *Microeconometrics using Stata* (2 ed.). College station: Stata Press.

de Almeida, A. T. C. and I. T. de Araújo Júnior (2017). Tabagismo e penalização salarial no mercado de trabalho brasileiro. *Economia Aplicada 21*(2), 249.

Eriksen, M., J. Mackay, H. Ross, et al. (2013). *The tobacco atlas.* Number 4. American Cancer Society.

Grafova, I. B. and F. P. Stafford (2009). The wage effects of personal smoking history. *ILR Review 62*(3), 381–393.

Halpern, M. T., R. Shikiar, A. M. Rentz, and Z. M. Khan (2001). Impact of smoking status on workplace absenteeism and productivity. *Tobacco Control 10*(3), 233–238.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica 47*(1), 153–61.

Heineck, G. and J. Schwarze (2003). Substance use and earnings: The case of smokers in Germany. *IZA Discussion Paper* (173).

Hoffmann, R. and R. C. S. Simão (2009). Determinantes do rendimento das pessoas ocupadas em Minas Gerais em 2000: o limiar no efeito da escolaridade e as diferenças entre mesorregiões. *Nova Economia 15*(2), 35–62.

Hotchkiss, J. L. and M. M. Pitts (2013). Even one is too much: The economic consequences of being a smoker. *FRB Atlanta Working Paper Series*.

Jann, B. (2008). The Blinder-Oaxaca decomposition for linear regression models. *The Stata Journal 8*(4), 453–479.

Justus, M., H. Kawamura, and A. L. Kassouf (2015). What is the best age to enter the labor market in Brazil today? *EconomiA 16*(2), 235–249.

Lee, A. J., I. K. Crombie, W. C. Smith, and H. D. Tunstall-Pedoe (1991). Cigarette smoking and employment status. *Social Science and Medicine 33*(11), 1309–1312.

Lee, Y. L. et al. (1999). Wage effects of drinking and smoking: An analysis using Australian twins data. *Working paper* (22), 1–29.

Leigh, J. P. (1995). Smoking, self-selection and absenteeism. *The Quarterly Review of Economics and Finance 35*(4), 365–386.

Leistikow, B. N., D. C. Martin, and C. E. Milano (2000). Fire injuries, disasters, and costs from cigarettes and cigarette lights: A global overview. *Preventive Medicine 31*(2), 91–99.

Levine, P. B., T. A. Gustafson, and A. D. Velenchik (1997). More bad news for smokers? The effects of cigarette smoking on wages. *Indutrial and Labour Relations Review 50*(3), 493–509.

Lye, J. N. and J. Hirschberg (2004). Alcohol consumption, smoking and wages. *Applied Economics 36*(16), 1807–1817.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review*, 693–709.

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.

Weng, S. F., S. Ali, and J. Leonardi-Bee (2013). Smoking and absence from work: Systematic review and meta-analysis of occupational studies. *Addiction 108*(2), 307–319.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

World Bank (1999). A epidemia do tabagismo: Os governos e os aspectos econômicos do controle do tabaco. *The World Bank, agosto*.

World Bank (2015). Taxes on tobacco can save a life every six seconds. Technical report, The World Bank.