UNIVERSIDADE FEDERAL DE SÃO CARLOS
Programa de Pós-Graduação em Economia

ANA HELENA NALLIN DAVINHA

Os Impactos Socioeconômicos da Prática de Atividades Físicas no Brasil: uma análise sobre renda e defasagem idade-série .

The Socioeconomic Impacts of Physical Activities Practice in Brazil: an analysis over income and age-grade gap.

SOROCABA
2020

Ana Helena Nallin Davinha

Os Impactos Socioeconômicos da Prática de Atividades Físicas no Brasil: uma análise sobre renda e defasagem idade-série .

The Socioeconomic Impacts of Physical Activities Practice in Brazil: an analysis over income and age-grade gap.

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Economia como parte dos requisitos para a obtenção do título de Mestra em Economia.

Master Thesis presented to the Economics Post-Graduation Program in partial fulfillment of the requirements for the degree of Master of Science in Economics..

Supervisor/Orientadora: Profa. Dra. Andrea Rodrigues Ferro

Este exemplar corresponde à versão final da Dissertação de Mestrado defendida por Ana Helena Nallin Davinha e orientada pela Profa. Dra. Andrea Rodrigues Ferro.

Sorocaba
2020

Profa. Dra. Andrea Rodrigues Ferro
Presidente da Comissão Julgadora

Prof. Dr. Adelson Martins Figueiredo
Universidade Federal de São Carlos

Prof. Dr. Henrique Coelho Kawamura
Universidade Federal da Integração Latino-Americana

Education must not simply teach work
- it must teach life.

(W. E. B. Du Bois)

Agradecimentos

À minha família, pelo apoio e compreensão nos momentos de maior tensão;

Aos professores, por todas as discussões proporcionadas durante as aulas, e por não medirem esforços para esclarecerem nossas dúvidas;

À minha orientadora, por solucionar os problemas *"insolúveis"* desta pesquisa num piscar de olhos;

À UFSCar, por todos os transbordamentos de conhecimentos e debates proporcionados;

Aos meus amigos, que ao compartiharem desta mesma experiência, dividiram toda a carga de estresse que se passa nesse *"mestradinho"*;

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Resumo

DAVINHA, Ana Helena Nallin. Os Impactos Socioeconômicos da Prática de Atividades Físicas no Brasil: uma análise sobre renda e defasagem idade-série. 2020. Dissertação (Mestrado em Economia) - Universidade Federal de São Carlos, *campus* Sorocaba, Sorocaba - SP, 2020.

A prática de atividades físicas vem se tornando fonte de amplos debates tanto no meio acadêmico, como no cotidiano dos indivíduos, no que diz respeito aos seus benefícios à saúde física e mental. Apesar disso, pouco foi discutido nestas duas esferas a respeito de seus impactos socioeconômicos. Assim sendo, o objetivo do presente trabalho é analisar a prática de atividades físicas como um problema econômico, na medida em que afeta a saúde e qualidade de vida (bem-estar) dos indivíduos, consequentemente causando impactos na sua produtividade e capacidade de auferir renda. Para os adultos (aqueles maiores de 18 anos, mas com idade inferior a 65 anos), analisou-se a relação entre a prática de atividades físicas e os rendimentos individuais. Já para indivíduos em idade escolar (adolescentes, de 15 a 17 anos), como a prática de atividades físicas, além de estar inserida na grade horária das escolas, também impacta seu bem-estar físico e metal, analisou-se os impactos da prática de atividades físicas no desempenho acadêmico dos adolescentes, isto é, na sua defasagem idade-série. Para tanto, empregou-se a metodologia estatístico-econométrica como forma de análise e avaliação destas relações, utilizando a equação de Heckman para a renda e modelos de dados em contagem para a defasagem idade-série. Os resultados sobre a renda mostram que a prática de atividades físicas (PAF) afeta positiviamente não só o estoque de saúde dos indivíduos, mas também sua renda, tornando-se um problema socioeconômico que pode afetar a qualidade de vida por diversos meios. Neste sentido, o fato do indivíduo praticar atividades físicas eleva seus ganhos de renda em aproximadamente de 10.2% (IV-Probit IV-GMM) a 12.3% (BP IV-GMM), na média. Esse resultado foi obtido controlando-se as estimativas para viés de seleção e para a presença de endogeneidade entre a PAF e a situação ocupacional, e a PAF e a renda dos indivíduos. Além disso, estes impactos variam entre os sexos e entre as faixas etárias, de modo que são observados maiores impactos da PAF sobre a renda dos homens, e um comportamento em forma de U-invertido ao longo das faixas etárias, semelhante aos próprios impactos da idade e experiência no rendimento dos indivíduos (MINCER, 1958; SCHULTZ, 1961). Quanto à defasagem idade-série, os efeitos da PAF sobre o desempenho acadêmico dos adolescentes foram bastante diferentes para meninos e meninas: enquanto, para as meninas, o fato de participar de atividades físicas fora do horário escolar reduz a defasagem idade-série em 24.8%, ou 0.281 anos (cerca de 3.4 meses) em média, para os meninos, a PAF foi, na maioria dos casos, estatisticamente zero em termos de afetar sua defasagem idade-série. Adicionalmente, considerando exclusivamente os meninos que frequentavam regularmente a escola, os efeitos marginais da estimação NBRM apontam

para um impacto positivo da PAF na defasagem idade-série desses meninos, de modo a afetar negativamente seu desempenho acadêmico.
Palavras-chave: Saúde e Desigualdade; Bem-Estar; Diferenciais Salariais; Seleção Amostral; Variáveis Instrumentais.

Abstract

The practice of physical activities has become a source of ample debates both in the academic environment and in the daily life of individuals, with respect to their benefits to physical and mental health. Despite this, little has been discussed in these two spheres regarding the socioeconomic impacts of physical activity practice. Therefore, the main objective of this study is to analyze the practice of physical activities as an economic problem, insofar as it affects the health and quality of life (well-being) of individuals, consequently causing impacts on their productivity and ability to earn income. For adults (those between 18 and 65 years old), we analysed the relationship between physical activity practice and individual income earning. As for school-aged individuals (adolescents, aged 15 to 17 years old), as physical activity, besides being included in the school hours, also impacts their physical and metal well-being, we analysed the impacts of the practice of physical activities on the academic performance of adolescents, that is, on their age-grade gap. In order to do so, the statistical-econometric methodology was used as a way of analyzing and evaluating these relations, using the Heckman equation for income and count data models for the age-grade gap. The results on income show that physical activity practice (PAP) not only can lead to positive impacts on the individuals' health stock, but can also affect their income earnings, becoming a socioeconomic issue that affects the quality of life of individuals through many means. Thus, we find that the practice of any kind of physical activity out of the work hours increases the individual's income earnings by 10.2% (IV-Probit IV-GMM) to 12.3% (BP IV-GMM), on average. This result was obtained while controlling the estimations for selection bias and for the presence of endogeneity between PAP and the individual's occupational situation and income. These PAP impacts vary across sexes and throughout age groups, so that PAP displays greater impacts on men's income earnings, and an inverted U-shaped behavior across age groups, similar to the very impacts of age and experience on income (MINCER, 1958; SCHULTZ, 1961). As for the age-grade gap, the effects of PAP on the academic performance of adolescents were quite different for boys and girls: while, for the girls, the fact of participating on out-of-school hours PAP reduces their age-grade gap by 24.8%, or 0.281 years (around 3.4 months), on average, for the boys, the PAP was, in most cases, statistically zero in terms of affecting AGG. Even more, considering exclusively those boys regularly attending school, the NBRM MEs point out to a positive effect of participating on PAP on their age-grade gap, so that it negatively affects the academic performance of these boys.

Keywords: Health and Inequality; Welfare; Wage Differentials; Sample Selection; Instrumental Variables.

List of Figures

List of Tables

List of Abbreviations and Acronyms

2SLS	Two Stage Least Squares
AGG	Age-Grade Gap
GMM	Generalized Method of Moments
IBGE	*Instituto Brasileiro de Geografia e Estatística*
IV	Instrumental Variables
ML	Maximum likelihood
OLS	Ordinary Least Squares
PAP	Physical Activity Practice
PNAD	*Pesquisa Nacional de Amostra por Domicílio*
PSU	Primary Sampling Unit
BP	Bivariate Probit
EJA	*Educação de Jovens e Adultos*
LDB	*Lei de Diretrizes e Bases da Educação Nacional*
LR	Linear Regression
ME	Marginal Effects
MEC	Ministry of Education
NBRM	Negative Binomial Regression Model
PE	Physical Education
PNS	*Pesquisa Nacional de Saúde*
POF	*Pesquisa de Orçamentos Familiares*
PRM	Poisson Regression Model
PSM	Propensity Score Matching
VIGITEL	Sistema de Vigilância de Fatores de Risco e Proteção para Doenças Crônicas por Inquérito Telefônico

WHO	World Health Organization
ZINB	Zero-Inflated Negative Binomial
ZIP	Zero-Inflated Poisson

Contents

Chapter 1

Introduction

Insufficient physical activity is the fourth of the 10 leading risk factors for global mortality (WHO, 2018b). Sedentary lifestyle is still a serious problem in Brazil - and worldwide. Data from the 2013 *Pesquisa Nacional de Saúde* (PNS) (IBGE, 2014) show that, in Brazil, one in two adults does not practice the minimum level of physical activity recommended by the World Health Organization (WHO) . In general, 46% of the interviewed population was insufficiently active, with this proportion being higher for women, and increasing with age group. Still, the WHO global statistics (WHO, 2018a) for 2010 indicate that 81% of teenagers (11 to 17 years) and 23% of adults (over 18 years) are insufficiently active.

The WHO defines physical activity as "*any body movement produced by the musculoskeletal system that requires energy expenditure.*" (WHO, 2018a, p. 14). Therefore, it can take different forms, from walks, races, collective and individual sports, and recreational activities (such as dances and yoga), to part of work (paid or not) in case of manual labor and domestic activities. For an adult to be considered active he/she must practice a minimum of 150 minutes of moderate physical activity per week, or something equivalent. The level of physical activity adequate for teenagers is 60 daily minutes, in an intensity that varies from moderate to high (WHO, 2018b).

As mentioned previously, failure in accomplishing these physical activities practice (PAP) recommendations - that is, active insufficiency - increases not only the risk of mortality, but also the risk of occurrence of cardiorespiratory and metabolic diseases (WHO, 2018a). Figures 1.1 and 1.2 present the worldwide prevalence of insufficient physical activity practice among adults and school going adolescents. In

Figure 1.1, it is noticed that, in relation to adults, Brazil is in an intermediate situation, where the prevalence of physical inactivity is not too high (between 20% and 29% of adults are insufficiently active), as it is in some African countries and in North America.

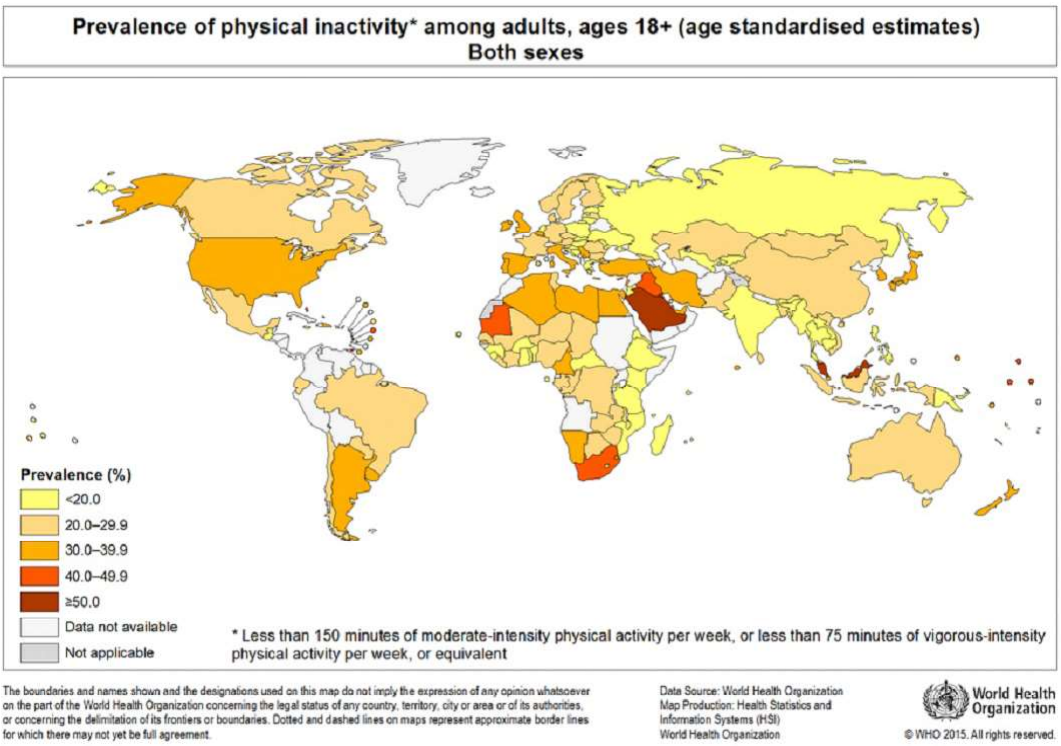

Figure 1.1: Prevalence (rates) of PAP insufficiency among adults.

WORLD, Health Organization (WHO), 2015.

On the other hand, regarding the adolescents, the reality is somewhat different: Figure 1.2 shows that Brazil is in one of the worst situations found worldwide, in which about 90% of the adolescents are insufficiently active.

Regular PAP (according to WHO recommended levels) reduces the risk of heart disease, diabetes, breast and colon cancer. It also reduces the risk of heart attacks, hypertension and depression, as well as acting as an important component in the individual energy balance, in order to control weight. For younger people - teenagers and children -, physical exercise improves cardiorespiratory fitness, muscle strength and endurance, and bone health. Other benefits include reducing body fat, reducing the risk of cardiorespiratory and metabolic diseases, and reducing the symptoms of anxiety and depression.

Recently, Chekroud et al. (2018), while analysing data from 1.2 million adults on the USA, found that individuals who had exercised had 43.2% fewer days of poor mental

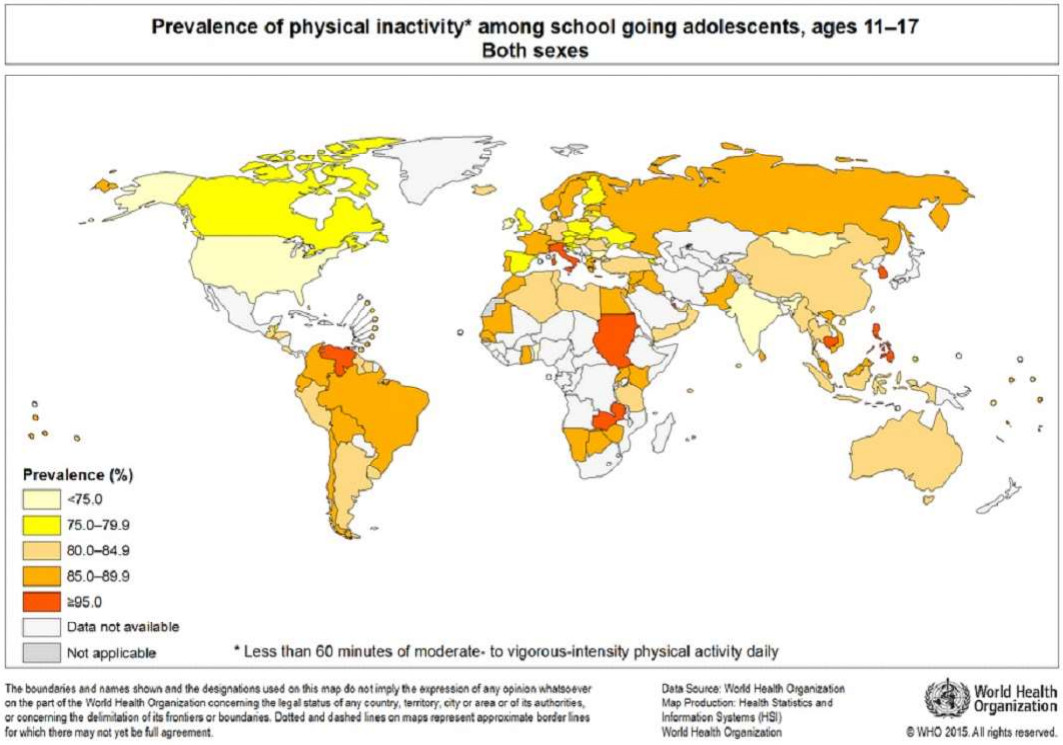

Figure 1.2: Prevalence (rates) of PAP insufficiency among school going adolescents. WORLD, Health Organization (WHO), 2015.

health in the previous month than individuals who did not exercise. Thus, individuals who exercised were "happier" than those who did not. They also found that physically active people feel just as good as those who don't do sports, but who earn around $\$25,000$ more a year.

When relating PAP to the countries' income levels, the WHO points out that:

> "The prevalence of insufficient physical activity rose according to the level of income. High income countries had more than double the prevalence compared to low income countries for both men and women, with 32% of men and 42% of women being insufficiently physically active in high income countries as compared to 13% of men and 19% of women in low income countries. In wealthier countries, the transition towards more sedentary occupations, recreation and motorised transport could explain the higher levels of inactivity, while in lower-income countries, more activity is undertaken at work and for transport. (...) The increased automation of life and work in higher income countries creates more opportunities for insufficient physical activity." (WHO, 2018b).

In addition to the consequences for physical and mental health, the PAP can also impact the individuals' well-being in an economic aspect. There is a range of

theories about the relationship between the PAP and the individuals' socioeconomic welfare. Starting from the notion that the individuals' physical and mental health affects the stock of human capital (BECKER, 2007), once the latter pervades issues of education and life expectancy, then the PAP would positively affect the individuals' productivity - and hence, their income earnings. With regards to PAP effects on adolescents, Fejgin (1994), anchored on the social importance of sports practice in the United States high schools, the author contrasts two theories regarding the impacts of such physical exercise practice on academic performance: the development theory (REHBERG, 1969; SCHAFER; ARMER, 1968; REES; HOWELL; MIRACLE, 1990), which highlights the positive impacts of PAP, due to its characteristics of socialization, and character and discipline building on students; and the zero-sum theory (COLEMAN, 1961; SPADY, 1970), which main ideas point to a trade-off between sports practice and the academic focus of individuals, regarding their time allocation and energy expenditure.

However, in Brazil, it is noticed that there are not many studies focused on the socioeconomic analysis of PAP. Although it has not been widely discussed as such, PAP is a socioeconomic problem whose characteristics and consequences must be studied so that:

> "Failure to recognize and invest in physical activity as a priority within NCD (Non Communicable Diseases)[1] prevention and treatment represents a missed opportunity. Ongoing inaction will see the costs of physical inactivity continue to rise, contributing to further negative impact on health systems, the environment, economic development, community well-being and quality of life for all." (WHO, 2018a, p. 16).

Therefore, a discussion regarding the PAP as an economic problem becomes very pertinent not only for the individual decision to practice physical activities, but also for a possible formulation of public policies, in order to encourage PAP and sports practice as a way to improve the individuals' well being and income earning. Still, the publication of the WHO report entitled *Global Action on Physical Activity 2018-2030: more active people for a healthier world* in 2018 also emphasizes the importance of studying these different aspects of PAP, highlighting its impacts on the environment and for income inequality reduction (WHO, 2018a).

[1]Such as: diabetes, heart attacks, cardiorespiratory problems and cancer (breast and colon cancer, mainly).

Thus, the following questions are raised: how does the practice of physical activities affect individual incomes? And what are its impacts on the adolescents' academic performance, that is, what are its impacts on their age-grade gap? In this sense, the objective of this research is to analyze the PAP as a socioeconomic problem, relating it to the well-being of individuals in Brazil. On one hand, for adults (those over 18 but under 65 years old), we intend to analyze the relationship between their income and the PAP, since greater well-being is related to higher productivity, which leads to higher income earnings (MINCER, 1958; SCHULTZ, 1961; BECKER, 1962; MUSHKIN, 1962). On the other hand, for school-aged individuals (teenagers, from 15 to 17 years old), since PAP, besides being included in the school hours, also impacts the physical and mental welfare of children and adolescents, we will analyze the age-grade gap of these individuals.

For that, we will use the statistical-econometric methodology as a way of analyzing and evaluating these relations, using the Heckman equation for income and count data models for the age-grade gap. We will also use the microdata from *Pesquisa Nacional por Amostra de Domicílio* (PNAD) of 2015, which includes a supplementary questionnaire referring to the practice of physical activities.

Whereas the hypothesis that the practice of physical activities has a positive relationship with the individuals' health stock (MINCER, 1958; MUSHKIN, 1962; GROSSMAN, 1972; BECKER, 2007), so as to indicate greater well-being and quality of life, it is expected, *a priori*, that there be positive impacts of PAP on the income earning of individuals, but negative impacts on the age-grade gap.

Finally, for the conduct of this research, this work is subdivided into four other chapters, besides this introduction: next, on chapter two, we discuss the determinants of PAP and its impacts on individuals' health stock through a descriptive statistical analysis; chapter three discourses over the socioeconomic impacts of PAP for adults, that is, the impacts of PAP on individuals' income earnings, while chapter four deals with these same impacts, but regarding teenagers, that is, analyzing the age-grade gap; lastly, we present some concluding remarks and the next steps that can be taken to complement this socioeconomic PAP analysis.

Chapter 2

Physical Activity Practice: A statistical analysis

2.1 Physical Activity Practice, Health and Socioeconomic Characteristics

Worldwide, nearly 23% of adults (over 18 years of age and younger than 65 years) are physically inactive (WHO, 2018b). In Brazil, data from 2013 PNS show that this figure is of 46% (IBGE, 2014). In spite of its benefits to mental and physical health, a decrease on PAP has been observed on several countries (TROIANO; MACERA; BALLARD-BARBASH, 2001). Since physical activity can be defined as "*any body movement produced by the musculoskeletal system that requires energy expenditure.*" (WHO, 2018a, p. 14), it can take many different forms, from walks, races, collective and individual sports, and recreational activities (such as dances and yoga), to part of work (paid or not) in case of manual labor and domestic activities. By taking different forms, it can have different impacts o both physical and mental health. In this research, we will focus on the out of the work hours PAP.

Whereas PAP can differently impact individuals' physical and mental health, it can also cause effects on socioeconomic characteristics, such as occupational situation, productivity and income earning (BECKER, 1962; GROSSMAN, 1972; MUSHKIN, 1962; BECKER, 2007). We, then, intend to analyze PAP as an economic problem, in the sense of testing for socioeconomic and behavioral differences between those individuals who

practice some kind of physical activity out of the working hours, and those who do not. That is, the main goal of this analysis is to test for statistical differences regarding these preferential, behavioral and socioeconomic variables between Brazilian physical activities practitioners and those who are not, in order to find preliminary evidence of characteristics that can affect and/or be affected by the PAP.

In general, regular PAP reduces the risk of heart disease, diabetes, breast and colon cancer. It also reduces the risk of heart attacks, hypertension, anxiety and depression, as well as acting as an important component in the individual energy balance, in order to control weight (WHO, 2018b). Wessel et al. (2004) and Ekelund et al. (2012), while analyzing samples for different groups and age groups, observed that greater physical fitness (calculated as an index) and greater time spent in physical activities have a negative and significant relationship with cardiometabolic risk factors and with the occurrence risk of cardiovascular diseases. Yet Ross and Hayes (1988) relate a psychological well-being index to the practice of physical exercises (also in an index form), controlling the results for sociodemographic and sociopsychological characteristics, variables that may be associated with both exercise and psychological well-being, including age, marital status, education, income, sex, religion, and instrumentalism[1]. Their results, for a random sample of adult individuals from Illinois in 1984, indicate that exercise is associated with decreased symptoms of depression, anxiety, and discomfort (discomfort feeling, difficulty on sleeping) in general population.

When it comes to socioeconomic characteristics, Mushkin (1962) and Grossman (1972), by analyzing health as a capital that can enhance both productivity and human capital, state that higher health stocks lead to higher income earnings. Since PAP increases individuals' health stocks, then PAP can lead to higher wages and income earnings. Still, Grossman (1972) highlights that the opposite effect can also exist: higher health stocks can be achieved by those with higher incomes. Thus, individuals who practice physical activities can do so because they earn higher incomes, which makes them able to pay for that benefit. Moreover, they also have more time for the practice of these activities (unlike the poorer ones, who usually work longer hours for a lower salary).

[1]Instrumentalism is an index created by the authors that measures the individuals' agreement with some statements regarding hard work, luck and success.

Regarding the behavioral impacts of PAP, the theoretical discussion relies on two different aspects related to sports practice: (i) on one hand, the development theory, introduced in Rehberg (1969) and Schafer and Armer (1968), and also discussed by Hanks and Eckland (1976) and Rees, Howell and Miracle (1990), points out that sports practice, once it develops in individuals characteristics such as teamwork, persistence, patience, leadership, commitment and organization, would lead to better academic - and also, socioeconomic - performance/accomplishments, alongside with being a central factor in adolescents' and young adults' character building and socialization. (ii) On the other hand, Coleman (1961), noting that sports practice may be detrimental to academic goals and success, discusses the zero-sum theory, whereas the idea that there is a finite amount of time and energy to be spent by individuals, then there would be a trade-off between the successful results in the academic field and sports practice.

King et al. (1992) and Weinberg and Gould (2016) study the reasons for adherence and withdrawal from PAP and point as reasons to exercise: body weight control, increased self-esteem, decreased risk of hypertension, decreased stress and depression, satisfaction (related to having fun while exercising) and socialization. King et al. (1992) also highlights some determinant factors of adult PAP, such as personal characteristics - sex, age, color/race, occupation, smoking habits, income level, medical condition, knowledge and attitudes or beliefs - behavioral attributes, skills, and environment conditions. However, Cousins (2014) states that lack of time to get acquainted with friends, discouragement of family and friends, not having the necessary equipment and accessories, the sense of exclusion in the exercise practice, and even previous experiences or lack of company would be detrimental and could even cause PAP withdraw.

Anchored on these previous framework, Santos and Knijnik (2009) find that the initial reasons for PAP among 40-60 years old adults are: medical order, leisure and quality of life, aesthetics and health (or physical conditioning). In order to do so, they conducted their own data collection and questionnaire application.

While trying to understand the determinants of PAP among adults, Florindo et al. (2009) used data from VIGITEL – a Telephone-based Surveillance of Risk and Protective Factors for Chronic Diseases collected in 2006- , and through average differences they found that the prevalence of inactive people at leisure was higher than 60%; men were

more active than women in leisure, work and displacement, while women were more active in heavy housekeeping and chores. People over 12 years of schooling were more active at leisure and less active at work, displacement and chores. Furthermore, the existence of a place near the residence for PAP or recreation has favored this practice for both sexes, of all age and schooling groups.

Focusing on gender differences as discussed by Vaitsman (1994) when stating that PAP gender differences are developed as part of certain social systems, within which differences and hierarchies between certain social categories are built, Salles-Costa et al. (2003) similarly analyze leisure PAP through average differences between individuals that practice physical activity and those who do not, using 1999 Pro-Health Study data, from Rio de Janeiro State University. Their results showed that women stated they practiced significantly less physical activities at leisure time than men, which may be due to women's double/triple working hours, which contributes to the limitation of their leisure time (AQUINO, 1996). Thus, the time spent on domestic chores and the number of dependent kids on the family, may affect the availability of PAP and leisure time of individuals and, even more of women. Still, these results could also indicate how income earning and time availability could importantly affect the PAP individual decision.

Regarding the displacement PAP and using the same average differences method for data from the 2008 PNAD[2], Knuth et al. (2011) stated that the physical activity on the commute to work was most common among men, young people with low schooling and in the most impoverished regions of the Country (that is, the North and Northeast). In that sense, this type of physical activity, in the Brazilian reality, was not practiced as an option to promote quality of life, but because of a need.

Furthermore, Sallis, Bauman and Pratt (1998) discuss the importance of physical environmental factors as essential elements on influencing individuals' behaviour regarding PAP. That is, the existence of physical environmental factors can influence the amount and the types of physical activities practiced by individuals. As stated by Powell, Slater and Chaloupka (2004) "*Positive associations between physical environments and levels of physical activities have been found for children, adolescents and adults.*" (POWELL; SLATER; CHALOUPKA, 2004, p. 137).

[2]The 2008 PNAD supplementary questionnaire was, too, about health and PAP.

In the sense of investigating the association between these physical environmental factors and socioeconomic status, ethnicity and race, Powell, Slater and Chaloupka (2004) run several Qualitative Response regressions for the availability of those first ones, while controlling for race, population density, region and socioeconomic status variables. Using a combination of the ImpacTeen Project and the Census Bureau data[3], they find that higher median household income and lower poverty rates are associated with increasing levels of available physical activity-related settings. Also, higher proportions of racial minorities (such as African-Americans and Hispanics), who are at most risk of being inactive, are associated with the fewest community-level physical activity-related settings.

Through a structural equation modeling (SEM) approach, Meyer, Castro-Schilo and Aguilar-Gaxiola (2014) simultaneously analyze self-rated health and mental health, with regards to socioeconomic status, PAP and neighborhood safety, controlling for age and ethnicity. For the 2009 California Health Interview Survey (CHIS), the largest statewide, population-based health survey in the USA, the estimation results showed that greater socioeconomic status and PAP were associated with better self-rated and mental health. Yet physical activity was positively associated with socioeconomic status, but negatively related to age and neighbourhood safety fears, which may indicate that more adequate neighbourhood conditions could lead to increased PAP.

Considering the PAP differences among socioeconomic and sociodemographic characteristics, both Powell, Slater and Chaloupka (2004) and Meyer, Castro-Schilo and Aguilar-Gaxiola (2014) highlight that policy interventions regarding PAP and public health should address disparities considering both socioeconomic status and race/ethnicity. Ergo:

> "(...) targeted interventions to improve access to neighbourhood physical activity-related opportunities may help to reduce the persistent disparities in health related to SES (socioeconomic status). Broadly speaking, in addition to income redistribution policies, interventions to improve the health of low-income individuals should include proactive urban planning policies to reduce barriers related to physical activity." (POWELL; SLATER; CHALOUPKA, 2004, p. 143).

[3]The authors end up with an USA nationwide sample of 8th, 10th and 12th grade students.

Thus, taking into consideration all these framework and previous results here presented, we intend to analyze average differences between physical activities practitioners and non-practitioners, regarding these socioeconomic, sociodemographic and behavioural characteristics, such as sex, color, age, schooling, marital status, occupational situation, income earning, time availability and neighborhood physical activity-related settings. Using 2015 PNAD microdata, we expect these differences to be statistically significant, so that PAP individuals should present increased averages for income earnings and schooling, but decreased averages for age, hours spent on domestic chores and number of dependent kids (KING et al., 1992; WEINBERG; GOULD, 2016; COUSINS, 2014; AQUINO, 1996). Still, we expect an increased average of men, individuals with spouses and individuals living in areas with more PAP-related settings among those that practice any kind of physical activities (POWELL; SLATER; CHALOUPKA, 2004; MEYER; CASTRO-SCHILO; AGUILAR-GAXIOLA, 2014).

With regards to the individuals' occupational situation, two possible results are expected: (i) if the average of occupied individuals is higher among those who are physical activities practitioners, then, on one hand, the income earn and employment security could have positive impacts on the individual decision of practicing physical activities; on the other hand, the PAP could enhance the individuals' health stock, which would have positive impacts on the individual's decision of participating on the labor market (GROSSMAN, 1972; MUSHKIN, 1962); still, (ii) if the average of occupied individuals is higher among those who are not PAP practitioners, then the effect of time unavailability could be observed, once the individual would spend most of their time working and/or commuting to work (COLEMAN, 1961).

Finally, it is expected a higher average of individuals with higher health stocks among those who are PAP practitioners. Also, when it comes to preferential and behavioural characteristics, we expect individuals that care about their health, their socialization and also about public investment on sports and physical activities development to be mostly inclined to be physical activities practitioners.

2.2 PAP Descriptive Statistics Analysis

As already discussed on this Chapter, PAP can impact both health and socioeconomic characteristics of individuals, in terms of their well-being. In order to preliminarily investigate if there is evidence of these impacts for Brazilian individuals, a descriptive statistic analysis will be performed. We will use data from 2015 PNAD that besides being the most recent Brazilian sample survey to which we have access, it also holds a supplementary questionnaire with regards to the practice of physical activities, which suits the main objectives of this analysis. That is, the main objective of this section is to describe the out-of-work hours PAP among adults in Brazil, as presented on the PNAD sample, for the purpose of finding preliminary evidence of characteristics/variables that affect and can be affected by the practice of physical activities.

We start by treating the data for keeping only the observations that concern to our object of analysis. Since we are testing the socioeconomic impacts of the PAP, first we kept on our sample only the individuals that were selected and answered the supplementary questionnaire regarding the PAP. As we are interested on analysing PAP impacts on adults' well-being - which inevitably relies on an income analysis -, we selected the observations of individuals that were between 18 and 65 years old, the most common ages of participation on the labor market. Ages above or below these limits could mislead the estimation results, since they correspond to phenomena that are not the object of study of this research, e.g. child labor (OLIVEIRA; JUSTUS, 2017).

According to Vaz and Hoffmann (2007) suggestions, we also dropped from our sample individuals with zero income and undeclared or greater than one hundred thousand *reais* income[4], as well as those individuals who were classified as "undetermined" on the *"years of study"* category (whose information on schooling was not declared or did not allow for the IBGE *years of study* classification).

After cleaning the dataset, of our 55,120 remaining observations sample, 61.4% of the individuals did not practice any kind of sports or physical activity out of their working hours, while 38.6% did - which corresponds to 21,255 observations of the sample. These types of sports and physical activity are presented on Table 2.1, along with their practice proportion.

[4]Those observations were dropped because they were outliers.

Table 2.1: Types and Practice Proportion of Physical Activities and Sports

	Proportion
Walking	39.5%
Soccer	24.7%
Fitness Activities	14.8%
Bodybuilding	5.66%
Other	5.30%
Bike Riding	2.75%
Cycling	2.34%
Martial Arts	2.28%
Running	2.22%
Gymnastics	1.9%
Swimming	1.77%
Dancing	1.55%
Volleyball	1.22%
Athletics	1.13%
Water Sports	0.58%
Skate	0.23%
Basketball	0.22%
Racquet Sports	0.21%
Sports with Animals	0.14%
Adventure/Nature	0.11%
Handball	0.098%
Parasports	0.01%

Source: Elaborate by the author, from PNAD dataset.

From Table 2.1, we can observe that the most practiced physical activity is walking, corresponding to nearly 40% of the total of individuals that stated they practiced some type of sport or physical activity, followed by soccer (24.7%) and fitness activities (nearly 15%). Walking is the easiest, most simple and with less requirements physical activity listed. Soccer is the most popular sport on the Country, and fitness activities, such as aerobic exercises classes and weight training, are also a very common type of energy spending currently. Still, it is important to note that the sum of these percentages displayed on Table 2.1 will exceed 100% due to doubled responses, in cases which the individual practiced more than one sport or physical activity.

With the PAP already discriminated within our sample, we can now turn to the analysis of the characteristics that may affect and also may be affected by the practice of physical activities. Based on the control variables used by Salles-Costa et al. (2003), Florindo et al. (2009), and Sallis, Bauman and Pratt (1998) when analyzing the prevalence of physical inactivity and factors associated with PAP, respectively, we select

and construct the following variables, in order to test for differences - with regards to the characteristics represented by these variables - between those individuals who did and did not practice any type of physical activity. The variables are subdivided in three categories, indicating if they should affect or be affected by the PAP, or if they are endogenous, according to what King et al. (1992) and Weinberg and Gould (2016) discuss. These variables are displayed and expatiated on Table 2.2 below:

Table 2.2: PAP Related Variables: Construction and Description

Variable	**Description**
Endogenous:	
p_i	A binary variable, which assumes the value 1 if the individual is occupied, and 0 in case it is unoccupied or inactive.
y_i	The income variable, represented by the per hour wage earned by each individual on its main work.
Affect PAP:	
sex	A binary variable for the individual's sex, which assumes the value 1 for women.
age	A discrete variable representing the individual's age.
young adult[5]	A binary variable that indicates if the individual fits into the young adults age category, that is, if their age is comprehended between 18 and 35 years.
middle age adult	A binary variable that indicates if the individual fits into the middle-aged adults category, that is, if their age is comprehended between 36 and 55 years.
old adult	A binary variable that indicates if the individual fits into the old adults age category, that is, if they are older than 56 years.
color/race	Four binary variables to distinguish the color/race declared by the individual: white (base category), indigenous, yellow, brown and black.
householder	A binary variable for the individual's condition in the family, being 1 for householder, and 0 for children/spouses/ others.

[5]Age categories were constructed based on Petry (2002).

spouse	A binary variable that indicates whether the individual does have a spouse (1) or doesn't (0).
chores hours	A continuous variable, representing the number of weekly hours spent on domestic work, such as chores.
working hours	A continuous variable, representing the number of weekly hours worked by the individual on their main job.
working_h categories	A categorical variable, representing five categories of weekly hours worked by the individual on all of their jobs, both main and secondary. The categories are: (1) up to 14h; (2) from 15h to 39h; (3) from 40h to 44h; (4) from 45h to 48h; and (5) more than 48h.
n_jobs	A categorical variable, indicating the number of jobs of the individual (both main and secondary jobs) on the week of reference[6], ranging from one to three, being these categories: one, two, or three or more jobs.
dependent kids	A variable indicating the individual's number of dependent kids younger than 14 years old.
public_neigh_space	A continuous variable representing the average of positive answers to the existence of public areas dedicated to sports or physical activity practice, by PSU (primary sampling unit). Since it is a categorical answer (*yes* or *no*), this variable is comprehended between 0 and 1, and gives an idea of how the neighborhoods are propitious/friendly to PAP .
metropolitan	A binary variable indicating if the individual's household is located in a metropolitan region (1) or if it isn't (0).
rural	A binary variable indicating whether the individual's household is located in a rural (1) or urban (0) area.
I_public_pap	Indicates if the individual thinks there should be public investment on sports or physical activity development on his/her neighborhood.

health	Indicates if the individual has chosen to practice/not to practice sports or physical activities because of health or well being reasons. In other words, this is a binary variable that seeks to grasp if the individual cares about his/her health.
Affected by PAP:	
school	Schooling, a variable indicating the completed years of study of the individual.
health stock	A binary proxy of the individuals' health stock, that indicates if they have decided to practice or to stop practicing physical activities or sports due to medical indication or health problems. Thus, if this variable has a positive answer (1), then the individual should have a lower health stock.

Source: Elaborate by the author.

Thus, we expect these variables to comprehend both socioeconomic and behavioral characteristics that can affect the individuals' decision to practice physical activities, but that can also be affected by the mental and physical health impacts of the PAP. For the purpose of comparing and analyzing the differences in the scope of these variables between individuals that practice physical activities and those who don't, two-sided t tests of mean differences are performed[7]. Thereby, we can check if these differences are statistically significant. Both groups means and standard deviations are presented on Table 2.3, along with t test statistics and significance results.

Almost all tested variables showed statistically significant average differences between the physical activities practitioners and not practitioners, except the fact of being indigenous and the categories of weekly worked hours.As expected, and similarly to what Salles-Costa et al. (2003) and Florindo et al. (2009) found, according to our t test results, the average of women (men) practicing physical activities is smaller (larger) than the average of women (men) that don't. Also as expected, a higher age average is found among those individuals that do not practice any kind of physical activities or sports. As for the age groups, a higher average of young adults is observed for those who are

[7]We also present on Appendix A, both right-sided and left-sided t tests p-values, for reasons of results confirmation.

physical activities practitioners, while the average is smaller for both middle-aged and old adults within this PAP group. These preliminary results make sense, since higher ages correspond to lower health stocks (MUSHKIN, 1962; MINCER, 1958).

Table 2.3: Descriptive Statistics and t Test Results.

	(1) NON PAP		**(2) PAP**			
	Mean	**SD**	**Mean**	**SD**	**Mean Diff**	***t* Statistic**
Endogenous:						
p_i	0.649	0.477	0.708	0.454	-0.058***	(-14.45)
y_i	58.906	260.44	91.493	536.36	-32.59***	(-6.92)
Affect PAP:						
sex	0.559	0.496	0.481	0.499	0.078***	(17.91)
age	40.763	12.975	38.184	12.915	2.579***	(22.78)
young adult	0.389	0.487	0.479	0.499	-0.090***	(-20.78)
middle age adult	0.442	0.496	0.391	0.488	0.050***	(11.76)
old adult	0.168	0.374	0.128	0.334	0.039***	(12.86)
indigenous	0.003	0.060	0.004	0.067	-0.0008	(-1.57)
yellow	0.003	0.055	0.005	0.074	-0.002***	(-4.14)
brown	0.495	0.499	0.437	0.496	0.057***	(13.19)
black	0.108	0.310	0.099	0.299	0.008***	(3.09)
householder	0.545	0.498	0.537	0.498	0.0075*	(1.72)
spouse	0.594	0.491	0.556	0.497	0.037***	(8.72)
chores hours	15.95	15.69	14.39	14.59	1.563***	(11.88)
working hours	25.52	21.21	27.66	20.44	-2.135***	(-11.76)
working_h categories	3.075	0.996	3.083	1.001	-0.008	(-0.739)
n_jobs	1.029	0.179	1.054	0.244	-0.025***	(-10.69)
dependent kids	0.771	1.005	0.628	0.895	0.142***	(17.33)
public_neigh_space	0.131	0.041	0.137	0.043	-0.005***	(-14.96)
metropolitan	0.372	0.483	0.392	0.488	-0.020***	(-4.88)
rural	0.134	0.341	0.088	0.284	0.046***	(17.10)
I_public_pap	0.688	0.463	0.845	0.361	-0.157***	(-44.42)
health	0.157	0.364	0.722	0.447	-0.565***	(-154.6)
Affected by PAP:						
school	9.036	4.379	11.155	3.941	-2.119***	(-58.83)
health stock	0.157	0.364	0.151	0.358	0.006*	(1.82)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The number of observations N is $33,865$ for the individuals that do not practice physical activity and $21,255$ for the practitioners, except for the income variable, y_i, in which the sample size is $22,006$ and $15,057$ observations for the physical activity non practitioners and practitioners, respectively, since, for this variable, there only are observations for the individuals who are occupied.

Source: Elaborate by the author, from PNAD dataset.

In terms of health stock, a statistically greater averaged can be observed for this variable on the NON PAP group. This result goes according to what we expected, since the successes on this categorical variable represent a loss of health stock. Thus, we find preliminary evidence that corroborate the positive impacts of PAP on the health and

well being of individuals (WHO, 2018b; ROSS; HAYES, 1988; WESSEL et al., 2004), enhancing their health stock (MUSHKIN, 1962; GROSSMAN, 1972).

Also as we expected, the average of people living in metropolitan areas is greater on the group of physical activities practitioners than on the group of non-practitioners. Still in relation to the environment in which the individual lives, a smaller average of individuals practicing physical activities out of their working hours is found in rural areas, and the average existence of public areas dedicated to PAP is greater within those individuals who practice physical activities, as discussed and observed by Meyer, Castro-Schilo and Aguilar-Gaxiola (2014), pointing out how the availability and the structure of a PAP-friendly environment can encourage PAP.

When it comes to individuals' socioeconomic characteristics, the average income and years of study of those who practice physical activity are greater than of those who don't, as argued by Becker (2007) when relating investment in human capital and health stock. Still, a larger average of occupied individuals is also found among those who are physical activities practitioners. As discussed on the previous section, these results may also show the importance of the individuals' income earning as a vehicle that allows for time and resources availability for the PAP (GROSSMAN, 1972). Adding up to this argument, the average number of jobs of the individuals on the PAP group is bigger than the average number of jobs of individuals who are not physical activities or sports practitioners.

Another characteristic that goes through the individuals' income earning, due to historical and social systematic construction[8] (VAITSMAN, 1994; POWELL; SLATER; CHALOUPKA, 2004) is their race: because of their job positions and their ability to earn income, there is a higher average of yellow individuals among those PAP practitioners, while this average is smaller for brown and black individuals[9].

Regarding the availability of time for the PAP (COLEMAN, 1961), similarly to the results showed by Salles-Costa et al. (2003), the average of time spent on house chores and the average of the individuals' dependent kids are greater for those who don't

[8]For a more broad and detailed discussions regarding how the social system construction may affect individuals' PAP, see Vaitsman (1994), Kington and Smith (1997), Crespo et al. (2000), and Powell, Slater and Chaloupka (2004).

[9]It is important to emphasize that, as stated by Kington and Smith (1997) and Powell, Slater and Chaloupka (2004), although the social construction and income earning may be very relevant when it comes to explaining these PAP differences among races and ethnicities, they can not be taken as the only and complete causes of those differences.

practice any kind of sports or physical activities, since time is limited. However, the weekly working hours exhibited higher average for the PAP group, demonstrating the importance of the income earning as a necessary resource for PAP, as opposed to Coleman's (1961) zero-sum theory. At the same time, taking into consideration the categorical variable for the individuals' worked hours on all their jobs, there was no statistical difference between the averages of the PAP and the NON PAP groups.

Still with respect to the availability of time, our results showed that the average of espoused individuals is greater for those who are not physical activities practitioners, as pointed out by Cousins (2014) when discussing that the lack of time to get acquainted with friends and family can be detrimental to the PAP. Similarly, the average of householders is greater within the NON PAP group, since this position usually comes with a higher burden of family and income provision responsibilities, also consuming more of these individuals' time and energy (COLEMAN, 1961).

Lastly, when it comes to the individuals' behavior and preferences regarding PAP, a greater average of individuals thinking that there should be public investment on PAP development on their neighborhood is observed for those who declared they practiced some kind of physical activity, than for those who did not. The t test results also showed that there is a higher average of individuals that care about their health and well-being within those who practice any kind of physical activities or sports.

Therefore, from the significant results of these average tests, we can state that there are statistical differences not only on socioeconomic, but also on behavioral and environmental characteristics of individuals that practice physical activities and those who do not. Thus, there might be impacts of these features on the individuals' decision of practicing physical activities, and/or impacts of the PAP on their socioeconomic status.[10]

2.3 Concluding Remarks

Once PAP already has proven impact on individuals' mental and physical health (ROSS; HAYES, 1988; WHO, 2018a; WESSEL et al., 2004; EKELUND et al., 2012), we intended to test, in this chapter, for preliminary evidence of PAP impacts also

[10]Although these average tests results showed that there are statistical differences with regards to individuals' socioeconomic and behavioural characteristics between the PAP and NON-PAP groups, it is important to note that these differences do not imply causal effects between PAP and these variables, or vice-versa. These causality effects will be tested on the following Chapter.

on socioeconomic and behavioural characteristics of individuals. Thus, in order to find out if there are evidences of socioeconomic, behavioral and preferential differences between individuals that are physical activities/sports practitioners and those who are not, we tested for average differences on variables that encompass those characteristics. For that, two-sided t tests of average differences were performed.

In short, our test results showed that the average of young, single men living on metropolitan areas, with public spaces dedicated to PAP practice and development is larger for those who practice physical activities than for those who do not, as suggested and discussed by King et al. (1992), Vaitsman (1994), and Meyer, Castro-Schilo and Aguilar-Gaxiola (2014).

We could also observe that, when it comes to time availability, the average of dependent kids and of hours spent on domestic chores are smaller for the individuals who are physical activities practitioners, once time is limited (COLEMAN, 1961; SALLES-COSTA et al., 2003). When it comes to preferential characteristics, the average of individuals that cared about their health and well being, and the average of individuals that thought there should be public investment on PAP development were higher within those who stated they practiced any kind of sports or physical activities. These results make sense, since individuals that practice physical activities should think of PAP as an important part of their lives.

With regards to the individuals' socioeconomic status, the average of occupied individuals, their years of study, and their income earnings were statistically higher on the PAP group than on the non-PAP group, as we expected *a priori*. In this sense, higher health stocks - obtained from PAP - could lead to enhanced productivity and human capital accumulation, which, in turn could result in higher income earnings (GROSSMAN, 1972; MUSHKIN, 1962; BECKER, 2007). Still, as stated by Grossman (1972), due to time availability and access to PAP resources, higher income earnings could, also, positively affect PAP. A similar relationship is valid for the individuals' occupational situation, which highlights the possibility of existence of endogeneity between PAP and the individuals' income earnings and labor market participation.

From these results, it is possible to observe that there are statistical differences not only on socioeconomic, but also on preferential and behavioural characteristics of PAP and non-PAP individuals, although we can not empirically infer over the causality

and the direction of the effects of these variables on one another. Therefore, in order to better understand the direction, the effects and the causality of these socioeconomic characteristics-PAP relationships, we proceed to the econometric approaches discussed on the following Chapters.

Lastly, for further research and also as a robustness and complimentary analysis to the one developed on this chapter, we suggest to perform a Propensity Score Matching (PSM) analysis, defining control and output variables, so that it is possible to evaluate PAP impacts on these socioeconomic variables.

Chapter 3

Physical Activity Practice as a Socioeconomic Factor: An analysis over income.

3.1 Introduction

Insufficient physical activity is the fourth of the 10 leading risk factors for global mortality (WHO, 2018b). Sedentary lifestyle is still a serious problem in Brazil - and worldwide. Globally in 2016, 23% of men and 32% of women aged over 18 years were insufficiently physically active (WHO, 2018b). According to the World Health Organization (WHO), these figures did not significantly improve on the past 15 years: on average, 28.5% of the adult population were insufficiently active in 2001, while in 2016 the prevalence of physical inactivity decreased to 27.5%.

In Brazil, the scenario is not very different: data from the 2013 *Pesquisa Nacional de Saúde* (PNS) (IBGE, 2014) show that one in two adults does not practice the minimum level of physical activity recommended by the WHO. In general, 46% of the interviewed population was insufficiently active, with this proportion being higher for women, and increasing with age group. WHO 2016 data corroborates these statistics: the NCDs[1] country profile report showed that 47% of Brazilian adult population were physically inactive (WHO et al., 2018).

[1]Non-Communicable Deseases.

For an adult to be considered active he/she must practice a minimum of 150 minutes of moderate physical activity per week, or something equivalent (WHO, 2018b). Regular physical activities practice (according to WHO recommended levels) reduces the risk of heart disease, diabetes, breast and colon cancer. It also reduces the risk of heart attacks, hypertension and depression, as well as acting as an important component in the individual energy balance, in order to control weight. Even more so, WHO et al. (2018) estimates that nearly 74% of 2016 total deaths were due to NCDs, which could have been prevented/reduced with the practice of physical activities.

Alongside with the consequences for physical and mental health, the physical activities practice (PAP) can also impact the individuals' well-being in an economic aspect. In this sense, the objective of this research is to analyze the PAP as a socioeconomic problem, relating it to the well-being of adult individuals (aged 18 to 65 years old) in Brazil, in terms of income earnings. Thus, it is intended to analyze the relationship between the individuals' income and the PAP, since greater well-being is related to higher productivity, which entails greater individual capacity to earn income. Anchored on Mincer (1958) and Mincer (1974), and Schultz (1961), Mushkin (1962), and Becker (2007) human and health capital theories, we expect PAP to have positive impacts on individuals' income, since it increases their health stock and their productivity.

Many researches have already been carried out when it comes to understanding the impacts of PAP on health and on diseases, such as NCDs. Wessel et al. (2004) and Ekelund et al. (2012), while analyzing samples for different groups and age groups, observed that greater physical fitness and greater time spent in physical activities have a negative and significant relationship with cardiometabolic risk factors. Ross and Hayes (1988) relate a psychological well-being index to the practice of physical exercises, finding that exercise is associated with decreased symptoms of depression, anxiety, and discomfort in the general Illinois population.

Following this line of analysing the effects of PAP on the well being of individuals, but from an economic aspect, Chekroud et al. (2018) found that individuals who had exercised had 43.2% fewer days of poor mental health in the previous month than individuals who did not exercise, on a dataset of 1.2 million adults on the USA. Thus, individuals who exercised were "happier" than those who did not. In addition,

they also found that physically active people feel just as good as those who don't do sports, but who earn around $\$25,000$ more a year.

However, when it comes to income, the effect of PAP is not well known, since it has not been studied this much. With regards to the latter, WHO points out that:

> "The prevalence of insufficient physical activity rose according to the level of income. High income countries had more than double the prevalence compared to low income countries for both men and women (...) In wealthier countries, the transition towards more sedentary occupations, recreation and motorised transport could explain the higher levels of inactivity, while in lower-income countries, more activity is undertaken at work and for transport. (...) The increased automation of life and work in higher income countries creates more opportunities for insufficient physical activity." (WHO, 2018b).

With that said, it is important to understand how the PAP affects the socioeconomic aspects of the well-being of individuals either than just its impacts on individuals' health. Failure to recognize these other PAP impacts and invest on incentives to reduce physical inactivity can increase its health and economic costs, "*(...)contributing to further negative impact on health systems, the environment, economic development, community well-being and quality of life for all.*" (WHO, 2018a, p. 16). In this sense, the productivity loss due to physical inactivity, according to the human capital approach, would be a loss of income potential not only for individuals but also for the whole society (OLIVEIRA; JUSTUS, 2017).

Therefore, this research intends to analyze PAP as an economic problem, - which can lead to socioeconomic impacts - relating individuals' PAP to their income earnings, through Heckman's sample selection method, similarly to what Oliveira and Justus (2017) and Teixeira (2016) performed in order to understand the impacts on income/wages of depression and obesity, respectively. We will use data from 2015 PNAD[2] that besides being the most recent Brazilian sample survey to which we have access, it also holds a supplementary questionnaire with regards to the practice of physical activities, which suits the main objectives of this analysis.

Hence, this chapter is divided in four other sections besides this introduction: section 3.2 explains the theoretical embasement behind our research problem and hypothesis, followed by section 3.3 in which we detail the estimation method and the

[2] *Pesquisa Nacional por Amostra de Domicílio.*

variables used; in section 3.4 the results are presented and discussed, which are concluded on section 3.5.

3.2 Literature Review

The practice of physical activities, regardless of their forms or motivations, always brings health benefits, when practiced in sufficient quantity and regularly (WHO, 2018a). Thus, while some physical activities are practiced by choice - which makes them pleasurable - others are practiced just because they are necessary or even mandatory, so that the latter may not generate the same benefits for mental health and social acquaintanceship than those first ones. WHO (2018a) mainly highlights the impact of PAP on reducing the incidence of cardiorespiratory diseases and heart attacks, as a support tool on individuals' weight control and energy balance, helping avoid obesity, besides having great importance as a form of reduction of depression and anxiety risks and symptoms.

On previous works, much has already been discussed about the impacts of PAP on physical and mental health. Wessel et al. (2004) and Ekelund et al. (2012), while analyzing samples for different groups and age groups, observed that greater physical fitness (calculated as an index) and greater time spent in physical activities have a negative and significant relationship with cardiometabolic risk factors and with the occurrence risk of cardiovascular diseases.

Regarding mental health, Ross and Hayes (1988) relate a psychological well-being index to the practice of physical exercises (also in an index form), controlling the results for sociodemographic and sociopsychological characteristics. Their results, for a random sample of adult individuals from Illinois in 1984, indicate that exercise is associated with decreased symptoms of depression, anxiety, and discomfort (discomfort feeling, difficulty on sleeping) in the general population. Similarly, taking as object of study the high school teenagers in the United States, the results of Babiss and Gangwisch (2009) indicate that:

> "As sports participation increases, the odds of suffering from depression decreases by 25% and the odds of having suicidal ideation decreases by 12%, after controlling for sex, age, race/ethnicity, public

> assistance, and physical limitations." (BABISS; GANGWISCH, 2009, p. 376).

In an attempt to quantify the benefits of PAP to mental health, Chekroud et al. (2018), while comparing the self-reported mental health burden between balanced samples of US adults who exercised in the previous month and those who did not, also accounting for self-selection of these adults into exercising[3], they found that, on average, individuals who had exercised had 43.2% fewer days of poor mental health than those who had not, on that referred previous month. All exercise types were associated with a lower mental health burden than not exercising at all, but larger associations were observed for popular team sports, as well as for exercises with durations of 45 minutes and frequencies of three to five times a week. It is interesting to note, however, that the individuals who exercised for more than 3 hours[4] were associated with worse mental health burden than not exercising at all. Still, they also found, on this research, that physically active people feel just as good as those who don't do sports, but who earn around $\$25,000$ more a year.

From these positive impacts of PAP on the health of individuals, as for treating it as a socioeconomic problem, we inevitably fall on health impacts over the individual's productivity and, as a consequence, on their ability to earn income - that is, on the Human Capital approach.

In the seminal idea regarding the factors that determine the individuals' ability to earn income and the theory of human capital, Schultz (1961) shows that there are investments made by individuals "in themselves" which will lead to higher wage returns in the future. In that sense, these kinds of expenditure - that are not pure consumption - that enhance individual skills and abilities can be seen as investment in "human capital", once they lead to future monetary returns. Thus, the remuneration of the labor factor would not be homogeneous among workers because of the difference on the quality of the capital they represent (BECKER, 1962).

Similarly, Mincer (1958) points, from a theoretical and empirical framework, that higher experience and schooling would increase the accumulation of human capital, which in turn would increase the productivity of individuals, so that the latter would be reflected in higher wages. Still in this discussion, Mincer (1958) emphasizes that:

[3]Through a propensity score estimation of engaging on a specific type of PAP/sport.
[4]That is, exercises duration of more than 3 hours.

> "The growth of experience and hence of productivity is reflected in increasing earnings with age, up to a point when biological decline begins to affect productivity adversely. The important difference among occupational groups is that, on the whole, increases in productivity with age are more pronounced, and declines are less pronounced, in jobs requiring greater amounts of training." (MINCER, 1958, p. 301).

From this last observation, Mushkin (1962) and Becker (2007) analyze health as part of human capital, being capable of affecting productivity and, consequently, the income of individuals. Mushkin (1962) points that health and education are not only consumption goods, which satisfy human wants, but also essential ingredients of human welfare. There are economic gains in preventing and curing sickness, which are more human labor, higher productive capacity and productive work time added (MUSHKIN, 1962), since greater health stocks are related to higher productivity and less time treating diseases (GROSSMAN, 1972). Therefore, the individual is more effective in society as a producer and as a consumer because of these investments in health.

While relating health, education and human capital, using the method of maximization of the utility of individuals, and considering their probabilities of survival and life expectancy, Becker (2007) states that: "*Higher survivorship at adult ages would induce greater investment in education because expected returns on education investment would be greater.*" (BECKER, 2007, p. 407). Thus, the longer the life expectancy, the greater the investments in education. That is, the accumulation of human capital is greater when treating health as part of this last variable. Consequently, the productivity and income of individuals are greater. In this sense, since PAP has positive impacts on individuals' health, then it should have positive impacts on their income earnings.

It is important to highlight, as pointed out by Oliveira and Justus (2017) and discussed by Becker (2007) and Grossman (1972), the existing possibility of endogeneity between the individuals' income and the PAP: on one hand, those who practice physical activities can do so because they have higher incomes, which makes them able to pay for that benefit. Moreover, they also have more time for the practice of these activities (unlike the poorer ones, who usually work longer hours for a lower salary). On the other hand, and explaining the relation that this work intends to analyze, the practice

of physical activities elevates the well-being of individuals, increasing their health stock[5], and leading to positive impacts on their income.

Following the idea of analyzing the costs of physical inactivity, - or PAP insufficiency - in the sense of the reduction on individuals' productivity, Oliveira and Justus (2017), while trying to find factors that reduce the individual well-being of women in Brazil, analyze depression as an economic problem that negatively affects productivity - and consequently, women's earnings - and their search for work stations, anchored on Schultz (1961), Mincer (1958), and Becker (2007) human and health capital theories. Whereas depression has negative impacts on health through the persistence of sadness and apathy, it would contribute to the fact that customary activities could not be performed with the same intensity as by a healthy person, which would reverberate on productivity loss (OLIVEIRA; JUSTUS, 2017). Using the Generalized Method of Moments (GMM) and instrumentalizing depression (with two binary variables for (i) the presence of back pain and (ii) if the woman had already suffered some type of physical or psychological violence), based on PNAD data from 2008, their results show a negative impact of depression not only on women's earnings but also on their participation in the labor market.

Analogously, and also as a way of measuring health impacts on the labor market, Teixeira (2016) seeks to examine whether there is a relationship between obesity and wages[6] through an empirical analysis of multivariate regression models and likelihood of employment, acknowledging the presence of endogeneity between obesity and wages. His results were heterogeneous between men and women: while, for women, obesity is associated with a penalty of 3.9% (IV[7]) to 9.1% (OLS[8]) in earnings, for men the impact is positive - overweighted men earn an average of 7.2% (IV) to 14.4% (OLS) higher income than non-obese individuals. This diversity in results may be led by the heterogeneity in occupation (of men and women), accentuating discriminatory differences in the labor market. In addition, Teixeira (2016) also emphasizes that the obesity impact on wages may be caused by health and productivity differences, but mainly by weight discrimination.

[5]Leading to greater Health and Human capital stocks (MUSHKIN, 1962; BECKER, 2007).
[6]For Brazil, using microdata from 2008-2009 *Pesquisa de Orçamentos Familiares* (POF).
[7]Instrumental Variables estimation.
[8]Ordinary Least Squares.

As for the social and individual costs of the lack of PAP, Ding et al. (2016) propose to analyze them through either country-specific and generalized estimates for the whole world, taking into account both the direct costs (of health system expenditures) and indirect costs derived from the productivity losses due to diseases caused by physical inactivity (treated as friction costs of labor replacement). Their results, presented on the WHO (2018) global PAP report, show that health system overheads, worldwide, represents $INT\$53.8$[9] billion, of which 57% is borne by the public sector and 14% is attributed to productivity loss. Thus, the productivity loss, according to the human capital approach, would be a loss on potential income earning not only for individuals, but also for the whole society (OLIVEIRA; JUSTUS, 2017).

3.3 Methodology

3.3.1 Econometric Procedures

In order to test the impacts of the PAP on individuals' income we will use the estimation method proposed by Heckman (1979) for income equations. Heckman (1979) argues that when estimating income equations only through Ordinary Least Squares (OLS) there is sample selection bias, since in a survey such as PNAD - sampled and stratified -, there is an answer for the individuals' wages only if they were working and earned income in the period of data collection. Therefore, one would not be considering the individual's reserve salary, or, in other words, its opportunity cost of being occupied/participating in the labor market, by estimating an income equation only for the occupied individuals.

Thus, the selection bias could arise for the following reasons: i) the self selection of the individuals or data that are investigated; or ii) the researchers' decisions about the sample selection (similar to the self selection) (ROCHA; SOUZA CAMPOS, 2011). In this case, as we intend to analyze the income of individuals, the selection bias occurs insofar as we can only analyze the income of those individuals who can receive it, that is, those who are employed. Therefore, a sample composed only of employed / income earning individuals - object of study in this analysis - would not be representative of the

[9]INT$ is an international currency, which was calculated by converting national currencies from the Purchasing Power Parity (PPP).

entire population, since it would not take into consideration the individual decision of working or not.

To solve this selection bias problem, Heckman (1979) suggests the estimation of two-stage income equations. In the first stage, a probit model of qualitative choice for the occupation of the individual is estimated (via Maximum likelihood - ML), in which the dependent variable assumes the value 1 if the individual is occupied, and 0 in case it is unoccupied or inactive. The model is estimated according to the following equation 3.1:

$$p_i = \beta_0 + \beta_k X_{ik} + \beta_{k+1} pap_i + \upsilon_i \tag{3.1}$$

Where:

p_i is the dependent variable, which assumes the value 1 if the individual is occupied, and 0 in case it is unoccupied or inactive. In terms of post-estimation, it represents the probability of the i-th individual being occupied;

β_0 is the intercept;

X_i is a vector of explanatory variables of individual characteristics that interfere in the individual's occupational situation, as suggested by Mincer (1996), Mincer (1974), and Mincer (1958), Oaxaca (1973), and Vaz and Hoffmann (2007). Those control variables will be listed and itemized in the following section;

pap_i is a dummy variable that indicates whether the individual practices any type of physical activity (1) or not (0);

and υ_i is the vector of random errors.

From the estimation of the parameters of this selection equation, it is possible to calculate λ_i, inverse of the Mills ratio, which represents the probability that an observation is selected in the sample (in this case, in the sample of the individuals that earn income). Values of λ_i can be predicted as described in 3.2:

$$\lambda_i = \frac{f(Z_i)}{1 - F(Z_i)} \tag{3.2}$$

Where $f(Z_i)$ represents the probability density function of Z_i, and $F(Z_i)$ is the (cumulative) probability distribution function of Z_i. Z_i follows a standardized normal distribution, defined from the parameters estimated in the probit model so that:

$$Z_i = -\frac{\beta_k X_{ik}}{(\sigma_k)^{\frac{1}{2}}} \tag{3.3}$$

Where σ_k represents the variance of the parameters β_k of the model, including the one referring to the physical activities practice. Thus, the inverse of the Mills ratio represents the probability of the i-th individual being occupied.

The inverse of the Mills ratio is calculated for each observation in the sample and is later added as an explanatory variable in the estimation of the income equation, characterizing the second stage of the estimation proposed by Heckman, correcting for the selection bias. It is important to note that the estimation of the income equation would be inconsistent if the variable λ_i was omitted. Thus, the Mincerian income equation, corrected for selection bias using the method proposed by Heckman (1979), can be estimated as shown in equation 3.4:

$$\ln y_i = \beta_0 + \beta_k X_{ik} + \beta_{k+1} pap_i + \gamma \lambda_i + \varepsilon_i \tag{3.4}$$

In which:

$\ln y_i$ is the dependent variable, the natural logarithm of the per hour wage earned by each individual, derived from its main work;

β_0 is the intercept;

X_i is a vector of explanatory variables that characterize the individual's ability to earn income, containing the same variables of the first stage qualitative model, excluding one of them for reasons of model identification (CAMERON; TRIVEDI, 2005). This excluded variable must affect the decision to work, but not necessarily the income of individuals. In this case, we chose to exclude two variables, *chores hours* and *dependent kids*, which represent the number of hours spent by the individual on house chores and the number of kids that depend on the individual, respectively. These variables will be better explained on the following section;

pap_i is a dummy variable that indicates whether the individual practices some type of physical activity (1) or not (0), the variable of interest for the purposes of this

research. *A priori*, it is expected for the parameter of this variable to be positive and statistically significant, indicating that the PAP has positive impacts on the individual's ability to earn income;

λ_i is the inverse of the Mills ratio, calculated from the model first-stage estimation;

and ε_i is the vector of random errors.

Despite correcting the inconsistency through the addition of the inverse of the Mills ratio, since we are estimating an income equation, it is very likely for our data to be heteroskedastic. Thus, in the presence of heteroskedasticity, the OLS estimations are still inefficient, causing the standard errors to be incorrectly estimated and the t statistics to be imprecise for running tests. In order to correct this inefficiency, White's covariance matrix must be used as the weighting factor to calculate the variance of the Least Squares estimator, along with the weight of the observations [10], once we are dealing with a stratified sample (GUJARATI, 2009).

Besides heterocedasticity, it is possible that there is endogeneity between the dependent variables (both occupation and the logarithm of the per hour wage) and the explanatory variable of interest, pap_i, as explained in the previous section. If so, the model estimators will be biased and inconsistent. In this case, it will be necessary to instrumentalize the endogenous variables, estimating the models via instrumental variables (IV) , using two-stage least squares (2SLS) or the Generalized Method of Moments (GMM). Thus, our final model would be estimated holding equations 3.1 and 3.4, along with equation 3.5, showed as follows:

$$pap_i = z\delta_1 + e \tag{3.5}$$

Equation 3.1 is the selection equation; equation 3.4 is the one in which results we are interested; and equation 3.5 describes the behavior of the endogenous variable. As shown in equation 3.5, z represents the instruments vector, which will be illustrated in the following section, and e is the non-observed random errors. It is important to note that for a variable (or a set of variables) to be a strong instrument, it must be highly correlated with the endogenous variable but not correlated with errors (ALMEIDA, 2012;

[10] Probability weight of individuals, calculated by IBGE and present in the PNAD microdata file.

GREENE, 2003). Therefore:

$$E(z, \varepsilon_i) = 0 \tag{3.6}$$

$$E(z, pap_i) \neq 0 \tag{3.7}$$

In terms of the model estimation, we assume that $E(z, e) = 0$, so that the income equation is identified under the presence of endogenous covariates. Also, z and p_i are always observed, and lny_i is observed when $p_i = 1$.

Since pap_i is present in both the selection and the income equation, it will be instrumentalized in the two of them. Because we intend to use more than one instrument for only one endogenous variable, we expect our model to be overidentified. Thus, the estimation method we intend to use is the IV-GMM, which is efficient under hetoreskedasticity (GREENE, 2003).

3.3.2 Sample, Data Treatment and Variables

For this analysis, we will use microdata from PNAD 2015 [11], an annual sample survey planned and conducted by IBGE, at household level, in order to produce results that cover the entire Country. These results include data about general population characteristics, education, labor market, income and housing (IBGE, 2016). Besides, the 2015 survey also includes a supplementary questionnaire regarding sports and physical activity practice, which suits the objectives of this research.

In order for the data to fit the requirements of this analysis, some treatment procedures and filters were applied to the sample. Since we are testing the socioeconomic impacts of the PAP, first we kept on our sample only the individuals that were selected and answered the supplementary questionnaire regarding the PAP. Also, since this paper seeks to analyse PAP impacts on adults' well-being, we selected the observations of individuals that were between 18 and 65 years old, which are the ages of participation in the labor market[12]. Ages above or below these limits could mislead the estimation results, since

[11] PNAD was shut down in 2016, with the disclosure of data from 2015, and replaced by Continuous PNAD, which provides a more embracing territorial coverage along with conjunctural quarterly information about workforce and labor market at national level. Still, the Continuous PNAD displays a more condensed questionnaire, containing less in-depth information.

[12] The age groups comprised in these age classifications may vary. One of the most commonly found in the literature and used by Talbot et al. (2005) is: young adults (20-45), middle-aged adults (46-65),

they correspond to phenomena that are not the object of study of this research, e.g. child labor (OLIVEIRA; JUSTUS, 2017).

Moreover, according to Vaz and Hoffmann (2007) suggestions, as well as done by Oliveira and Justus (2017), we also dropped from our sample individuals with zero income and undeclared or greater than one hundred thousand *reais* income, in addition to those individuals who were classified as "undetermined" on the *"years of study"* category (whose information on schooling was not declared or did not allow for the IBGE *years of study* classification). These cuts on the sample are justified once the information of the explanatory variables on the income equations must be valid and must not be biased by outliers.

Thus, after these treatments, our sample ended up with $55,120$ observations for the unrestricted equation (the occupation equation, the first part of the Heckman estimation) and $37,063$ observations for the restrict estimation (the income equation, which does not consider the unoccupied individuals). These sample sizes correspond to a population of $123,839,853$ and $81,509,696$ respectively, when expanded by the frequency weights calculated by the IBGE. However, it is important to state that these frequency weights were not used on the model estimations, they were only used to estimate the population size correspondent to our final sample. As stated on the previous section, we used probability weights - instead of frequency weights - on our estimations, in order to correct it for heteroskedasticity and to suit our sample to the Brazilian population, according to the estimations from IBGE.

With the dataset ready, we now turn to the construction of the variables used on this two-part estimation. They are presented and explicated on Table 3.1 below:

Table 3.1: Variables Construction and Description

Variable	**Description**
p_i	The binary dependent variable, which assumes the value 1 if the individual is occupied, and 0 in case it is unoccupied or inactive.
y_i	The dependent income variable, the natural logarithm of the per hour wage earned by each individual on its main work.

and older adults (over 65 years old). Still, Petry (2002) uses the following classification: young adults (18-35), middle-aged adults (36-55), and older adults (older than 56 years).

pap	A binary variable that indicates whether the individual practices some type of sports or physical activity out of their working hours (1) or not (0).
sex	A binary variable for the individual's sex, which assumes the value 1 for women.
age	A discrete variable representing the individual's complete years of age.
color/race	Four binary variables to distinguish the color/race declared by the individual: white (base category), indigenous, yellow, brown and black.
school	Schooling, a discrete variable indicating the complete years of study of the individual.
school_9	A variable with the objective of capturing changes in the rate of return of schooling as from 9 years of study (that is, the impacts on income of years of study after middle school). It is calculated as follows: $school_9 = \alpha_i(school - 9)$, Where: $\alpha_i = 1$ if $school > 9$ and $\alpha_i = 0$ if $school \leq 9$.
other_income	A binary variable indicating whether the individual has any other source of income besides their work, such as receiving interest, rent, pensions or social benefits (1), or if they do not (0).
exp	A variable representing the labor market experience of the individual, calculated as follows: $exp = age - school - 6$, once mandatory school attendance starts at the age of 6, in Brazil.
exp^2	Experience squared, in order to analyse the rate of change of each additional year of experience, since the effect of this variable on income is not linear.

union	A binary variable indicating if the individual was (1) or was not (0) associated with a trade union by the month in which the questionnaire was applied.
occupation type	Eight binary variables to distinguish among nine categories of occupation: (a) CEO and managing positions, (b) jobs related to sciences and arts, (c) technical jobs of high school level, (d) administrative positions, (e) sales and commerce services jobs, (f) services in general (base category), (g) farm work positions, (h) manufacturing industry jobs, and (i) army and military positions.
retired	A binary variable indicating whether the individual was (1) or was not (0) retired by the time in which the questionnaire was applied.
householder	A binary variable for the individual's condition in the family, being 1 for householder, and 0 for children/spouses/ others.
spouse	A binary variable that indicates whether the individual does have a spouse (1) or doesn't (0).
dependent kids	A variable indicating the individual's number of dependent kids younger than 14 years old. This variable will be used only on the occupation equation, since it affects the individuals' decision to participate on the labor market, but not necessarily their income earning.
chores hours	A continuous variable, representing the number of hours spent on domestic work, such as chores. This variable will also be used only on the occupation equation.
rural	A binary variable indicating whether the individual's household is located in a rural (1) or urban (0) area.
metropolitan	A binary variable indicating if the individual's household is located in a metropolitan region (1) or if it isn't (0).

macro_5	Four binary variables to distinguish the five Macro-Regions of Brazil: North, Northeast, Central West, South and Southeast (taken as base category).

(a) Occupation types sorted according to IBGE categories presented on the PNAD questionnaire.

(b) Occupation categories and trade union association taken as explanatory variables only on the income equation since these characteristics are only available for occupied individuals.

Source: Elaborate by the author.

For that first dependent variable, *pap*, those individuals who claimed they practiced sports or physical activities that did not require much body movement and energy expenditure were classified as non-practitioners of physical activity (that is, $pap = 0$). Those sports/physical activities were: fishing, bowling, pool, billiards, card games, dice games, chess, checkers, other sports with motor use and other cards sports. Also, because they were practiced by very few people, we also classified as non-practitioners of physical activity those who declared they practiced the following sports: motoring, motorcycling, *enduro karting*, motocross, rally and powerboat.

In terms of the Macro-Regions control variables, we also tested different binary variables for the Southeast and the Central West excluding the State of *São Paulo* and the Federal District, respectively (and generating another category for each of them), since they are considered outliers when it comes to income earning. This happens because the State of *São Paulo* concentrates the majority of the industrial, highly specialized, and human capital applicant work stations, which leads to higher wages and income earnings. In the case of the Federal District, the highest salaries come from the high Executive and Judiciary positions that exist in the Country's capital. Even with those modifications, the estimation results did not change much, with only a minimal magnitude variation on the parameters of these variables.

Also, we end up not using *age* as an explanatory variable on both income and occupation regressions because it is highly correlated with the schooling and experience variables, due to how we chose to construct them, exceeding the tolerated levels indicated by the TOL_j multicollinearity index (that is, if $TOL_j < 0.1$).

As discussed on the previous section, it is possible that we find endogeneity between the dependent variables (p_i and y_i) and the *pap*, since the dependent variables

depend on the occupational situation and income-earning potential of each individual, capacities that are largely non-observable - depending on factors such as skills, vocation, ambition, family situation, availability of time - making them correlated with some of the main observed determinants of wages and occupation, such as schooling and type of occupation (ROCHA; SOUZA CAMPOS, 2011).

If that's the case, - if there is endogeneity - then the OLS model estimators will be biased and inconsistent. In order to correct these problems it will be necessary to use instrumental variables as the estimation method. It is important to note that for a variable to be a strong instrument it must be highly correlated with the endogenous variable - in this case, the *pap* - but not correlated with errors - thus, not correlated with the dependent variables, p_i and y_i.

Since the individuals' behaviour, preferences and perceptions regarding health and lifestyle are strong determinants of their decisions over practicing physical activity (SNYDER, 1985; FEJGIN, 1994), we aim to construct instruments that represent this behaviour, as it seems to be uncorrelated with occupation or income. In order to do so, we rely on several qualitative questions regarding the reasons why the individuals practiced (or did not practice) physical activities or sports (health, figure, time, socialization), and also regarding their preferences on government areas of investment and expenditure (healthcare, education, public security, sports incentives), so that we end up with the four IV displayed on Table 3.2, which should represent these behaviours and preferences of the individuals regarding PAP.

Table 3.2: PAP Instrumental Variables Construction and Description.

IV	**Description**
public_neigh_space	A continuous variable representing the average of positive answers to the existence of public areas dedicated to sports or physical activity practice, by PSU (primary sampling unit). Since it is a categorical answer (*yes* or *no*), this variable is comprehended between 0 and 1, and gives an idea of how the neighborhoods are propitious/friendly to PAP.

ipublic_pap	Indicates if the individual thinks there should be public investment on sports or physical activities development on his/her neighborhood.
health	Indicates if the individual has chosen to practice or not to practice sports or physical activities because of health reasons. That is, if the individual cares about his/her health when deciding to practice - or not to practice - physical activities.
others_social	Indicates if the individual has chosen to practice, not to practice or quit practicing sports or physical activities due to having fun/socialization reasons (or lack of will/socialization issues).

Here, we consider PSU as a proxy variable for neighborhood.

Source: Elaborate by the author.

From that, we end up using four instrumental variables for only one endogenous covariate (*pap*), so that we estimate an overidentified model (GREENE, 2003). Under these conditions - and also under heteroskedasticity -, the most efficient estimation method is IV-GMM (GREENE, 2003; WOOLDRIDGE, 2010), which will be used here. Finally, with all the variables and econometric procedures already highlighted, we now step in to the results analysis.

3.4 Results and Discussion

Before we indeed move to the analysis of the estimations results, alongside estimating these income and occupation models for the full sample of individuals, we will also run them for each one of the sexes and age groups, once PAP is much differently faced - and is also such a different custom, developed within some societal hierarchies - by men and women, and throughout different ages (KING et al., 1992; VAITSMAN, 1994; AQUINO, 1996; POWELL; SLATER; CHALOUPKA, 2004). We built our age groups based on Petry (2002)'s definitions, so that we ended up with three age categories: young adults (aged 18 to 35), middle-aged adults (36 to 55 years old), and older adults (older

than 56 years)[13]. Thus, we end up with nine subsamples, so that we can compare the *pap* effects results among ages and between sexes.

As stated on the previous section, in order to find the model that best fits our data, we first test for the presence of endogeneity between *pap* and the dependent variables, p_i and y_i. The tests statistics are calculated under the null hypothesis of exogeneity. The results are shown on Table 3.3. It is worth to note here that because our endogenous covariate is binary - and not continuous - there might be some fitting problems if we estimate the models using regular IV-GMM, which applies an OLS estimation on its first stage, by default. In this sense, non-linearity of the first-stage dependent variable would not be taken into account, leading to inaccurate *pap* predictions - which, in turn, would invalidate the IV process (ANGRIST; PISCHKE, 2008; FREEDMAN; SEKHON, 2010).

In order to correct for these estimation problems, we adopt two different procedures - one for the occupation equation, in which the dependent variable is also binary, and another for the income equation, which dependent variable is continuous - as suggested by Freedman and Sekhon (2010), Roodman (2009), Chiburis, Das, Lokshin, et al. (2011), and Angrist and Pischke (2008):

a. According to Angrist and Pischke (2008), to account for this non-linearity specification of the IV estimation on the income equation, first, we should estimate a binary choice model where the binary endogenous variable, *pap*, is regressed on the instruments (including the exogenous explanatory variables of the model), and save the predicted *pap* probabilities. Second, we run a standard IV-GMM estimation, using as instruments for the binary endogenous variable the instruments used in the binary choice model as well as the fitted probabilities obtained from it, $\hat{pap}$.

   Angrist and Pischke (2008) argue that the problem with simply estimating the second stage of the IV endogeneity correction with $\hat{pap}$ instead of the actual *pap*[14] is that only OLS estimation guarantees to produce first-stage residuals that are uncorrelated with fitted values and covariates. Thus, a simple alternative that avoids problems due to an

[13]The *retired* variable will only be used on the estimations for the all-ages samples and for the older adults subsamples, for both sexes. We chose to use this variable only on these discriminated subsamples due to matters of representativeness and significance: on the all-ages sample, the proportion of retired individuals was on the order of 7.57%; for the young adults, this proportion did not even reach 1%, while almost 3.6% of the middle-aged adults were retired. On the other hand, at around 39.2% of the older adults were already retired by the time of the application of the PNAD questionnaire.

[14]A procedure that, in theory, would purge endogeneity.

incorrect nonlinear first-stage is to use the nonlinear fitted values, $\hat{pap}$, as instruments on the OLS first-stage regression.

b. To instrumentalize the binary p_i equation, we follow a procedure that computes maximum - likelihood (ML) estimates of a Bivariate Probit (BP) model, which assumes that the outcome and the endogenous covariate are each determined by latent linear index models with jointly normal error terms (CHIBURIS; DAS; LOKSHIN, et al., 2011). Although this model corrects for the non-linearity problem of the OLS regression, the BP estimator is not robust if the the BP model is misspecified, that is, if the data generating process does not follow the normal distribution assumed by the Probit estimation. In other words, when the error terms exhibit excess skewness or excess kurtosis, it often leads to highly biased BP estimates, and tests based on BP estimates greatly overreject a true null hypothesis (CHIBURIS; DAS; LOKSHIN, et al., 2011).

   Thus, to account for this possible misspecification problem, we will also correct non-linearity using the econometric procedure described for the income equation (ANGRIST; PISCHKE, 2008), i.e., including the predicted $\hat{pap}$ as an instrument on the IV-Probit estimation of the occupation equation, so that we can compare the results from both procedures.

After these specifications of the econometric procedures, we actually turn to the endogeneity tests statistics, displayed on Table 3.3.

These results evince that both dependent variables - p_i and y_i - are endogenous to the practice of physical activity, at least in some subsamples, as we previously expected, since we reject the test null hypothesis of exogeneity. It is interesting to note that pap is endogenous to the individuals' occupational situation in almost all subsamples, meaning that time allocation affects both the decisions to engage in the labor market and to practice physical activities regardless of the age or sex of the individual. That is, the practice of physical activities affects the individuals' time availability, which, in turn, affects their occupational situation, and vice-versa: the individuals' occupational situation affects their decision over practicing physical activities or sports, mostly due to time availability (COLEMAN, 1961; KING et al., 1992; WEINBERG; GOULD, 2016).

Table 3.3: Endogeneity Test Statistics for All Subsamples.

		All Ages		Young Adults		Middle-Age		Older Adults	
	Test Statistic	p_i	y_i	p_i	y_i	p_i	y_i	p_i	y_i
All	**Wald Test/ GMM C Statistic**	79.39***	6.790***	4.95**	2.936*	83.49***	6.166**	18.88***	0.500
	F-statistics of Strong Instruments	-	13,014***	-	11,487***	-	3,615***	-	275.4***
	Hansen's J Overrindentification	-	26.561***	-	25.039***	-	15.709**	-	7.309
Men	**Wald Test/ GMM C Statistic**	12.05***	3.079*	2.96*	0.253	17.28***	3.366*	0.94	0.370
	F-statistics of Strong Instruments	-	4,686***	-	4,702***	-	917.5***	-	208.3***
	Hansen's J Overrindentification	-	22.895***	-	39.365***	-	6.719	-	5.975
Women	**Wald Test/ GMM C Statistic**	56.58***	0.471	6.52**	0.730	53.65***	0.684	20.02***	1.781
	F-statistics of Strong Instruments	-	15,401***	-	8,704***	-	5,760***	-	337.5***
	Hansen's J Overrindentification	-	17.473***	-	4.392	-	14.899***	-	11.525**

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Source: Elaborate by the author.

The only subsample in which *pap* and p_i were not endogenous is the Men Older Adults subsample. Thus, in this specific case, *pap* would affect the individuals' decision over participating on the labor market, but not the other way around: these individuals' occupational situation would not affect their decision over engaging on physical activities or sports out of the working hours. As discussed by King et al. (1992) and Vaitsman (1994), this result highlights how much PAP is intrinsic to men's behavior, and how it depends much more on health, lifestyle and predisposition characteristics when it comes to the older adults.

When it comes to income, the results are quite different, but also very interesting: while taking into consideration both men and women, *pap* and income are endogenous for almost all age groups, except for the older adults - which makes sense since, for those individuals, the decision over engaging on PAP is much more related to health, lifestyle preferences and predisposition matters (KING et al., 1992; WEINBERG; GOULD, 2016). Still to reinforce this argument, when looking at the men and women subsamples, the *pap* was also not endogenous for this latter age group, in any of the sexes subsamples.

As we expected, there were some differences between the PAP-income endogeneity results for men and women. On one hand, for women, *pap* was not statistically endogenous to income in any age-group subsample, highlighting how the

women's decision over practicing physical activities is much more related to social hierarchical determinations, regarding duties and sports "sexual division" (VAITSMAN, 1994; FLORINDO et al., 2009), and also related to their time availability. In this sense, PAP is much more common and intrinsic to male behaviour, than it is for women. On the other hand, with regards to the men's subsamples, *pap* presented itself endogenous for middle-aged men, and also when we considered men of all ages. In this sense, we can state that, as for the older adults, the decisions over engaging on PAP for the young adults are much more associated to the individuals' behaviour and preferences regarding PAP and lifestyle, than it is to income earning.

Also, though we have strong instruments (which comes from the rejection of the null hypothesis of weak instruments of the F-statistic test), Table 3.3 overidentification test shows that our set of instruments are not valid, and that the model is not correctly specified[15], for almost every subsample. Still, because the tests regarding the validity of the instruments point to opposite directions - so that one of them indicates that the IV we chose to instrumentalize the *pap* are strong -, and due to the difficulty of finding other *pap* instruments that are not correlated to the individuals' income within the PNAD questionnaire[16], we continue to use on our estimations these instruments referred to on the previous section.

With the instrumental variables already defined - those variables that encompass the individuals' behaviour and preferences regarding health, lifestyle and PAP -, we can turn to the estimation of the main proposed models. All five models displayed here were estimated following Heckman's specifications (HECKMAN, 1979), and all equations were estimated using White's robust standard errors in order to correct for heteroskedasticity.

Model (1) instrumentalizes *pap* on both occupation and income equations, through Angrist and Pischke (2008)'s method, so that we have an IV-Probit (p_i) IV-GMM (y_i) estimation; model (2) also instrumentalizes *pap* on both equations, but using the BP method for the occupation equation (CHIBURIS; DAS; LOKSHIN, et al., 2011), so that we end up with a BP-ML (p_i) IV-GMM (y_i) model; the third model (3), however,

[15]That is, the models are overidentified - which they actually are, since we have more instruments than endogenous variables on the estimations.

[16]Due to individual privacy matters, PNAD observations are only identified by numeric codes, which vary across IBGE - and other agencies - researches and across PNAD periodic questionnaire applications, not making it possible to join datasets and use data from external researches.

instrumentalizes *pap* only on the income equation, so that we estimate IV-Probit (p_i) OLS (y_i) equations. Model (3) is only estimated for those subsamples in which *pap* was not statistically endogenous to y_i at a significance level of 10%. Lastly, models (4) and (5), do not instrumentalize *pap* on any of the two parts of the estimation, so that model (4) is estimated via Maximum Likelihood (ML) and model (5) is Heckman's two-step estimator. Because *pap* and p_i were not statistically endogenous on the Men Older Adults subsample, the models we estimated, for this specific subsample, were: (1) Probit IV-GMM, (2) BP IV-GMM and (3) Probit OLS. The PAP results of both parts of these estimation methods are displayed on Table 3.4.[17]

Table 3.4: PAP Coefficients for Occupation and Income Equations, for All Subsamples.

		All Ages		Young Adults		Middle-Age		Older Adults	
	PAP	p_i	y_i	p_i	y_i	p_i	y_i	p_i	y_i
All	**(1)** $\hat{pap}$ **IV-Probit IV-GMM**	-0.139***	0.102***	-0.015	0.046***	-0.292***	0.170***	-0.358***	0.153*
	(2) $\hat{pap}$ **BP IV-GMM**	-0.337***	0.123***	-0.238***	0.063***	-0.405***	0.186***	-0.584***	0.211*
	(3) $\hat{pap}$ / *pap* **IV-Probit OLS**	-0.139***	-	-0.015	-	-0.292***	-	-0.358***	0.109**
	(4) *pap* **Heckman - ML**	0.043***	0.072***	0.039	0.030**	0.045*	0.118***	0.085**	0.053
	(5) *pap* **Heckman - 2S**	0.049***	0.067***	0.052***	0.023**	0.066***	0.112***	0.068*	0.046
Men	**(1)** $\hat{pap}$ **IV-Probit IV-GMM**	-0.062	0.101***	-0.012	0.044*	-0.269***	0.172***	0.162**	0.111
	(2) $\hat{pap}$ **BP IV-GMM**	-0.373***	0.133***	-0.305***	0.053	-0.451***	0.193***	-0.690***	0.489*
	(3) $\hat{pap}$ / *pap* **IV-Probit OLS**	-0.062	-	-0.012	0.034**	-0.269***	-	0.162**	0.017
	(4) *pap* **Heckman - ML**	0.078***	0.079***	0.102***	0.044**	0.093**	0.120***	0.164**	0.033
	(5) *pap* **Heckman - 2S**	0.077***	0.069***	0.063**	0.029**	0.113***	0.099***	0.130**	0.044
Women	**(1)** $\hat{pap}$ **IV-Probit IV-GMM**	-0.119***	0.095***	-0.057	0.043	-0.221***	0.149***	-0.337***	0.077
	(2) $\hat{pap}$ **BP IV-GMM**	-0.248***	0.111***	-0.133***	0.051*	-0.328***	0.163***	-0.510***	0.079
	(3) $\hat{pap}$ / *pap* **IV-Probit OLS**	-0.119***	0.086***	-0.057	0.030	-0.221***	0.144***	-0.337***	0.102
	(4) *pap* **Heckman - ML**	0.028	0.072***	0.002	0.024	0.024	0.117***	0.047	0.066
	(5) *pap* **Heckman - 2S**	0.043**	0.065***	0.032	0.013	0.038	0.123***	0.040	0.028

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

(a) $\hat{pap}$ stands for instrumentalized PAP, while *pap* stands for non-instrumentalized PAP.

(b) Men's Older Adults models are: Probit IV-GMM; BP IV-GMM and Probit OLS.

Source: Elaborate by the author.

With regards to the selection equation, i.e., the occupation equation estimations, we can observe from the results on Table 3.4 that the estimation methods

[17]Full estimation results of all the explanatory variables employed on these models are displayed on Appendix B, along with other model adjustment statistics.

(1), (2) and (3) - in which *pap* is considered endogenous to p_i - present a negative and significant impact of the practice of physical activities out of the work hours on the individual's probability of being occupied, for almost all subsamples, except for the Men Older Adults[18]. This relationship can be viewed through Coleman (1961)'s idea of limited time, so that the choice of practicing any physical activity would depend on the availability of time, which is more scarce for those individuals who work. On the other hand, when not instrumentalizing *pap*, the results - from models (4) and (5) - either showed a positive relationship between this variable and the individual's probability of being occupied, or the relationship was not statistically significant.

Since p_i and *pap* are endogenous on the majority of our subsamples - making IV estimation necessary -, these results highlight the importance of accounting for endogeneity, so that we do not end up with biased and inconsistent estimations (GREENE, 2003). Once again, the exception here is the Men Older Adults subsample, to which models (3)[19], (4) and (5) are more adequate, since they do not account for endogeneity. In this case, participating on PAP for a male older adult would have positive impacts on his probability of being occupied, mostly due to lifestyle, health and predisposition matters (KING et al., 1992; MUSHKIN, 1962; WEINBERG; GOULD, 2016).

Although the results from Table 2.3 of the previous Chapter show that the average of occupied people is greater within those individuals that practice physical activities - countering the results presented on Table 3.4 while estimating the occupation equation with instrumental variables - , the average *t*-tests do not indicate the direction of this causality effect between p_i and *pap*. Thus, since *pap* and p_i are endogenous (making the IV estimation necessary, and consequently, models (1), (2) and (3) more adequate to the data[20]), the mean difference tests results might be reflecting a bidirectional causality.[21]

When it comes to income, all five estimations results displayed on Table 3.4 corroborate the theoretical framework and the hypothesis discussed on this paper, since

[18]Subsample in which *pap* and p_i are not endogenous.

[19]Probit OLS.

[20]Again, except for the Men Older Adults subsample.

[21]That is, since *pap* effects on p_i were found to be negative, the *t*-tests results might be indicating the impact of p_i on *pap*, once these variables are endogenous. In this sense, being occupied might increase the individuals' probability of engaging on PAP, mostly due to income and employment security (GROSSMAN, 1972).

pap coefficients were positive and statistically significant, for most subsamples.[22] Thus, the higher the accumulation of health capital (MUSHKIN, 1962) through PAP, and the higher the investment in human capital due to higher health stocks (BECKER, 2007), the higher is the individuals' income earning. In other words, the practice of any kind of physical activity out of the work hours increase the individual's income earning. As argued by Mushkin (1962) and empirically discussed by Oliveira and Justus (2017) and Teixeira (2016), behaviours that increase the individuals' health stock tend to lead to increases on their income earnings.

Even though all estimation methods and all subsamples results point to the same direction of the impact of the PAP on the individuals' income earning, the magnitude of this impact significantly varied both among subsamples and among estimation methods. In terms of comparison of the different models we estimated, it is noticed that models (1), (2) and (3) show quite similar results with regards to the *pap* variable within each subsample. Likewise, the *pap* estimated coefficients from models (4) and (5) - that do not consider *pap* endogenous in any step of the Heckman's sample selection estimation - also were very resembling to one another. When compared to the IV estimations - that is, when compared to models (1), (2) and (3) -, models (4) and (5) displayed an underestimation of the PAP effects on income earnings, for all subsamples. However, because we found endogeneity between *pap* and p_i, then the results from models (1), (2) and (3) tend to be more robust and consistent[23].

It is important to state that for those subsamples in which *pap* was not endogenous to y_i (all three *"older adults"* subsamples, as well as all four age-groups *"women"* subsamples - including that of the *"older adults"*-, along with the *"men - young adults"* subsample), the most adequate and parsimonious model is model (3)IV-Probit OLS, that does not account for endogeneity on the second part of the sample selection estimation. As for the other subsamples, in which *pap* is endogenous on both parts of the estimation, either models (1) and (2) best fit the requirements of the data, although it should be reminded that non-OLS first stage estimates do not guarantee to produce residuals that are uncorrelated with fitted values and other covariates (ANGRIST; PISCHKE, 2008). Also, if model (2) is misspecified, then the BP estimates are biased

[22]None of our results showed a negative impact of PAP on the individuals' income earnings, so that when this *pap* coefficient was not positive and significative, it was just not statistically significant.

[23]Except for the Men Older Adults subsample, to which models (3), (4) and (5) are more adequate.

(CHIBURIS; DAS; LOKSHIN, et al., 2011). From that, the results from (1) are preferred to (2).

Taking this model selection into consideration, the results from Table 3.4 show that, on average and for all individuals, engaging on PAP or sports out of the work hours increases the individuals' income earning by around 10.2% (1) to 12.3% (2). These figures are similar for men - from 10.1% (1) to 13.3% (2) - and smaller for women, being of the order of 8.6% (3). This behaviour is repeated within age-groups, so that for the young adults, while PAP increases men's income earning by 3.4% (3), for women in this age-group, the effect of participating on PAP is not statistically significant. For middle-aged men, the fact of practicing physical activities out of their working hours increases their income earnings, on average, by around 17.2% (1) to 19.3% (2); for middle-aged women, this income increase figures around 14.4% (3). However, for the older adults - those over 56 years of age -, a similar result can be observed for both men and women: on average, the PAP does not affect older women's and older men's income earnings.

Vaitsman (1994) argues that this difference of the PAP effects between sexes is derived from the hierarchical construction of PAP and sports participation within certain characteristics of determined social systems, meaning that PAP is much differently faced by the individuals of opposite sexes. That is, PAP is much more common and intrinsic to men's lifestyle than it is to women's - to whom other "obligations" and characteristics are "socially" designated. In order to illustrate this argument, we highlight the proportion of PAP practitioners in our sample, between sexes: in general, 42.5% of men engaged on PAP, while only 35% of women participated on physical activities practice. Thus, this quite dissimilar way of assessing PAP between men and women may have an effect on how it impacts the individuals' health stock, productivity and, therefore, their income earnings.

As for the different PAP impacts on income throughout the age groups, our estimation results corroborate the health capital theory discussed by Mushkin (1962), in the sense that the lower the individuals' amount of health stock - which reduces the older they get (BECKER, 1962; 2007) -, the higher the effect on their income earning of a behaviour that increases their health stock, that is, participating on PAP. However, the PAP impacts on individuals' income earnings seem to have an inverted U-shaped behavior

throughout the age groups, similar to the very impacts of age and experience on income, as discussed by Mincer (1958) and Schultz (1961) on the human capital approach.

That is, the higher the individuals' age and experience - and, hence, the higher their productivity and income earnings (MINCER, 1958) - the greater are the PAP impacts on the individuals' per hour wage, up to a point when biological and health issues start to adversely affect their productive performance, and again, their income earnings - mainly in jobs where physical effort or motor skills are involved (MINCER, 1958; SCHULTZ, 1961) -, so that PAP positive impacts on income decrease, being even non-significant. In this sense, there should be health problems and diseases, that come after a certain aging point, that hinder higher PAP impacts on individuals' well being, and hence, on their income earnings.

Lastly, from Tables 2 to 10 displayed on Appendix B, with regards to Heckman's two-part econometric procedure, it is observed that the probability of being occupied (λ) is statistically significant on the income equation, meaning that if we did not consider the individual decision of participating on the labor market, then the estimations would be biased by sample selection.[24] As for the other Mincerian equation consolidated variables, they presented, on average, both similar and expected results on all five estimated models: positive impacts of schooling and experience, as well as positive impacts of being householder, having a spouse and living on metropolitan areas on the individuals' income earning. Still, women, black and brown people, and people living on rural areas tend to earn lower income than a white man living on an urban area.

3.5 Final Remarks

Insufficient physical activity is the fourth of the 10 leading risk factors for global mortality (WHO, 2018b). Still, data from the 2013 PNS (IBGE, 2014) show that, in Brazil, one in two adults does not practice the minimum level of physical activity recommended by the World Health Organization (WHO). Alongside with having multiple positive impacts on physical, mental and perceived health (WHO, 2018a;

[24]The only subsample in which the Wald test of independent equations showed that $\rho = 0$ - that is, that the occupation and income equations were statistically independent - was the *Older Women* subsample. However, estimating it through regular OLS did not quite change the coefficients results.

WESSEL et al., 2004; BABISS; GANGWISCH, 2009; MEYER; CASTRO-SCHILO; AGUILAR-GAXIOLA, 2014), PAP can also affect socioeconomic characteristics of individuals.

Thus, anchored on Mincer (1958), Mushkin (1962), Grossman (1972), and Becker (2007) human and health capital theories, we intended to analyze PAP as an economic problem, so that it can affect individuals' productivity and, hence, their income earnings. This theoretical framework considers that, once PAP enhances individuals health stock, then it can affect their productivity by two means: (i) with a higher health stock, individuals loses less working days, and also are more productive, once they feel better (MUSHKIN, 1962; GROSSMAN, 1972); (ii) an elevated health stock also enhances individuals' life expectancy, which makes their investment in human capital more profitable. Hence, a higher stock of human capital leads to higher earnings (MINCER, 1958; SCHULTZ, 1961; BECKER, 2007).

Corroborating our hypothesis that PAP has positive impacts on individuals' income earnings, the results, after the estimation of an income equation through the sample selection method proposed by Heckman (1979), controlling for sociodemographic characteristics and for endogeneity, for a 2015 PNAD Brazilian sample, show that PAP coefficients were, in general, positive and statistically significant. Thus, we can remark that the practice of any kind of physical activity out of the work hours increases the individual's income earnings - that is, increases their per hour wage - by 10.2% (IV-Probit IV-GMM) to 12.3% (BP IV-GMM), on average.

Therefore, the higher the accumulation of health capital (MUSHKIN, 1962) through PAP and the higher the investment in human capital due to higher health stocks (BECKER, 2007), the higher is the individuals' income. These PAP impacts vary across sexes and throughout age groups, so that PAP displays greater impacts on men's income earnings, and an inverted U-shaped behavior across age groups, similar to the very impacts of age and experience on income, as described and argued by Mincer (1958), Schultz (1961), and Mincer (1974).

These findings show that PAP not only can lead to positive impacts on the individuals' health stock, but can also affect their income earnings, becoming a socioeconomic issue that affects the quality of life of individuals through many means. As stated by the WHO, "*Failure to recognize and invest in physical activity as a*

priority." (WHO, 2018a, p. 16) can lead to multiple costs to society, so that ongoing inaction with regards to PAP investments will contribute to further negative impacts on health systems, the environment, the individuals' productivity and income earning, economic development, community well-being and quality of life for all. Therefore, engaging on PAP may not only positively affect the individuals' well-being, but also the well-being of society as a whole, insofar as it affects their productivity, and thus, economic development.

For further work, we suggest to find new variables or other methods, such as treatment effect and propensity score matching (PSM), or adding a variable for controlling for the individuals' health stock when estimating the economic impacts of PAP, so that the results can be compared to the ones found on this research, as in a robustness analysis.

Furthermore, since PAP impacts differ among socioeconomic and sociodemographic groups, Powell, Slater and Chaloupka (2004) and Meyer, Castro-Schilo and Aguilar-Gaxiola (2014) highlight that policy interventions regarding PAP and public health should address disparities considering both socioeconomic status and sex/race/ethnicity, so that it can reduce barriers related to the practice of physical activities, mainly to low-income individuals. In this sense, PAP public policies and interventions can also be used as tools for public health and income inequality reduction.

Chapter 4

The Socioeconomic impacts of Physical Activity Practice on Teenagers: The age-grade gap approach

4.1 Introduction

According to WHO (2018b), in 2010, nearly 80% of school going adolescents (aged 11 to 17) worldwide were insufficiently active. When it comes to Brazil, this reality is sustained: the same WHO data show that 86.7% of Brazilian teenagers are considered inactive, being that number smaller for boys (82%), and larger for girls (91.4%). More recent data gathered and estimated by Guthold et al. (2019) considering the PAP trend from 2001 to 2016 show that nearly 78% of boys and 89% of girls are insufficiently physically active in Brazil. That is, generally, in 2016, the prevalence of insufficient PAP was on the order of 83.6% for Brazilian school going adolescents aged 11 to 17.

The adequate level of physical activity for adolescents is a minimum of 60 daily minutes, in an intensity ranging from moderate to high (WHO, 2018b). By following these recommendations, there are many benefits, in terms of health, that the PAP can provide, such as: improvements in cardiorespiratory fitness, muscle strength and endurance, and bone health. PAP also helps reducing body fat, reducing the risk of cardiorespiratory

and metabolic diseases, and reducing the symptoms of anxiety and depression (WHO, 2018b), along with providing the experience of social and self enjoyment (GUTHOLD et al., 2019).

These positive impacts of PAP on physical and mental health have already been widely studied and tested, as have Babiss and Gangwisch (2009) and Ekelund et al. (2012) recently done. Babiss and Gangwisch (2009), while being concerned with the mental health of adolescents, find that the higher the participation of teenagers on sports (and the higher the per week frequency of this sports participation), the smaller the odds of them suffering from depression and having suicidal thoughts. Similarly, Taliaferro et al. (2008) results also show that, for teenagers[1], sports participation reduced the risk of suicidal ideation for both boys and girls. When it comes to physical health, both Ekelund et al. (2012) and Chaput et al. (2013), for a US and a Canadian sample, respectively, find that more time spent on the practice of moderate to vigorous physical activities reduces the children's and adolescents' cardiometabolic risks, regardless of their amount of sedentary behaviour.

In addition to the consequences for physical and mental health, the PAP can also impact the quality of life of individuals in an economic aspect. The publication of the WHO (2018a) global report on physical activity, with a goal of reducing by 15% the global prevalence of physical inactivity in adults and in adolescents by 2030, highlights these socioeconomic impacts of PAP - or the lack of PAP - in terms of not only individual health benefits, but also in terms of health care systems cost, productivity loss, and environmental spillovers. In other words, the practice of physical activity affects not only the individuals, but the society as a whole.

Approaching these economic impacts of PAP for children and adolescents, we rely on their academic performance. In this sense, Fejgin (1994) associates academic performance and sports participation for US high school students, finding positive impacts of sports participation on grades and on adolescents' educational aspiration. However, when also analysing an US sample, but controlling for sample selection and for endogeneity, Ransom and Ransom (2017) results point out to non-significant impacts of sports participation on students' academic performance.

[1]Aged 12 to 18.

Also aiming to understand the association between extracurricular PAP and cognitive performance on teenagers, Esteban-Cornejo et al. (2014) show that, for 13 to 18 year old Spanish adolescents, the practice of vigorous extracurricular physical activities presents positive impacts on the cognitive variables related to verbal, numeric and reasoning abilities. Thus, participating on in multiple, organised physical activities out of the school hours may have benefits for the adolescents' cognitive performance, which may reflect on their academic results. On the other hand, Knaus, Lechner and Reimers (2018)' research shows positive impacts of an additional physical education class on German schools on both boys' and girls' cognitive (i.e. Math and German grades) and motor skills, and on girls' emotional control. However, when it comes to boys' social behaviour, the additional physical education class presented negative impacts, with regards to fights and rivalry.

Therefore, considering these prior evidences that PAP and sports practice can affect the individuals' cognitive, social and motor skills, the main objective of this paper is to analyze the PAP not only as a health problem, but also as a socioeconomic problem in Brazil, in that it can affect the individuals' well being, and so the academic performance of Brazilian adolescents.

The theoretical discussion of this research problem, as presented by Snyder (1985), relies on two different aspects related to the academic impact of sports practice: (i) on one hand, the development theory, introduced in Schafer and Armer (1968) and Rehberg (1969), and also discussed by Hanks and Eckland (1976) and Rees, Howell and Miracle (1990), points out that sports practice, once it develops in individuals characteristics such as teamwork, persistence, patience, leadership, commitment and organization, would lead to better academic - and also, socioeconomic - performance/accomplishments[2], alongside with being a central factor in adolescents' and young adults' character and discipline building, and socialization. (ii) On the other hand, Coleman (1961), noting that sports practice may be detrimental to academic goals and success, discusses the zero-sum theory, whereas the idea that there is a finite amount of time and energy to be spent by individuals, then there would be a trade-off between the successful results in the sports and in the academic fields.

[2]In addition to the results in the athletic field itself.

We build our hypothesis on the idea that, since PAP raises the health stock of individuals, besides enhancing character (as discussed by Rees, Howell and Miracle (1990)), then we expect a positive impact of PAP on academic performance. Thus, by seeking the objective of analyzing PAP as an economic problem, which affects the Brazilian adolescents' welfare, we intend to test PAP impacts on their age-grade gap, a measure of academic performance. Since the age-grade gap is a measure of grade retention, we expect the PAP to have negative effects on this variable.

We will use 2015 PNAD microdata, which also includes a supplementary questionnaire regarding sports and physical activity practice, in order to estimate regression models taking as dependent variable the presence/absence and the number of years of age-grade gap. Our sample will comprehend adolescents aged 15 to 17, which are the ages included on the supplementary questionnaire of the survey.

4.2 Literature Review

4.2.1 The Brazilian Educational System

Before we start our statistical analysis, it is important that we highlight some specific features of the Brazilian Educational System. According to the Law of Guidelines and Bases of National Education (LDB) (BRASIL, 1996), the current structure of the regular Brazilian education system comprises basic education - that includes early childhood education, primary and secondary education, and that is mandatory - and higher education. As it is common in most countries, there are public and private schools operating on the Brazilian school system. Still, the LDB resolutions apply to both of them.

Early childhood education, the first stage of basic education, is offered in kindergartens for children up to three years old and in preschools for children four to five years old. In terms of public schools, its offer is of municipal competence. The main purpose of early childhood education, as stated on article 29 of the LDB (BRASIL, 1996), is the integral development of children up to five years old, in their physical, psychological, intellectual and social aspects, complementing the action of the family and the community.

Primary education, with a minimum duration of nine years[3], ordinarily starts for children that complete six years of age by March 31^{st} of the referred school year[4] (BRASIL. Ministry of Education, 2006), but it also has the obligation to be freely offered by the Public Power - in this case, it is offered by municipal as well as by state governments - to all, including to those who did not have access to primary education at the regular age. The main objective of primary education is the basic formation of the citizen. For this, according to the article 32 of the LDB[5] (BRASIL, 1996), the primary education must comprise, along its nine years:

I. the development of the ability to learn, having as basic means the full mastery of reading, writing and calculating;

II. the understanding of the natural and social environment, of the political system, of technology, of the arts and the values on which society is based;

III. the development of the ability of learning, for the acquirement of knowledge and skills, and for the shaping of attitudes and values;

IV. the strengthening of family bonds, the bonds of human solidarity and mutual tolerance, on which social life is based.

Continuing in the school cycle, secondary education, namely high school, is the final stage of basic education, lasts at least three years and attends the general education of the student. Its main purposes are, as stated on article 35 of the LDB (BRASIL, 1996):

I. the consolidation and deepening of the knowledge acquired in primary education, allowing for the continuity of studies;

II. the basic preparation for work and citizenship of the learner, so that he/she continue learning so as to be able to adapt flexibly to new conditions of occupation or further improvement;

[3]Until the publication of the Law 11,274, from February 6^{th}, 2006 (BRASIL. Ministry of Education, 2006), primary education consisted of eight years, regularly starting for kids at the age of seven. Since then, and especially since 2010, when the regulation of this law became mandatory to all schools, primary education began to last 9 years, starting for children at the age of six years old.

[4]In Brazil, the school year starts by the end of January and it is finished by mid-December. It also must account at least 200 school days (BRASIL, 1996).

[5]Translated by the author.

III. the improvement of the learner as a human being, including ethical training and the development of intellectual autonomy and critical thinking;

IV. the understanding of the scientific-technological foundations of productive processes, relating theory to practice, in the teaching of each discipline.

Secondary education may include programs of general preparation for the labor market and, optionally, professional qualification - namely, technical schools. This technical education is separately offered from ordinary high school, although they can be allocated in the same school space, specially in the case of public schools. Thus, by opting for this technical secondary education, the student may finish high school with a professional diploma and a technical qualification. If the student had a regular trajectory through primary education - i.e. without any grade retention -, he/she should start high school by the age of 15, finishing basic education by the age of 17, again, if there are no grade retention.

Throughout the whole basic education, physical education (PE) classes are mandatory, as highlighted on the article 26, third paragraph, of the LDB (BRASIL, 1996). Thereby, it shows how important those PE classes are, not only for the physical and body awareness development of students, but also contributing to the human development and schooling of children and young people (MARCASSA; BUSS, 2016).

Although PE classes must be understood as a cultural manifestation which should incorporate the affective, cognitive and socio-cultural dimensions, Marcassa and Buss (2016) still point out that the "location" of those classes as a curricular component on schools are being questioned, i.e., if they should be taken as a regular discipline or just a complementary activity, despite its many benefits to socialization, and mental and physical health[6].

As previously highlighted on this section, if the individual did not have access to basic education at the regular age, it is a duty of the Public Power to provide that education at any age. This teaching modality, in which the content of primary and secondary education is condensed, is called Youth and Adult Education (EJA) . Primary education EJA is aimed at young people aged 15 and over who did not complete the stage between grades 1 and 9. On average, it lasts two years. Yet, secondary education EJA

[6]For further discussion on the "location" of PE in schools, see Marcassa and Buss (2016) and Pinto and Vaz (2009).

lasts around one and a half years, and is intended for 18 year olds (and older) that did not complete high school. Thus, it is possible for the individual to complete basic education, even out of the regular age range.

Still on the Brazilian educational system, after completing basic education, the individual can choose to start higher education, which comprehends both Undergraduate and Graduate courses, but that is not mandatory. Higher education can be joined at any age, as long as the individual has a degree of completed basic education, and after being approved on a general knowledge test[7].

Back to the basic education characteristics, - which is the analysis focus of this research, since we are interested on the academic performance of teenagers on their regular and mandatory school cycle - as we already pointed out on this discussion, primary education starts at the age of six, and if there are no grade retention, the student finishes the mandatory cycle (primary and secondary education) at the age of 17.

However, if the students' accomplishments on that school year do not meet the minimum level of knowledge and activities required, they are demanded to take that grade again on the next school year, that is, the students are held back in that grade. Therefore, they will not finish the basic cycle at the age of 17, and they will hold an age-grade gap, meaning that the age of the individual when coursing a determined grade is not the regular age assigned for that specific grade. With that said, this age-grade gap is the variable of interest on this research, since it can somehow measure the academic performance of students, without requiring their discipline grades.

For the matters of analyzing this age-grade gap, there is a singular characteristic of the Brazilian educational system that may affect this variable of interest: the continued progression. The continued progression is based on the idea of education in which the learning cycles last more than a year, that is, students must obtain skills and competencies in a cycle that is generally longer than a year or grade. Therefore, instead of being held back on a grade, students may continue through the cycle, given the chance to recover the content on the next grade (but on the same cycle) and through reinforcement classes.

[7]The majority of public Higher Education Institutions require this general knowledge test for the selection of their students, due to vacancy availability - which does not occur to some private institutions.

There are defenders and critics of this educational practice on the Brazilian system[8], regarding the effective knowledge acquired by the students, their real ability to learn, and their actual fitting into a new grade due to continued progression, but the main purpose of this practice is to regularize student flow over the school cycle, reducing the school failure rates, also as an attempt to reduce school withdraw due to grade repetitions.

However, relative to the main objective of this research, it is worth noting that the continued progression may affect the measurement of the age-grade gap as it may reduce grade repetitions. With that said, the age-grade gap may be underestimated, in terms of knowledge acquirement and academic performance throughout a regular school year. Nevertheless, we still intend to use this variable as our dependent variable while trying to understand the impacts of PAP on academic performance, since this underestimation may affect all students similarly, once the continued progression applies to all.

4.2.2 PAP Impacts on Academic Performance

In addition to being included in the school hours, the PAP provides many benefits for physical and mental health, when practiced according to WHO recommendations regarding regularity and time spent on these activities[9]. For younger people - teenagers and children -, these benefits include improvements in cardiorespiratory fitness, muscle strength and endurance, and bone health. PAP also helps reducing body fat, reducing the risk of cardiorespiratory and metabolic diseases, and reducing the symptoms of anxiety and depression (WHO, 2018b).

In the previous literature, as a way of identifying and empirically evaluating these benefits, Ekelund et al. (2012) seek to understand the independent and joint impacts of the time spent on PAP (from moderate to vigorous) and the time in which the individual is sedentary on the cardiometabolic risk factors. Using pooled section data from 1998 to 2009 for children aged 4 to 18 years[10], through descriptive statistics and exploratory data analysis, the authors analyzed as cardiometabolic health measures:

[8]For further discussions on the benefits and problems of continued progression, see Bertagna (2003), Arcas (2009), and Paro (2011).

[9]For teenagers (11-17 years old), the recommendations are at least 60 daily minutes of PAP, in an intensity varying from moderate to high.

[10]The data comes from The International Children's Accelerometry Database, and contains information from 14 surveys conducted in Australia, Europe, the United States and Brazil.

abdominal circumference, blood pressure, insulin levels, and HDL[11] cholesterol and triglycerides levels. Based on their results, they concluded that more time spent in the practice of physical activities, both for children and adolescents, leads to better health outcomes in related to these cardiometabolic risk factors (less factors in risk situation), regardless of the time spent on sedentary activity.

In an analogous study, but for a Canadian dataset of children aged 8 to 10 with at least one obese parent, using both covariance analysis and regression analysis, Chaput et al. (2013) also try to identify the combined association of time spent on moderate to vigorous physical activities and time spent on sedentary activities on children's cardiometabolic risk factors (waist circumference, fasting levels of triglycerides, systolic and diastolic blood pressure, and HDL cholesterol levels). As found by Ekelund et al. (2012), their results show that a high level of moderate to vigorous PAP was associated with reduced cardiometabolic risk, regardless of the children's amount of sedentary behaviour, that is: in linear regression, moderate to vigorous PAP was inversely associated with waist circumference and diastolic blood pressure and positively associated with HDL cholesterol, independent of the covariates related to sedentary time. However, the authors also point out that the type of sedentary behaviour might be more important than overall sedentary time in relation to cardiometabolic risks.

For a similar age group, but turning to the mental health benefits of the PAP, taking as object of study the high school adolescents in the United States, Babiss and Gangwisch (2009) aimed to identify factors that are inherent to the physical and mental health of these individuals that make sports practice an important tool in the prevention of depression and suicidal thoughts. The authors used an ordered logit for the frequency of sports practice in order to analyze if more time spent on sports practice is associated with depression and/or suicidal ideation, and also if the physical exercise itself, self-esteem, body weight, social support/acceptance and illicit substances abuse are mediators of these relationships. Their results showed that:

> "As sports participation increases, the odds of suffering from depression decreases by 25% and the odds of having suicidal ideation decreases by 12% after controlling for sex, age, race/ethnicity, public assistance, and physical limitations. Substance abuse, body weight, and

[11] High-density lipoprotein cholesterol.

> exercise did not mediate these associations. Consistent with self-esteem and social support acting as mediators of these relationships, the inclusion of these variables in the multivariate models attenuated the associations for depression and suicidal ideation." (BABISS; GANGWISCH, 2009, p. 376).

Thereby, Babiss and Gangwisch (2009) concluded that not only incentives, but also opportunities for sports practice should be created, as they can prevent depression - and its symptoms - and suicidal thoughts, through increases in adolescents' self-esteem and social support. In this same direction of analysis, and highlighting that suicide represented the third leading cause of death for youth 15 to 24 years in the United States, in 2007, Taliaferro et al. (2008) also aim to test if sports practice may be a protective factor against this adolescent suicide. Using data from 2005[12] and modeling logistic regressions for both sexes, controlling for age and race, the authors analyze the relative risk of hopelessness and suicidality with regards to PAP and sports participation.

Their findings showed that sports participation protected against hopelessness and suicidality for both males and females. However, with regards to the individual's level of physical activity, while frequent and vigorous activity reduced the risk of suicidal ideation for boys, low levels of activity increased the risk of having hopelessness feeling among girls. Also, the relative risk rates of suicidality and hopelessness were lower for athletes than for non-athletes. Thus, Taliaferro et al. (2008) state that the social support and integration created by sports practice may account for a very significant part of the protection system against suicidality and hopelessness among adolescents.

Explaining the health benefits of PAP for children and adolescents by following this same research line and applying it in Brazil, Coledam et al. (2014) analyze the physical activity practice in an interrelated way to participation in school physical education. In this sense, PAP would be positively related - within the scope of physical education classes - to school attendance.

Whereas the PAP raises the individuals' health stock, since it has positive impacts on the physical and mental health of children and adolescents, in order to analyze the socioeconomic impacts of PAP in adolescents - who do not earn their own income - Fejgin (1994), while making several regression analysis to relate the practice of high school sports in the United States to academic performance and success, and to behavioral

[12]Data from the 2005 Youth Risk Behavior Survey, comprehending adolescents aged 12 to 18, on grades 9 to 12.

discipline of students, observes positive effects of sports participation in grades, in self-knowledge, and in control and search for educational aspirations of these adolescents, as well as a negative effect when it comes to discipline problems.

Yet Ransom and Ransom (2017) find results that contradict those presented by Fejgin (1994). By controlling for selection bias and endogeneity between sports practice and academic performance of young adults/adolescents in the United States, their results point to a statistical non-significance of sports practice on academic performance. Thus, they argue and conclude that sports reveals character, but it does not build it.

Still, Esteban-Cornejo et al. (2014), also trying to understand how the practice of extracurricular physical activities could affect the cognitive performance of Spanish adolescents aged 13 to 18, they run a covariance analysis controlling for age, sex, type of school, active commuting to school and BMI (Body Mass Index) characteristics. Considering as cognitive performance the adolescents' results on the SRA-Test of Educational Ability (which comprehends verbal, numeric and reasoning abilities), Esteban-Cornejo et al. (2014)' findings showed that vigorous extracurricular PAP was positively associated with all cognitive variables, and that individuals who participated in more than one extracurricular physical activity had higher cognitive performance in all variables than the ones who participated in only one. Thus, the participation in multiple and organized extracurricular PAP can positively impact adolescents' cognitive performance (ESTEBAN-CORNEJO et al., 2014).

With a similar goal, Knaus, Lechner and Reimers (2018) sought to investigate the effects of an additional PE class on five realms: cognitive skills, non-cognitive skills (related to the individuals' social behavior), motor skills, PAP out of the school hours and health. Using data from the Motorik-Modul Study, a submodule of the German Health Interview and Examination Survey for Children and Adolescents, and anchored in the institutional construction of the German educational system and in the required number of PE classes established by law, the authors use a Propensity Score Matching (PSM) method (specifically the Inverse Probability Tilting method) to evaluate if there are differences in characteristics regarding the five spheres mentioned above for individuals who have more PE classes (treatment) than others (controls).

Controlling for individual and regional characteristics[13], and also controlling for endogeneity between the number of PE classes taken by students and the variables of interest in the analysis (the outcomes, represented by 49 variables that are encompassed on those five spheres of analysis), the authors instrumentalize the number of actual PE classes taken by students with the number of classes required for that school in that state. Thereby, they estimate an IV-PSM[14] for each one of those 49 outcomes, in order to estimate the Local Average Treatment Effect (LATE) of taking an additional physical education class. That is, they estimate a two-stage PSM, which estimates the Average Treatment Effect (ATE) of the instrument on the variables of interest, and the ATE of the instrument on the actual number of PE classes taken by the individual.

Their results highlight that an additional PE class has positive impacts on math and German grades for both boys and girls (cognitive skills). However, regarding non-cognitive skills, positive effects of the additional PE class are observed on girls' emotional control, while negative effects are observed on the social behavior of boys, regarding fights and disputes. Positive impacts on motor skills and on out of school hours PAP are also found for boys and for girls, but with greater impacts for girls. There were not found any statistically significant result regarding the health variables, since, as reported by the students, PAP on PE classes do not reach the necessary intensity levels to affect the quality of health of individuals (KNAUS; LECHNER; REIMERS, 2018).

The theoretical discussion in which these previous results are grounded, as presented by Snyder (1985), relies on the adolescents' behaviour regarding their need to succeed and to be recognized within their peers - which is highly associated with sports participation (COLEMAN, 1961; WERTS, 1967; SCHAFER; ARMER, 1968; SPADY, 1970) -, and on two different aspects of the academic impact of sports practice: (i) on one hand, the zero-sum theory, discoursed and highlighted by Coleman (1961) and Spady (1970); (ii) On the other hand, the development theory, introduced in Rehberg (1969) and

[13]Knaus, Lechner and Reimers (2018) used as control variables on their analysis: binary variables for grade, school type, sex, being a foreigner, eight income classes, four categories for year of birth, three categories for the parents' schooling, five categories for number of siblings, a binary variable indicating whether the parents were physically active, weight at birth, four categories for community size, educational spending per student, and a binary for East Germany.

[14]Instrumental Variable Propensity Score Matching.

Schafer and Armer (1968) and discussed by Hanks and Eckland (1976) and Rees, Howell and Miracle (1990)[15].

The zero-sum theory, discussed and tested on *The Adolescent Society* by Coleman (1961) is based on the idea that time and energy are limited. In this sense, a students' participation in sports and in athletic activities would divert them from their school work and academic pursuits. Considering that adolescents seek recognition, high status and success on environments within their peers, then Coleman's (1961) zero-sum idea arises from the facts that [16]: (a) athletes do get greater social rewards (popularity, prestige, leadership) than good students; (b) most high-school athletes are boys; and (c) boys value these informal rewards more than grades, or possible long-term gains (going to college, better jobs) (SCHAFER; ARMER, 1968). Therefore, we could expect boys to spend as much time and energy as possible on athletics. If time, energy, and attention are limited resources on students' lives, then their studies must suffer if they invest these resources in athletics, i.e., there is a trade-off between boys' - and adolescents, in general - academic performance and sports participation.

Thus, taking into consideration Coleman's zero-sum theory, it should occur that athletes would not perform as well academically as non-athletes, once the greater the student's participation in sports, the greater the detriment to his/her studies; also, a student's participation in those sports that are given the greatest recognition and attention should harm his/her academic performance more than the minor sports that do not require so much time, or give as great social rewards.

As stated by Spady (1970), corroborating the theory discussed by Coleman (1961), and also assuming that adolescents search for recognition and for high status perception within their peers, which are strongly associated with athletics:

> "Since the most visible and widely accepted form of success-striving is college, educational aspirations become for many a proxy for high status and personal recognition. The extracurriculum serves as a means toward this end both by providing opportunities for success that lie outside the formal academic structure and by helping students to develop attitudes and skills that will bolster those aspirations.The system backfires,

[15]The sociological concept that discourses about the behaviour of the adolescents within their peers is discussed and taken as a foundation on both of these sports/PAP theories here presented, that is, on the zero-sum and on the development theory.

[16]These statements from which the zero-sum theory arises are made after Coleman's (1961) data analysis of questionnaires applied on several US High Schools, regarding students' and their parents' perceptions of the school environment and their academic performances.

> however, when activities such as athletics stimulate students' status perceptions and future goals without providing the skills and orientations requisite for their fulfillment. Since exaggerated status perceptions and educational goals are often accompanied by marginal intelligence and weak academic commitment, the disappearance of the peer status structure and its institutionalized conventions of prestige allocation at the end of the senior year suddenly forces these students in particular to reexamine their resources for achievement." (SPADY, 1970, p. 700).

In this sense, sports participation would not develop leadership, commitment or organization characteristics on students, and so it would only limit even more their available time and energy dedicated to studying and to school work, being detrimental to academic performance. It is also worth mentioning that once Coleman (1961) finds these negative impacts on academic performance for sports participation on US high schools, the author generalizes these results from athletics to other forms of extracurricular activities regulated by the school, so that the participation on any school extracurricular activity or class would divert the students' time, energy and attention from their studies and academic goals.

Yet, regarding the different types of school extracurricular activities, the development theory analyses them separately, so that it were developed models to asses the role and impacts of each one of them on academic performance. Thus, as highlighted by Hanks and Eckland (1976), *"athletic participation and social participation in other extracurricular programs are largely orthogonal to one another and, moreover, have both different antecedents and outcomes"* (HANKS; ECKLAND, 1976, p. 273).

When it comes to athletics participation, the development theory is built upon the idea that participation on schools sports is beneficial to students' academic pursuits and performance, once commitment and organization characteristics, the competitive spirit and the desire to win, which are learned on the field, are carried into the classroom as a desire for better grades, better education and more prestigious occupation (REHBERG, 1969).

According to Schafer and Armer (1968), Rehberg (1969), and Hanks and Eckland (1976), within the framework of the development theory, the participation in sports programs seems to serve two important functions, with regards to students' academic performance: (a) it generates and reinforces educational success goals by

exposing students to a network of social relations, consisting in part of school personnel and achievement-oriented peers, with the immediate benefit of binding these students to the school and to its normative structure; and (b) it facilitates the achievement of such goals by students acquiring the kind of knowledge, interpersonal skills, self-confidence, and other attitudes that store them with resources needed in the longer run to turn goals into effective action.

Still, Werts (1967) speculates that positive effects of participation in sports on academic productivity and performance may occur due to: (a) more lenient grading, because teachers may see them school athletes as special or more deserving; (b) benefit from more help in schoolwork from friends, teachers, and parents; (c) exposure, in the sports subculture, to effort, hard work, persistence, and winning characteristics that spills over into nonathletic activities, such as schoolwork; (d) eligibility for certain sports, which may be conditional to getting good grades, or even the aspiration to get a sports college career, which also may be conditional to great academic performance; and (e) a more efficient and effective use of their limited study time, since the athletes spend a lot of time on practice and in the field.

As for the data analysis designed along with the development theory[17] in order to test if the theory applied to the US schools and colleges, Hanks and Eckland (1976) found that the effect of both school and college sports participation on all academic performance variables - grades, school curriculum, peer and teacher contact, and educational attainment - were weak. Thus, there was no evidence that supported the thesis that either high school or college athletics were detrimental to academic performance - as sustained by Coleman (1961)-, moreover they even were mildly beneficial for boys.

However, when it comes to participation in other extracurricular activities,- recalling that these other activities were separately analysed from sports participation by Hanks and Eckland (1976), other than Coleman (1961) - Hanks and Eckland (1976)'s results showed relatively significant direct and total positive effects on academic performance in both school and college for the participation on other extracurricular activities. Similarly, Werts (1967) also found that students with high grades usually won

[17]Hanks and Eckland (1976) and Rees, Howell and Miracle (1990), while discussing the development theory, also tested it empirically, to verify if it applied for the US high schools and colleges, using econometric and structural models, controlling for individual and family characteristics, and socioeconomic status.

recognition in several of these extracurricular areas, whereas the majority of students with low grades did not show any extracurricular achievement.

Therefore, in the light of this development theory framework, the participation in most extracurricular and sports programs[18] would clearly encourage students compliance to school and enhance their academic performance, doing so in a variety of ways, from the simple need to stand out and to be recognized within their peers, to the development of organizational, effort and commitment characteristics, bringing into the classroom the will to win and to achieve higher academic aspirations, or to the better management of their available time, or even if it is simply keeping them from early dropping out of school before graduation, as found by Schafer and Armer (1968).

Along with the development theory, according to Mushkin (1962) and Becker (2007) health capital discussion, the PAP would also result in positive impacts on students' academic performance. Whereas health can affect the stock of human capital[19], and that PAP entails positive impacts on individuals' health stock, Mushkin (1962) points out that health and education are not only consumption goods that satisfy human wants, but they also are essential ingredients of human welfare.

Thereby, there would be economic gains in preventing and curing sickness, which are, in the academic field, higher productive capacity and productive study time added (MUSHKIN, 1962), since greater health stocks are related to higher productivity and less time treating diseases (GROSSMAN, 1972). Also, less days being sick and treating diseases could result in less absent days from school, which, in turn, may enhance the adolescents academic performance. Still, it is important to note that better academic productivity and results may lead to higher stocks of human capital (MINCER, 1958; 1974; BECKER, 1962), which may turn into higher income earnings in the future of these adolescents, showing how PAP can have long term impacts not only on the individuals physical and mental health, but also on their economic well-being.

As discussed by Ransom and Ransom (2017), it is also worth highlighting that there might be endogeneity between PAP and academic performance, since both of them

[18]Due to the main purpose of this research, we will focus on the impacts of sports and PAP participation, rather than other extracurricular activities.

[19]For a broader discussion on the Human Capital Theory, see Mincer (1958), Mincer (1974), Schultz (1961), Becker (1962), and Mincer (1996).

depend on personal unobserved characteristics (such as motivation, determination and skills), and also on family background characteristics and socioeconomic status.

Finally, taking into consideration all this theoretical background and previous findings, this research intends to analyse the academic impact of out of school hours PAP, for Brazilian teenagers, aged 15 to 17, anchored in both development and zero-sum theories. A priori, we expect PAP to have positive impacts on academic performance - as found by Fejgin (1994), Esteban-Cornejo et al. (2014), and Knaus, Lechner and Reimers (2018)-, since it can enhance the individuals' health stock and the development of commitment and organization characteristics (MUSHKIN, 1962; WERTS, 1967; SCHAFER; ARMER, 1968; REHBERG, 1969).

4.3 Methodological Procedures

4.3.1 Econometric Approach

In order to accomplish the main objective of this analysis, that is to test the impacts of PAP on adolescents' academic performance, we chose to use as our academic performance dependent variable the age-grade gap, defined as the incompatibility between completed total years of study and age of the child/adolescent at the beginning of the school term. Since the data that we have available for use (the 2015 PNAD) do not comprehend any specific grade result, we opted on using this age-grade gap variable, instead of using school frequency, which is biased due to mandatory school enrollment.

According to the Ministry of Education of Brazil (MEC) , a child must enter the first year of elementary school when he/she turns six by March 31^{st} of the referred school year[20] (BRASIL. Ministry of Education, 2006). If the child proceeds through the system continuously (throughout elementary, middle and high school), that is, without repeating a grade or leaving the course, he or she will not exhibit age-grade gap (MACHADO; GONZAGA, 2007). Here, it is important to highlight that, since our available data refers to teenagers aged 15 to 17 on 2015, they should have started primary education around

[20]Since the publication of the Law 11,274, from February 6^{th}, 2006 (BRASIL. Ministry of Education, 2006), and especially since 2010, when the regulation of this law became mandatory to all schools, primary education began to last 9 years, starting for children at the age of six years old. Before that, primary education used to last eight years, regularly starting for kids at the age of seven.

the 2000's - that is, before 2006 - when the mandatory age of enrollment was seven years old. Taking this into consideration, the age-grade gap can be defined as follows:

$$agg_i = (age_i - 7) - school \tag{4.1}$$

In Equation 4.1 above agg_i indicates the presence/absence and the magnitude of age-grade gap, *school* represents the individuals' completed school years, and *age* represents the completed years of age of the individual. In this sense, if $age - 7 = school$, the individual does not present any age-grade gap, and so is proceeding continuously through mandatory education; however, if $age - 7 > school$, then he/she exhibits age-grade gap, meaning that the individual must have been held back in some school year. Still, $age - 7 < school$ means that the individual is advanced in school, mainly due to month of birth reasons[21].

Thus, the analytical model we intend to estimate is exemplified in 4.2:

$$agg_i = \beta_0 + \beta_k X_{ik} + \beta_{k+1} pap_i + \varepsilon_i \tag{4.2}$$

In which:

agg_i is the dependent variable, which represents the presence ($agg_i > 0$) or absence ($agg_i \leq 0$) of age-grade gap for the i-th individual, and also its magnitude;

β_0 is the intercept;

X_i is a vector of explanatory variables that characterize the socioeconomic and familiar conditions of each individual, which might interfere on the individual's academic performance, as suggested by Rehberg (1969), Spady (1970), Hanks and Eckland (1976), Fejgin (1994), Machado and Gonzaga (2007), Ransom and Ransom (2017), and Knaus, Lechner and Reimers (2018). These control variables will be listed and itemized in the following section;

pap_i is a binary variable that indicates whether the individual practices any type of physical activity (1) or not (0); as in Werts (1967), Schafer and Armer (1968),

[21]There are two month-based reasons for this age-grade gap to be negative: (i) kids that are born after March 31^{st} should start school only on the following year, according to the MEC (BRASIL. Ministry of Education, 2006), although there are exceptions - which implies in younger children starting primary education months before completing seven years of age; (ii) since the PNAD questionnaire was applied in the end of September of 2015, the individuals whose birthday was on the following months (October, November and December) had not yet completed another year of age, which could also have caused the age-grade gap to be negative.

and Rehberg (1969) and also as pointed out by Mushkin (1962) and Becker (2007), the *pap* is expected to have a positive impact on the academic performance of adolescents, so as to reduce the individuals' age-grade gap. Thus, we expect the parameter referring to this variable to be negative and statistically significant;

and ε_i is the vector of random errors.

Because our dependent variable is discrete, the model we intend to estimate should reflect the distributional characteristics of this type of data, that is discreteness and nonlinearity (CAMERON; TRIVEDI, 2010; LONG; FREESE, 2006). Still, as we described above, our dependent variable, besides being discrete, it also exhibits negative values - which drives it away from the principles of count data models. Therefore, a few different estimation strategies must be taken into consideration.

We start with a simple Linear Regression estimated through Ordinary Least Squares (OLS), as a way of including in our analysis the negative values of the age-grade gap. OLS methodology is based on the idea of estimating the parameters of the linear conditional mean, minimizing the sum of squared errors, that is, the difference between the estimated and the observed values of the dependent variable (GUJARATI, 2009; CAMERON; TRIVEDI, 2010):

$$E(agg|X, pap) = \beta_0 + \beta_k X_k + \beta_{k+1} pap \tag{4.3}$$

With the sum of the squared errors being:

$$\sum_{i=1}^{N} [agg_i - (\beta_0 + \beta_k X_{ik} + \beta_{k+1} pap_i)]^2 \tag{4.4}$$

So that the OLS estimator of the parameters - which also represents the Marginal Effects of the explanatory variables[22] - can be written, in matrix form, as:

$$\hat{\boldsymbol{\beta}} = (X'X)^{-1} X'agg \tag{4.5}$$

[22]Because the OLS estimates a linear regression - requiring linearity on the parameters - the marginal effects are given directly by the estimates of the slope coefficients of each variable (GUJARATI, 2009; CAMERON; TRIVEDI, 2010), so that:

$$\frac{\partial E(agg|X)}{\partial x_k} = \beta_k$$

.

In which the vector $\hat{\beta}$ contains the estimated coefficients for all the explanatory variables, including the intercept and the *pap*.

However, due to OLS normality and linearity assumptions and to the fact that our data is discrete, the OLS estimation might not present a good fit to the data, once the normal probability distribution does not represent the actual data generating process and does not account for discreteness, and OLS does not account for nonlinearity. To deal with these particular features we turn to count data models.

Count data models are used when the outcome of interest is a nonnegative integer, so as the response variable is discrete. Thereby, to accommodate our data to these models, we replace all the negative values of *agg* by zero, meaning that there is no age-grade gap for these individuals. Another characteristic of this type of data is that they are usually represented by a nonlinear function - that is, the data generating process is nonlinear, which results in nonlinear count regressions (CAMERON; TRIVEDI, 2010). There are some variants of the basic count data model, but here we rely on the most common ones, and those that are more adequate to the data and to the problem we intend to analyse.

The most common starting point for modeling count data is using the Poisson distribution and the Poisson model. In the univariate Poisson distribution, the number of occurrences of the event *agg* - in this case, the number of years of age-grade gap that the individuals exhibit - over a fixed exposure period has the probability mass function as in 4.6:

$$Pr(Y = agg) = \frac{e^{-\mu}\mu^{agg}}{agg!}, agg = 0, 1, 2, ... \tag{4.6}$$

Where μ is the intensity parameter, and also the first two moments of the distribution:

$$\begin{aligned} E(Y) &= \mu \\ Var(Y) &= \mu \end{aligned} \tag{4.7}$$

The definitions on 4.7 shows the Poisson property of mean and variance equality, that is, the equidispersion property of the distribution. Since μ standard parameterization is $\mu = exp(X'\beta)$ to make sure that $\mu > 0$, then from 4.7 the Poisson

model is heteroskedastic by its definition (CAMERON; TRIVEDI, 2010). This intrinsic heteroskedasticity along with the fact that count data are often overdispersed break the primary assumption of equidispersion of the Poisson model. In other words, besides varying across the sample, the (conditional) variance of count data exceeds its (conditional) mean.

To correct these assumption-breaking problems, two strategies can be taken into consideration (LONG; FREESE, 2006; CAMERON; TRIVEDI, 2010): (i) one approach is to maintain the conditional mean assumption ($E(agg|X) = exp(X'\boldsymbol{\beta})$), but to relax the equivariance assumption, estimating a model with robust standard errors - similar to White's heteroskedasticity correction on linear regression models -, which keeps the consistency of the Poisson model estimation, and also corrects for the heteroskedasticity problem; (ii) another approach is to estimate a model that explicitly takes into consideration overdispersion, that is to use the Negative Binomial distribution and the Negative Binomial model.

Overdispersion, that is, additional variability in agg, is often generated due to unobserved heterogeneity. To account for this feature, the Negative Binomial regression model adds a parameter (α) that reflects unobserved heterogeneity among observations. This additional parameter comes from the introduction of multiplicative randomness to the Poisson model, replacing μ for $\mu\nu$, with $\nu \sim Gamma(1, \alpha)$, so that the marginal distribution of agg is a Poisson-Gamma mixture with a closed form (LONG; FREESE, 2006; CAMERON; TRIVEDI, 2010) - that is, the Negative Binomial distribution: $agg \sim NB(\mu, \alpha)$.

From the addition of ν, the Negative Binomial distribution is such that $E(agg|\mu, \alpha) = \mu$ and $Var(agg|\mu, \alpha) = \mu(1 + \alpha\mu)$, hence $Var(agg) > E(agg)$, which allows for overdispersion of the data. The Negative Binomial regression model holds $\mu = exp(X'\beta)$ and leaves α as a constant, being a more general case than the Poisson regression model, and that can be reduced to this last one as $\alpha \to 0$ (CAMERON; TRIVEDI, 2005). Therefore, the probability mass function of the Negative Binomial distribution - denoted by $NB(\mu, \alpha)$ - can be written as in 4.8:

$$Pr(Y = agg|\mu, \alpha) = \frac{\Gamma(\alpha^{-1} + agg)}{\Gamma(\alpha^{-1})\Gamma(agg + 1)} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu}\right)^{\alpha^{-1}} \left(\frac{\mu}{\mu + \alpha^{-1}}\right)^{agg} \tag{4.8}$$

Where $\Gamma(.)$ denotes the gamma integral, that actually narrows to a factorial for an integer argument, as in count data. In equation 4.8, the variance function parameter, (α), enters the probability equation, meaning that the probability distribution over the counts depends upon α, even though the conditional mean does not. Ergo, if the data are indeed overdispersed, then the fitted probability distribution of the Negative Binomial can be quite different from that of the Poisson, even though the conditional mean is similarly specified in both, and so the Negative Binomial model is preferred [23]. Lastly, it is still worth to highlight here that the heteroskedasticity problem can be corrected on the Negative Binomial regression model as on the Poisson and on the Linear Regression models: through the robust estimation of the variance-covariance matrix (CAMERON; TRIVEDI, 2010).

Despite correcting the overdispersion of data by increasing the conditional variance without changing the conditional mean, the Negative Binomial regression model still does not account for the excess of zeros that can exist within a count dataset. To solve this problem, Lambert (1992) introduced the possibility of the Zero-inflated count models, which take into consideration not only overdispersion, but also the excess zeros, by changing the mean structure to allow zeros to be generated by two distinct processes (LONG; FREESE, 2006): a count process, of density $f_2(.)$, and a binary process, with the density of $f_1(.)$.

If the binary process takes on the value of 0, with a probability of $f_1(0)$, then $agg = 0$. If the binary process takes on a value of 1, with a probability of $f_1(1)$, then agg takes on the count values $0, 1, 2, ...$ from the count density $f_2(.)$. This lets zero counts occur in two ways: as a realization of the binary process and as a realization of the count process when the binary random variable takes on a value of 1 (CAMERON; TRIVEDI, 2010). Therefore, the zero-inflated model has a density function as in equation 4.9:

$$f(agg) = \begin{cases} f_1(0) + \{1 - f_1(0)\} f_2(0) & \text{if } agg = 0, \\ \{1 - f_1(0)\} f_2(agg) & \text{if } agg \geq 1 \end{cases} \tag{4.9}$$

[23] If one's goal is only to model the conditional mean - and not the probability distribution of the data -, then the Negative Binomial and the Poisson models present quite similar results (LONG; FREESE, 2006; CAMERON; TRIVEDI, 2010)

The probabilities $f_1(.)$ can be parameterized through a binomial model, like the logit or probit, which are regressions such as:

$$E[agg_i|X] = 0[1 - f_1(X'\beta)] + 1[f_1(X'\beta)] = f_1(X'\beta) \tag{4.10}$$

Where $f_1(X, \beta)$ is the cumulative density function, so that the set of parameters β reflects the impact of changes in X, the explanatory variables, on the individual's probability of exhibiting age-grade gap (GREENE, 2003). It is worth to mention that these binomial models (logit and probit) are chosen to be used here due to the categorical and probabilistic characteristics of the dependent variable, on this first part of the zero-inflated models. Yet, on the second part, the probabilities of each count (including zeros), $f_2(.)$, can be estimated by either a Poisson or a Negative Binomial regression, as previously presented on this section[24].

All these count data models - Poisson, Negative Binomial and Zero-Inflated - will be estimated through the Maximum Likelihood (ML) method, given the discrete, probabilistic and non-linear characteristics in the distribution of the dependent variable. ML is an iterative method that seeks to maximize the logarithmized likelihood function. In the ML estimation, even if we do not assume normality on the error distribution, the estimators exhibit the desired properties of consistency, asymptotic normality and asymptotic efficiency, provided that the estimated model represents the real data generating process and that the sample is large enough[25] (ALMEIDA, 2012).

Still, it should be noted that, under the presence of heteroskedasticity, there is no guarantee of consistency of the ML estimators (LIN; LEE, 2005). Thus, if that is the case, then White's covariance matrix must be used as a weighting factor to calculate the variance of the ML estimators (that is, the robust estimator), along with the weight of the observations [26], once we are dealing with a stratified sample (GREENE, 2003). This heteroskedasticity correction method will be used in all four regression models so far presented on this section, that is: the Linear Regression (even though it is estimated through OLS, which also presupposes homoskedasticity, making the correction necessary

[24]The density function $f_2(.)$ also depends on a set X of variables, which does not necessarily need to be the same as in $f_1(.)$.

[25]In this case, the method is called *Quasi* Maximum Likelihood (QML).

[26]Probability weight of individuals, calculated by IBGE, and present in the PNAD microdata file.

so that the estimators remain efficient and t-tests and z-tests remain valid), the Poisson model, the Negative Binomial regression model and the Zero-Inflated model.

Even though all these models deal with several specific characteristics of count data, there is still one particular trait of our dataset that might require another type of data treatment: because we are dealing with a sample of adolescents and we intend to analyse their academic performance, we should take into consideration if they are attending school, otherwise our estimation of the *agg* would be biased by their last school year attended. Therefore, the *agg* would only truly represent the age-grade gap of the individuals if they were still in the school system, that is, if they were attending school by the time of the questionnaire application[27].

This could be simply solved by removing from the sample the individuals that dropped out of school, so that we would end up analysing PAP impacts on academic performance only for adolescents that were attending school by the time of the application of the PNAD questionnaire. However, if the AGG determination depends on the individual choice of dropping-out of or staying-in school, then a sample selection problem could arise. If that's the case - what we will test for -, then the *agg* sample would not be randomly selected, leading to biased parameters estimation.

To correct this sample selection bias we use Heckman (1979)'s two-part estimation technique, which is further discussed on the previous Chapter 3. The method suggests the estimation of two-stage equations. In the first stage, a probit model of qualitative choice for school attendance is estimated (ML), in which the dependent variable assumes the value 1 if the individual is regularly attending school, and 0 in case it has dropped off. From the estimation of this selection equation, it is possible to predict λ_i, inverse of the Mills ratio, which represents the probability that an observation is selected in the sample, that is, the probability of the dependent variable assuming the value of 1. In this case, it represents the probability that the adolescent is regularly attending school. Values of λ_i can be predicted as described in 4.11:

$$\lambda_i = \frac{f(Z_i)}{1 - F(Z_i)} \tag{4.11}$$

[27] Another point of view that would justify these AGG estimations with the whole sample, not only for the individuals that were attending school is that, because they are adolescents, they dropping out of school would result in an age-grade gap, even if they did not continue through the school system, since they remain in school age.

Where $f(Z_i)$ represents the probability density function of Z_i, and $F(Z_i)$ is its cumulative distribution function. Z_i follows a standardized normal distribution, defined from the parameters estimated in the probit model. Ergo, λ depends on the parameters estimated on the selection equation.

The inverse of the Mills ratio is calculated for each observation in the sample and is later added as an explanatory variable in the estimation of the academic performance equation, characterizing the second stage of Heckman's method. It is important to note that the use of this method is only valid if the two equations - the selection and the AGG equation - are dependent on one another[28]. If so, then the estimation of the AGG equation would be inconsistent if the variable λ_i was omitted. Thus, this second stage AGG equation, corrected for selection bias using the method proposed by Heckman (1979), can be estimated as shown in 4.12, similar to equation 4.2:

$$agg_i = \beta_0 + \beta_k X_{ik} + \beta_{k+1} pap_i + \gamma\lambda_i + \varepsilon_i \tag{4.12}$$

Another possible misspecification effect, which breaks down the modeling assumptions and would still lead the estimators to be biased and inconsistent, is endogeneity. As underlined on the previous section, since both PAP and academic performance depend on personal unobserved characteristics (such as motivation, determination and skills), and on family background characteristics, it is possible that we find endogeneity between them. If so, it will be necessary to instrumentalize the endogenous variables, estimating the models via instrumental variables (IV), using the Generalized Method of Moments (GMM), which is consistent under heteroskedasticity (GREENE, 2003).

In terms of the instruments, it is also important to highlight that, in order for them to be strong instruments, they should be highly correlated with the endogenous variables but not correlated with errors (CAMERON; TRIVEDI, 2005). Thus, the models would be estimated holding equation 4.13 - which describes the behavior of the endogenous variable (*pap*) -, showed as follows:

$$pap_i = z\delta_1 + \eta \tag{4.13}$$

[28]If the two equations are not dependent, then this technique proposed by Heckman should not be used, since there wouldn't be any sample selection bias, and we should return to the estimation of the other count data models.

Where z represents the instruments vector, which will be illustrated on the next section, and η is the non-observed random errors.

Finally, since we are interested in analyzing the impacts of PAP on academic performance, we will both interpret parameters and estimate marginal effects for this explanatory variable. Because the dataset and the dependent variable are the same for all estimation strategies here presented, we expect *pap* marginal effects on *agg* to be negative and statistically significant, according to the Development Theory (WERTS, 1967; REHBERG, 1969) and as found by Esteban-Cornejo et al. (2014) and Knaus, Lechner and Reimers (2018), once PAP can enhance the individuals' health stock and the development of commitment and organization characteristics.

In this sense, for the Linear Regression and for the Heckman model with sample selection the parameters already represent the marginal effects of the regressor, because of their method of estimation. Still, we will also estimate semielasticities for these models, so that we can analyse PAP impacts on a proportional/percentage form.

For the Poisson and the Negative Binomial models, the parameters and marginal effects are interpreted in the same way for both models, because they have the same conditional means. Thus, the coefficients can be interpreted as semielasticities, that is, a coefficient of 0.030 means that a unit change on the explanatory variable can be associated with a 3.0% increase in the dependent variable. Yet, the marginal effects of a unit change in a continuous regressor, x_j, equals:

$$\frac{\partial E(agg|X)}{\partial x_k} = \beta_k exp(X'\beta) \tag{4.14}$$

.

Which depends on the evaluation point, X - usually, marginal effects are estimated at sample means (CAMERON; TRIVEDI, 2010). A marginal effect of 0.030 can be interpreted as a unit change on the regressor being associated to a 0.030 increase on the dependent variable, on its unit of measure.

For the Zero-Inflated regression model, the parameters do not have any direct interpretation on the final effect of the explanatory variable, because the variables are used as regressors on both parts of the model. Thus, we can separately interpret the coefficients of each generating process: for the count data process, the parameters represent semielasticities, as in the Poisson and on the Negative Binomial model; for the

binary process, parameters can be interpreted as in the logit/probit models, as impacts on probabilities of "certain zeros" success. Also due to this feature, marginal effects of a given explanatory variable are complicated to estimate, since they have to compute the effects of both parts of the model - but still, they have a much more straightforward and direct interpretation, and will be analysed.

As in a complementary analysis, we will also estimate and analyse the PAP marginal effect for the binary school attendance equation, which is given by:

$$\frac{\partial E[f\_school|X]}{\partial X} = \left\{ \frac{dF(X'\beta)}{d(X'\beta)} \right\} \beta = f(X'\beta)\beta \tag{4.15}$$

Where $f(.)$ is the density function that corresponds to the cumulative distribution, $F(.)$, for the probit binary model. In this model, we expect *pap* marginal effects on school attendance to be positive and statistically significant, once we expect the PAP to reduce school drop out, leading to positive impacts on the individuals' probability of regularly attending school.

To sum it up, we will estimate several models, testing them to check which one fits best our data, so that we can evaluate the impacts of PAP on the adolescents' AGG. The models that will be estimated are: (1) a Linear Regression (LR), holding negative values of the *agg*; (2) starting the count data models, with *agg* ranging only on positive values, a Poisson Regression Model (PRM); (3) a Negative Binomial Regression Model (NBRM); (4) a Zero-Inflated Poisson model (ZIP); (5) a Zero-Inflated Negative Binomial model (ZINB); and lastly, (6) a Heckman estimation technique, that accounts for sample selection - in which we will also analyse the PAP impacts on the school attendance Probit estimation, as for a complementary analysis.

All these models will be tested for heteroskedascity and endogeneity, and corrected - as described above - if needed. For a sensitivity analysis, we will estimate them by gender, and also taking into consideration in our sample only the adolescents that were attending school by the time in which the questionnaire was applied.

4.3.2 Sample, Data Treatment and Variables

To accomplish our goal of analysing PAP impacts on adolescents' AGG it would be required of the data to contain information regarding sports and physical

activity practice. Hence, we will use microdata from the 2015 PNAD, an annual sample survey planned and conducted by IBGE, at household level, in order to produce results that cover the entire Country. These results include data about general population characteristics, education, labor market, income and housing (IBGE, 2016), alongside with a supplementary questionnaire, which subject-matter was PAP on the 2015 edition, suiting the objectives of this research quite well.

Since we are interested on the academic performance as a dependent variable (e.g., the age-grade gap), we will keep in our sample only the individuals aged 15 to 17 [29] that answered the PNAD supplementary questionnaire. As previously stated, we will run the models for samples both considering adolescents that had already dropped out of school, and only for the ones that were regularly attending school in 2015. Thereby, the control and characteristics variables we will use in this analysis, as suggested by Rehberg (1969), Hanks and Eckland (1976), Fejgin (1994), Machado and Gonzaga (2007), and Ransom and Ransom (2017), are displayed on Table 4.1 below:

Table 4.1: Variables Construction and Description

Variable	**Description**
agg	The discrete dependent variable, which shows if the individual exhibits age-grade gap ($agg > 0$), or if he does not ($agg \leq 0$). If the individual is advanced in school, then $agg < 0$.
pap	A binary variable that indicates whether the individual practices some type of sports or physical activity out of the school hours (1) or not (0).
sex	A binary variable for the individual's sex, which assumes the value 1 for women.
color/race	Four binary variables to distinguish the color/race declared by the individual: white (base category), indigenous, yellow, brown and black.
chores hours	A continuous variable, representing the number of hours spent on domestic work, such as chores.

[29] Originally, we intended to analyze individuals aged 11 to 17, but the PNAD supplementary questionnaire only covers individuals from the age of 15. This happens probably due to individual assent and independent individual decisions of practicing any kind of sports or physical activity.

working hours	A continuous variable indicating the number of worked hours on the week of reference, i.e., in which the questionnaire was applied.
mom_household	A binary variable that receives the value 1 if the mother is present in the individual's household.
householder_school	A variable representing the complete schooling of the householder.
dependent kids	A variable indicating the number of kids younger than 10 year old in the adolescent's family.
ln_income	A continuous variable, which is the natural logarithm of the monthly per capita family income.
rural	A binary variable indicating whether the individual's household is located in a rural (1) or urban (0) area.
metropolitan	A binary variable indicating if the individual's household is located in a metropolitan region (1) or if it isn't (0).
macro_5	Four binary variables to distinguish the five Macro-Regions of Brazil: North, Northeast, Central West, South and Southeast (taken as base).

Source: Elaborate by the author.

For that first dependent variable, *pap*, those individuals who claimed they practiced sports or physical activities that did not require much body movement and energy expenditure were classified as non-practitioners of physical activity (that is, $pap = 0$). Those sports/physical activities were: fishing, bowling, pool, billiards, card games, dice games, chess, checkers, other sports with motor use and other cards sports. Also, because they were practiced by very few people, we also classified as non-practitioners of physical activity those who declared they practiced the following sports: motoring, motorcycling, *enduro karting*, motocross, rally and powerboat.

Moreover, this *pap* variable encompasses the practice of physical activities that occurs outside physical education classes, since these classes are mandatory. In other words, we do not consider the time spent on physical education classes as PAP, because it is not our goal to analyse the impact of these classes on the AGG - but yes of the out of schools hours PAP.

For the school attendance equation, we also considered as explanatory variables the *agg*, a binary variable - *job_week* - indicating whether the adolescent had worked on the week of reference (1), and a discrete variable indicating the size of the family of the adolescent (fam_size), along with the other variables used on the AGG equation, as previously expatiated.

Due to the variables we chose to use, and to the way they were constructed, some observations had to be dropped, so that they wouldn't exhibit misleading values. Hence, we dropped from the sample those individuals who were classified as "undetermined" on the *"years of study"* category (whose information on schooling was not declared or did not allow for the IBGE *years of study* classification), the individuals with undeclared per capita family income, and also the individuals to which there were not declared information regarding their parents/householder schooling.

Thus, after these treatments, our sample ended up with $3,402$ observations of 15 to 17 years old adolescents, whereas $2,865$ (84.2% of the sample) were regularly attending school by the time the PNAD questionnaires were applied. These sample sizes correspond to a population of $10,392,265$ and $8,818,146$ (84.85% of the referred population) respectively, when expanded by the frequency weights calculated by the IBGE. However, it is important to state that these frequency weights were not used on the model estimations, they were only used to estimate the population size correspondent to our final sample. As stated on the previous section, we used probability weights - instead of frequency weights - on our estimations, in order to correct them for heteroskedasticity and to suit our sample to the Brazilian population, according to the IBGE population estimations.

With respect to endogeneity, if its presence is verified [30], then we will need to instrumentalize the *pap* variable. Since for the instruments to be strong they have to be highly correlated with the endogenous variable, but not correlated with the errors (CAMERON; TRIVEDI, 2005; GUJARATI, 2009), we build them so that they are related to the PAP but not to the AGG. In order to do so, we construct instruments that represent the individuals' behaviour, preferences and perceptions regarding health, lifestyle and socialization, which are strong determinants of their decisions over practicing physical

[30] Through Hausman's endogeneity test.

activities (WERTS, 1967; SCHAFER; ARMER, 1968; SNYDER, 1985; KING et al., 1992).

Thus, we rely on the reasons why the adolescents practiced (or did not practice) physical activities or sports (health, figure, available time, socialization), and also regarding their preferences on government investment and expenditure regarding sports incentives. These instruments are described on Table 4.2.

Once we only aim to instrumentalize the *pap* variable - the only one within our set of explanatory variables that would have theoretical reasons to exhibit an endogenous behaviour regarding *agg* (RANSOM; RANSOM, 2017) -, then we would end up with six instruments to an unique endogenous covariate, so that the models we intend to estimate would be overidentified (GREENE, 2003).

Table 4.2: Instrumental Variables Construction and Description

IV	**Description**
public_neigh_space	A continuous variable representing the average of positive answers to the existence of public areas dedicated to sports or physical activity practice, by PSU (primary sampling unit). Since it is a categorical answer (*yes* or *no*), this variable is comprehended between 0 and 1, and gives an idea of how the neighborhoods are propitious/friendly to PAP.
ipublic_pap	Indicates if the individual thinks there should be public investment on sports or physical activities development on his/her neighborhood.
health	Indicates if the individual has chosen to practice or not to practice sports or physical activities because of health reasons. That is, if the individual cares about his/her health when deciding to practice - or not to practice - physical activities.
no_time	Indicates if the motive for the individual not to practice - or quit practicing - physical activities/sports was lack of time.

others_social	Indicates if the individual has chosen to practice, not to practice or quit practicing sports or physical activities due to having fun/socialization reasons (or lack of will/socialization issues).
sports_before	A binary variable that indicates if the individual has practiced any sports or physical activity before the researched period - that is, before a year previously to the application of the PNAD questionnaires.

Source: Elaborate by the author.

Finally, with all the econometric procedures and data treatment already described and expatiate, we can now follow through to the results analysis and discussion, regarding the PAP impacts on adolescents' AGG.

4.4 Results and Discussion

4.4.1 Descriptive Statistics Analysis

Because we intend to analyse the adolescents' behaviour regarding both PAP and AGG, it is important to start with a descriptive statistics analysis of our sample, so that we can have some insights on these referred behaviours and on the adolescents average characteristics. As stated on the previous section, after filtering the data, we ended up with a sample of $3,402$ observations of adolescents aged 15 to 17. Of these, $1,817$ - 53.4% - stated they practiced any sports or physical activities out of the school hours, while $1,585$ - which corresponds to 46.6% of the sample - were not engaged in any of these two.[31]

For those who answered they did participate in any type of sports or physical activity, the most commonly practiced one was soccer, corresponding to over 56% of these PAP adolescents, most likely because it is the most popular sport in the country. As we can see on Table 4.3, there is a great figure difference between this first most practiced

[31]These sample proportions and those results presented on Table 4.3 were quite similar to the ones estimated for the Brazilian population (exhibiting differences only on the third decimal place), by weighing the observations with the frequency weights calculated by the IBGE.

sport and the other activities, since soccer is directly followed by walking - the most simple and with less requirements physical activity listed -, which is practiced by around 10% of the PAP adolescents in the sample.

Table 4.3: Types and Practice Proportion of Sports and Physical Activities Among Adolescents

	n	**Proportion**
Soccer	1,023	56.3%
Walking	192	10.56%
Fitness Activities	151	8.31%
Volleyball	129	7.10%
Martial Arts	92	5.06%
Bodybuilding	80	4.40%
Other	64	3.52%
Bike Riding	45	2.47%
Swimming	43	2.37%
Dancing	42	2.31%
Running	28	1.54%
Skate	25	1.37%
Cycling	25	1.37%
Handball	24	1.32%
Basketball	23	1.26%
Gymnastics	15	0.82%
Athletics	8	0.44%
Racquet Sports	5	0.27%
Sports with Animals	3	0.16%
Water Sports	2	0.11%
Adventure/Nature	2	0.11%
Parasports	0	0%

Source: Elaborate by the author, from PNAD dataset.

Differently from the adults, adolescents tend to practice collective activities, that involve playing with friends and engaging in socialization, such as soccer and volleyball, and also swimming, dancing, and martial arts, which are usually taken in collective classes. It is likewise notable from Table 4.3 that the sports that are not so popular in Brazil and that have very specific equipment requirements - such as sports with animals and racquet sports - are quite unrepresentative in the sample. As an extreme, we can see that the engagement in parasports is not representative at all.

Still, it is important to highlight that on Table 4.3 the sum of these sports and physical activities practice proportions exceeds 100%, since multiple answers to multiple sports are allowed, in cases which the individual practiced more than one sport or physical activity. Hence, there are $2,021$ positive answers to sports and PAP, for $1,817$ observations

of PAP adolescents, meaning that around 11.2% of these adolescents practiced more than one sport or physical activity.

With regards to the personal, socioeconomic and family characteristics of the adolescents, that is, the control variables we intend to use on our estimations, we present their means on Table 4.4, along with group means for those PAP and NON PAP adolescents. Also, to get some insight on how the variables behave between these two groups - and to initially check for differences between them -, since we want to understand how the PAP can affect the adolescents' academic performance, we carry out and present t Tests of mean differences for these control variables between the PAP and NON PAP adolescents in our sample. The results are also displayed on Table 4.4.

As we can see from Table 4.4, almost all the control variables showed statistically significant mean differences, except for the color/race variables and the adolescents' residence location binary variables, that is, the rural and metropolitan area dummies. From that, we can state that there are not statistical differences for the color/race of the individuals and their location of residence between the PAP and NON PAP groups, so that these variables may not be determinant for the individual decision of practicing sports/physical activities or not.

The number of weekly worked hours and the number of dependent kids on the household also did not exhibit statistically significant mean differences for the two groups analysed. Only 18.8% of the adolescents stated they worked on the week of reference, and the distribution of these worked hours are quite similar for the observations on both PAP and NON PAP groups. Thus, there should be no mean difference with regards to this characteristic. The same behaviour appears on the number of dependent kids feature: similar distribution between the two groups, with even smaller dispersion - so that the observations are concentrated on zero and one (around 91%), reaching a maximum value of five dependent kids on the same household.

On the contrary of what we see on the working hours variable, the number of hours spent on chores differs from one group to another: the average of hours spent on chores by the adolescents on the PAP group is statistically smaller than the average chores hours for the adolescents that did not practice sports or physical activities. Still, the proportion of individuals that stated they did not spend any time on chores was substantially smaller when compared to this same proportion on the worked hours variable

Table 4.4: Descriptive Statistics and t Test Results of Mean Differences over PAP

		(1) NON PAP		(2) PAP		
Variable	**Sample Mean**	**Mean**	**SD**	**Mean**	**SD**	**Mean Diff**
agg	1.261	1.326	1.997	1.204	1.726	0.122*
agg_p	1.384	1.452	2.117	1.325	1.853	0.126**
sex	0.491	0.671	0.469	0.334	0.471	0.337***
indigenous	0.006	0.005	0.075	0.007	0.084	-0.001
yellow	0.003	0.002	0.050	0.003	0.052	-0.000
brown	0.522	0.522	0.499	0.522	0.499	0.000
black	0.086	0.087	0.282	0.085	0.279	0.002
chores hours	9.034	10.67	11.01	7.638	9.324	2.996***
working hours	5.191	4.926	12.18	5.421	12.44	-0.494
mom_ household	0.784	0.746	0.435	0.817	0.386	-0.070***
householder_school	7.304	6.923	4.324	7.636	4.497	-0.713***
dependent kids	0.471	0.491	0.755	0.454	0.756	0.036
ln_income	6.034	5.909	1.237	6.143	1.094	-0.023***
metropolitan	0.354	0.367	0.482	0.342	0.474	0.025
rural	0.161	0.154	0.154	0.167	0.373	-0.012
public_neigh_space	0.133	0.132	0.039	0.135	0.041	-0.003**
ipublic_pap	0.787	0.702	0.457	0.862	0.344	-0.160***
health	0.211	0.048	0.215	0.353	0.478	-0.304***
no time	0.125	0.209	0.407	0.051	0.221	0.157***
others_social	0.701	0.650	0.476	0.745	0.435	-0.095***
sports_before	0.124	0.208	0.406	0.051	0.221	0.157***
attending school	0.842	0.789	0.407	0.888	0.315	-0.098***

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

(a) The number of observations N is $1,585(46.6\%)$ for the individuals that do not practice physical activities, and $1,817(53.4\%)$ for the practitioners, on a total sample of $3,402$ observations.

(b) t Test null hypothesis: $NONPAP - PAP = 0$.

(c) The variable *agg* comprehends both negative and positive values for the age-grade gap, while the variable *agg_p* considers only positive values - to account for the dependent variable of the count data models -, replacing the negative values by zero.

Source: Elaborate by the author, from PNAD dataset.

distribution, being around 30% of the observations of the whole sample. Thus, the more free time available, the easy it is for the adolescents to engage on sports or PAP. This same relationship can be interpreted from the *no time* variable: the average of individuals who stated that lack of time was one of the reasons they did not practice sports or physical activities is larger among those on the NON PAP group.[32]

[32]This *no time* variable, as well as the *sports_before* variable, are derived from PNAD questions that exist only for the individuals that did not practice sports, but might as well have practiced physical activities. Therefore, although there might be positive answers in both PAP and NON PAP groups, it is straightforward that the proportion of successes is larger on the NON PAP category.

Regarding the other reasons why the adolescents might choose to engage or not to engage on PAP, the results also show that the proportion of adolescents that care about their health and that think of sports and PAP as a fun way of socialization is larger in the PAP group, in opposition to the individuals that do not participate on PAP due to heath or socialization issues. However, when it comes to the previous to the research sports practice, the average of individuals that had practiced sports before is greater within those who are NON PAP - as stated above, this may be due to the fact that the questions used to construct this variable are available only for the individuals that did not practice sports, but might as well have practiced physical activities. Therefore, it should follow that the proportion of successes with regards to this variable to be larger on the NON PAP group.

Still when it comes to adolescents' preferences regarding PAP, the average of teenagers that think there should be public investment on PAP development on their neighborhood is statistically - and substantially - larger for the PAP group, once PAP must be an important part of their life. On the other hand, the lack of public spaces dedicated to and that enable different sorts of PAP may be detrimental to the practice of these activities, once there are significantly less areas like these on the neighborhoods where the NON PAP teenagers of our sample live.

As we expected, boys are more inclined to participate on sports and physical activities than girls, so that the proportion of boys (girls) on the PAP group is statistically greater (smaller) than on the NON PAP group. Vaitsman (1994) deeply discusses this result as a development of certain social systems, within which differences and hierarchies between certain social categories - in this case, the gender categories - are built. In this sense, since PAP is such a different custom for boys and girls, then it is possible for its impacts on AGG to be also different between sexes.

Once again looking at the availability of time and at the family support to the adolescents, the proportion of individuals whose mother lives in their same household is greater for those who practice physical activities than for those who do not. In this same sense, the average per capita income is also greater for those adolescents on the PAP group, exceeding the average per capita income of the NON PAP group by 30.4%, which can mean that a family with greater financial conditions can provide a better support and incentive to sports and physical activities practice (WERTS, 1967), especially because

some of these sports are not publicly offered and available in Brazil, and might require some specific equipment for training.

Still with regards to the family influence and support to PAP, one variable that might as well be very important to the academic performance of the adolescents is the schooling of the householder, insofar as the higher the schooling of the householder, the higher should be the incentive and the family influence for the adolescent to stay in and do well in school. In terms of the results from Table 4.4, the schooling of the householder is, on average, higher for those families in which the adolescent is engaged on PAP. This can also be seen through Mushkin (1962) and Becker (2007)'s perspective of health as being part of and positively affecting the individuals' stock of human capital.

Finally checking for PAP and NON PAP differences on our dependent variable, i.e. AGG, the t test results show that the average age-grade gap is smaller on the PAP group[33], meaning that there is statistical difference on the academic performance of those adolescents that participate on physical activities and sports out of the school hours, and those who do not. Since the AGG mean is greater within the NON PAP group, this previous result may preliminarily indicate that our hypothesis of negative PAP impacts on AGG is corroborated, accordingly to the Development Theory. As for the school attendance, as we expected, the proportion of individuals regularly attending school is higher within the PAP group, which also may be due to time availability and family support.

Still regarding the school attendance, it is interesting to note that the older the adolescents, the higher the percentage of school drop out. For the 15 years old, nearly 8% of these adolescents did not attend school, while for the 16 years old the percentage of drops out reached the figure of 14%. Even more worryingly, almost 26% of the adolescents that were 17 years old did not regularly attend school by the time of the questionnaires application. This increase on the percentage of school drop out with the increase on the adolescents' age makes sense as the older they get, the easier it is for them to enter the labor market and become a relevant income source in the family, since regular labor becomes allowed according to the Brazilian law at the age of 16, and partial time apprentice labor is allowed from the age of 14 (BRASIL. Constitution (1988), 1998).

[33]The variable agg was statistically significant at a significance level of 10% (p-value of 0.07), while the variable agg_p exhibited statistical significance at a 5% level (p-value = 0.04).

Thus, this primary mean differences investigation and data description analysed on this section imply that there might be negative PAP impacts on adolescents' AGG, which we intend to econometrically test for causality on the following section.

4.4.2 Econometric Estimations Results

Because we estimated a series of different models that could fit our data and its generating process, we start the results analysis by running some adjustment tests in order to check for evidences of which model best fits the data. Therefore, we run tests for the presence of endogeneity, tests of overidentification and model specification, tests of equation independence (for school attendance and AGG), we also test, in the count data estimations, if the data distribution is more similar to the Poisson or to the Negative Binomial, and if its generating process differs among the counts (i.e., is zero-inflated). The results of these tests are displayed on Table 4.5.

Under the null hypothesis of exogeneity, we can see from Hayashi (2000) GMM C statistic that there is no evidence, in any subsample, of endogeneity between the variables PAP and AGG. Since Stata [34] can only perform this test for linear regressions, we test endogeneity under these estimations for both variables *agg* and *agg_p*, that is, respectively taking into consideration the negative values of AGG, and considering the dependent variable of the count data models.

Because the main problem with estimating linear regressions for count data dependent variables is prediction, we still rely on the results of Hayashi (2000)'s test, but we also run NB and ZINB regressions taking as an explanatory variable the predicted error term from the regression of the PAP - the allegedly endogenous covariate - on the other independent variables, and also the instruments (WOOLDRIDGE, 2010, p. 665).

That is, we perform a Hausman endogeneity test, testing for the significance of this error term - η. If it is statistically significant on the AGG regressions, then there is information about AGG being held on PAP and vice-versa, that is, they are endogenous.

[34] All the estimations and statistical tests of this research are run on software Stata 14 (STATACORP, 2015).

Table 4.5: Summary Tests and Statistics of Model Fitting

Test	**Whole Sample**			**Attending School**		
	All	Boys	Girls	All	Boys	Girls
Endogeneity Test (GMM C)	0.2547	0.2203	0.6076	0.0091	0.6546	0.1017
Overridentification Test (Hansen's J)	15.86***	11.82**	10.67*	17.15***	9.228	10.40*
Endogeneity Test - Count Data	0.4021	0.3698	0.6543	0.0075	0.9343	0.1924
Overridentification Test - Count Data	15.57***	10.53*	11.04*	17.66***	8.913	10.75*
Overridentification Test - Poisson	16.00***	10.28*	12.35**	18.33***	9.023	11.12**
Likelihood-Ratio Test of $\alpha = 0$	919.88***	442.86***	451.82***	532.99***	241.17***	284.59***
Vuong NB ZINB Test	4.87***	4.04***	3.77***	4.06***	3.80***	2.99**
Heckman Wald Test of Independent Eq.	0.81	208.0***	0.01	-	-	-
Heckman Wald Test - NB Count Data	0.36	-	-	-	-	-

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

We did not present Heckman *agg_p* Wald test results for boys and for girls because the estimation of the models did not converge.

Source: Elaborate by the author.

These *"by-hand"* Hausman tests results, as well as the linear regressions Hayashi (2000)'s tests, show that there is very little evidence of endogeneity between AGG and PAP, since the null hypothesis of exogeneity is not rejected in almost all subsamples.[35] Therefore, it will not be needed to instrumentalize the PAP explanatory variable on our estimates.

When it comes to model specification, almost all the instrumentalized equations came back as overidentified, either for the *agg*, the *agg_p* or the *agg* on the Poisson regression model. The Null hypothesis of this overidentification test is that the instrument set is valid and that the model is correctly specified. Thus, by rejecting the null hypothesis it is possible to say that the IV models are not correctly specified, pointing to the same conclusion as on the endogeneity tests, that is, for the non existence of endogeneity.

Regarding the distribution that could best fit our data, the likelihood-ratio test indicates that in none of the subsamples the variance is equal to the subsample mean, and, thus, the generating process that has the best fitting is the Negative Binomial distribution, instead of the Poisson distribution.

For checking if the data generating process is different when *agg* assumes zero, that is, if the data is zero-inflated, we first take a look at the proportion of zeros in each of our subsamples, as seen on Table 4.6 below:

Table 4.6: Proportion of Zeros in Each Subsample

	Whole Sample			**Attending School**		
	All	Boys	Girls	All	Boys	Girls
agg	32.42%	28.63%	36.34%	37.28%	32.82%	42.07%
agg_p	43.59%	38.11%	49.25%	46.11%	40.43%	52.21%

Source: Elaborate by the author, from PNAD dataset.

First, it is worth noting that because the variable *agg_p* replaces the negative values of *agg* by zero, then it makes sense that the proportion of zeros on that first variable is greater. In general, we can see that the proportion of girls that do not exhibit AGG is greater than the proportion of boys with zero years of age-grade gap. Among those attending school, the proportion of individuals with zero AGG is greater than on the full sample, which considers the adolescents that were not attending school. However, as seen

[35]The only subsamples in which the error term was statistically significant - at a 10% level - was the NBRM estimation for all the adolescents attending school, and for the boys who attend school. Still, all these tests results are presented on Appendix C.

on Table 4.6, the proportion of zeros within our sample is smaller than 50%, in almost all subsamples, not making it clear whether the sample is zero-inflated or not. To account for that, we analyse Vuong (1989)'s closeness test of model selection.

The statistic tests the null hypothesis that the two models are equally close to the true data generating process, against the alternative that one model is closer. It cannot make any decision whether the "closer" model is the true model, so that the rejection of the null hypothesis simply implies that either the ZINB model fits the data better than the NB model, or vice-versa, not that zero-inflation is present or absent (WILSON, 2015). Still, from the results on Table 4.5 it is possible to infer that there are differences on data fitting between ZINB and NBRM regression models, although we can not tell which one produces better adjustment to the true data generating process.

Even though we present and analyse Vuong's test, it is important to highlight, as discussed by Wilson (2015), that this test is not adequate to non-nested models[36], such as ZINB and NBRM. Wilson (2015) argues that because ZINB reduces to NBRM only when the parameter of the binomial model is fixed at zero, the distribution of the log-likelihood ratios would be non-normal, leading to non-normality of the sampling distribution of the zero-inflation parameters, which, in turn causes misleading test results, once Vuong's statistic is developed under normality assumptions. Thereby, model selection will be done through information criteria and prediction errors analysis.

Lastly with regards to model fitting, when it comes to sample selection, we test for equation dependence between AGG and school attendance. From Wald test results on Table 4.5 we observe that the estimation of the AGG would be biased by the decision of dropping-out of school only for boys, but not for the girls. In this case, we also estimate Heckman NBRM-ZINB models for this subsample.

Therefore, after and according to all these adjustment tests, the models we estimate and present their results are: (1) a Linear Regression (LR), holding negative values of agg; (2) a Negative Binomial Regression Model (NBRM); (3) a Zero-Inflated Negative Binomial model (ZINB); (4) a Heckman-NBRM estimation method, for the boys subsample; and (5) a Heckman-ZINB estimation method, also for the boys subsample. The PAP coefficients on the AGG equations are shown on Table 4.8. The coefficients and marginal effects of the other explanatory variables employed on these models are displayed

[36]Davison (2003) defines nested models as: *"two models are said to be nested if one reduces to the other when certain parameters are fixed."*.

on Appendix C. Still, it is worth to remember that all these models are estimated holding robust standard errors to correct for heteroskedasticity, and expanding the sample for the Brazilian population, weighting the estimates by the probability weights calculated by the IBGE.

However, before we move on to the actual results analysis, there are still some adjustment statistics that are worth highlighting. As previously stated on this section, to check if either the ZINB or the NBRM estimation best fits our data, since Vuong's test is not reliable for non-nested models, we construe some model fitting and prediction statistics, as displayed on Table 4.7.

Table 4.7: Models Goodness-of-fit Criteria

	Whole Sample			**Attending School**		
	All	Boys	Girls	All	Boys	Girls
(1) LR						
AIC	13,602.31	7,034.07	6,523.62	10,734.10	5,630.44	5,082.39
BIC	13,718.82	7,132.27	6,621.22	10,847.35	5,725.88	5,176.54
R^2	0.143	0.165	0.125	0.129	0.155	0.107
(2) NBRM						
AIC	10,507.41	5,708.44	4,788.21	8,223.47	4,575.82	3,645.36
BIC	10,630.05	5,817.54	4,891.23	8,342.68	4,676.57	3,744.74
Pseudo R^2	0.050	0.052	0.048	0.049	0.048	0.042
Predicted Count Error	0.059	0.074	0.046	0.039	0.069	0.015
(3) ZINB						
AIC	10,453.56	5,689.33	4,770.47	8,201.13	4,564.65	3,650.22
BIC	10,692.71	5,902.09	4,971.10	8,433.58	4,760.85	3,843.75
Predicted Count Error	0.037	0.046	0.030	0.035	0.048	0.038

Boys whole sample NBRM and ZINB models are Heckman's sample selection estimations.
Source: Elaborate by the author.

Because the dependent variable differs from LR to NBRM and ZINB, their goodness-of-fit statistics are not directly comparable. Thereby, from (1), we can see that model fitting from information criteria (BIC and AIC) are slightly better on the attending school subsample, although explainability (R^2) is greater for boys, and on the whole sample. Still, as boys and girls take PAP and academic performance as such different customs and habits, goodness-of fit is better when estimating separate models between sexes.

With regards to NBRM (2) and ZINB (3), Table 4.7 shows that, when it comes to information criteria, while AIC points to a slight better fitting of the ZINB models[37], the

[37]Except for the girls attending school subsample.

Schwarz criterion (BIC) tends to the opposite way: indicating that the NBRM estimations best fits our AGG count data. In spite of this adjustment "tie" between AIC and BIC, the predicted count error throughout the probability mass function is smaller on the ZINB estimations, meaning that the AGG prediction is closer to its actual value when using the ZINB model estimations, instead of using the NBRM. The only subsample in which the count error is smaller for NBRM model is the girls attending school, which can also be seen from both AIC and BIC criteria.

From these model fitting analysis, we now turn to the actual models PAP results, taking into consideration both NBRM and ZINB estimations, but keeping in mind that the ZINB results are more accurate. Still, we compare these count data results to the LR estimation, which considers negative values of AGG.

Table 4.8: Models Estimation Results: PAP Coefficients on AGG Equations

	Whole Sample			**Attending School**		
PAP	All	Boys	Girls	All	Boys	Girls
(1) LR	-0.049	0.051	-0.192*	0.023	0.155	-0.156
(2) NBRM	-0.049	0.046	-0.168*	0.011	0.128	-0.145
(3) ZINB - NB	-0.146**	-0.024	-0.371***	-0.054	0.117	-0.313
(3) ZINB - Inflate	-0.621**	-0.380	-1.52***	-0.475	-0.066	-1.92
(4) Hekman - NBRM	-	-0.012	-	-	-	-
(5) Heckman - ZINB - NB	-	0.033	-	-	-	-
(5) Heckman - ZINB - Inflate	-	-0.360	-	-	-	-
N	3,402	1,729	1,673	2,865	1,484	1,381

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Source: Elaborate by the author.

As displayed on Table 4.8, for both LR and NBRM models, the PAP coefficients were statistically zero on almost all the subsamples - except for the girls on the full sample, in which the referred coefficients were negative at a 10% significance level - while for the ZINB some results were statistically significant: for the full sample, and for the girls on the full sample, the practice of physical activities or sports out of the school hours exhibited negative impacts on their AGG.

That is, corroborating our initial hypothesis, and according to the Development Theory (REHBERG, 1969; SCHAFER; ARMER, 1968; HANKS; ECKLAND, 1976; REES; HOWELL; MIRACLE, 1990), the fact of participating on physical activities or sports reduces the girls' age-grade gap by 37% - or by a factor of 0.69 -, holding all other variables in the model constant, for those who might have some

AGG (those who are not "certain zeros"). Then, in the light of this theory, the perseverance, commitment and organization characteristics developed on the PAP field (REHBERG, 1969; REES; HOWELL; MIRACLE, 1990), and the increase in the health stock also brought forth by the PAP (MUSHKIN, 1962; BECKER, 2007), would overflow to the academic field, enhancing the adolescents' - in this case, the girls' - academic performance.

However, as a result of the inflate equation, the fact that a girl practices physical activities or sports reduces the probability of her not displaying AGG, that is, reduces the odds of her being a "certain zero" by a factor of 0.21[38]. In this sense, the girls' participation on out-of-school hours PAP would reduce her available time and energy, being detrimental to her academic performance (COLEMAN, 1961; SPADY, 1970) and reducing the odds of her being a "certain zero" on the matters of AGG, that is, not exhibiting age-grade gap.

The same interpretation can be made for the full sample ZINB results: for those individuals who might present AGG, the fact of practicing physical activities or sports reduces the age-grade gap by 14.6%, or by a factor of 0.86, holding constant all the other variables; still, being part of the PAP group reduces the individuals' odds of certainly not presenting AGG by a factor of 0.54. These both-sided PAP impacts show the complexity of the relationship between AGG and PAP, meaning that both health stock, commitment and organization enhancement and time and energy expenditure caused by the PAP can affect the individuals' academic performance.

In order to evaluate the magnitude of PAP impacts on the AGG, we also present PAP Marginal Effects (ME) and semielasticities, as follows on Table 4.9. Marginal effects represent the total impact of an unit change on the explanatory variable PAP on the AGG, in terms of years - the measurement unit of AGG -, so that for the ZINB models, in which we have PAP as an explanatory variable on both equations, the MEs account for the PAP effects of both generating processes. Semielasticities are similar to MEs, with the only difference that semielasticities represent the percentage change in AGG due to an unit change on PAP. It is important to note that, because the NBRM and ZINB estimations are not linear, MEs and semielasticities are conditional to the values of the explanatory variables, varying across them. Thus, the results on Table 4.9 refer

[38]Both factor changes for unit increase from ZINB-NBRM and ZINB-Inflate calculated as $exp(\beta_{k+1})$.

to MEs and semielasticities at the means of the explanatory variables employed on the estimations.

Table 4.9: PAP Marginal Effects and Semielasticities of Selected Models, at Means of Explanatory Variables.

	Whole Sample			Attending School		
PAP	All	Boys	Girls	All	Boys	Girls
(1) LR						
ME	-0.049	0.051	-0.192*	0.023	0.155	-0.156
Semielasticity	-0.041	0.036	-0.201*	0.022	0.129	-0.186
(2) NBRM						
ME	-0.056	-0.015	-0.155*	0.010	0.138*	-0.118
Semielasticity	-0.049	-0.011	-0.167*	0.011	0.127	-0.144
(3) ZINB						
ME	-0.090	0.124	-0.281**	0.002	0.159	-0.302
Semielasticity	-0.072	0.086	-0.248**	0.000	0.118	-0.311

Boys whole sample NBRM and ZINB models are Heckman's sample selection estimations.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Source: Elaborate by the author.

As we can see on Table 4.9, at the sample mean values of the explanatory variables, there were statistically significant magnitude impacts of PAP on AGG on the model estimations for the girls subsample. For this group, the fact of participating on out-of-school hours PAP would reduce the girls' age-grade gap by 16.7% (2) to 24.8% (3) 24.8% (3), or by 0.155 (2) to 0.281 (3) years, that is, from around 1.9 to 3.4 months, on average.[39]

However, when it comes to PAP impacts on AGG for boys, the majority of the results were not statistically different from zero. Therefore, there would be no impacts of sports or physical activities on boys' academic performance. The only exception was the NBRM boys attending school estimation, in which *pap* displayed a positive and significant impact on AGG. In this sense, the participation of boys on out-of-school hours PAP would corroborate Coleman (1961)'s zero-sum theory, so that it would increase by at around 1.6 months (0.138 years) their age-grade gap, on average.

Similarly to what Knaus, Lechner and Reimers (2018) found, PAP impacts on academic performance are quite different for boys and girls. As discussed by Vaitsman (1994), sports and physical activities are institutions and social interactions that emulate the social hierarchy prevailing in other spheres of the society. By emulating this societal

[39]It is worth to highlight that ZINB estimation was the method that presented the best fit to our data.

behaviour, PAP ends up with a sort of "sexual division of sports": in which boys are much more inclined to PAP than girls, and that, in turn, can lead to much different impacts of the PAP on academic performance.

This hierarchical difference can be clearly seen within our sample: for very similar boys and girls subsamples (respectively, of 1,729 and 1,673 observations), almost 70% of the boys stated they practiced any sort of sports or physical activities, while only 36.3% of the girls were engaged on PAP. Although the PAP frequencies and the time spent on each PAP session are very similar for both sexes[40] - for boys, PAP frequencies are at around 2.2 times a week, lasting one hour and ten minutes each, on average; for girls, the frequency of PAP is a little smaller, 2.15 times a week, on average, lasting around one hour and five minutes -, sports and PAP are much more intrinsic to the routine and behaviour of boys than for the girls', which makes it a lot more common for boys to engage on PAP, and that can lead to quite different impacts of PAP on AGG between the sexes.

Therefore, considering that PAP is much more intrinsic to the behaviour of the boys than the girls', we have that the impact of PAP on boys' academic performance can be positive - that is, participating on PAP increases boys' AGG -, resembling the zero-sum theory described in Coleman (1961) insofar as engaging on PAP would reduce the availability of time and energy of the boys, since those two are finite resources. Still, according to Knaus, Lechner and Reimers (2018)'s results, more time spent on PAP would negatively affect the social behaviour of boys, with regards to fights and competitions.

On the other hand, because sports and PAP are not the mainstream for girls according to the social hierarchy in which they are inserted (VAITSMAN, 1994), participating in PAP and getting physically and mentally involved in sports would bring benefits to the girls' mental and physical health (BABISS; GANGWISCH, 2009; EKELUND et al., 2012), and would develop their cognitive and organizational skills, as discussed on Rehberg (1969) and Rees, Howell and Miracle (1990)'s development theory, entailing negative impacts on their AGG. Alongside this negative impacts on AGG found in our estimations - that is, engaging on PAP would reduce the girls' age-grade gap, as seen on Table 4.9 -, Knaus, Lechner and Reimers (2018) similarly conclude that additional PAP (as in additional PE classes) also helps on the emotional control of the girls.

[40] According to the data in our PNAD sample of adolescents.

Because PAP effects are so different for boys and girls, then this should entail the full sample estimations not to be statistically significant. In this case, and as found by Ransom and Ransom (2017), sports and physical activities practiced out of the school hours are non-significant in terms of affecting the adolescents' academic performance, in general. From that, sports and PAP would not build character, but just reveal it (RANSOM; RANSOM, 2017).

Still in analysing the results among our subsamples, when it comes to the differences between all the adolescents and those who were regularly attending school, we can see from Tables 4.8 and 4.9 that, while some PAP results were negative and statistically significant on the full sample estimations, the majority of the coefficients and the MEs were not different from zero on the attending school subsamples[41], except for the boys attending school subsample - which NBRM estimation result point to a positive impact of PAP on AGG.[42] That is, for the boys' full sample, PAP was not relevant in terms of causing AGG, while for those boys attending school, PAP could have significative and adverse impacts on their academic performance, in terms of raising their age-grade gap.

This may occur because of a smaller AGG variability among the attending school adolescents, which results on non-significative PAP coefficients. However, including in the sample the adolescents that dropped out of school is justifiable because they remain in school age, so they not continuing through the school system would simply result in an age-grade gap that is different from zero. Still, these not attending school adolescents correspond to only 15.7% of the full sample.

Thus, the results from both samples are relevant and analytically significant, since they can have similar, yet somewhat different, interpretations: when considering exclusively the adolescents attending school, we account for the actual AGG of the students that were still enrolled on the school system, but we also end up with less variability and diversity of family and socioeconomic characteristics of the individuals; on the other hand, if we include on our estimations the adolescents that dropped out of school, along with greater characteristics variability, we may end up with some AGG

[41] That is, while for the attending school adolescents the PAP was not significant in terms of affecting their academic performance, when we take into consideration the individuals that dropped out of school, the PAP becomes relevant in terms of reducing their AGG.

[42] On the other hand, while looking at the boys' subsample, the PAP was not significant in terms of affecting their AGG, unless we do not take into consideration the boys who dropped out of school. In this last case, the PAP was relevant in terms of raising the attending school boys' AGG.

overestimations, once we would consider the last school year attended by the adolescent as their schooling - which is also valid because they are still on school-going age.

In this sense, PAP could not only affect the adolescents' academic performance, but also their probability of staying-in or dropping-out of school. In order to check if such relationships can be statistically significant, we also estimate school attendance binary equations, taking within the explanatory variables both PAP and AGG.[43] The coefficients and MEs[44] from these probit estimations are displayed on the following Table 4.10.

As we can observe, engaging on sports and PAP enhances the adolescents' probability of attending school, for all boys and girls. On average, the fact of participating on PAP enhances the individual probability of attending school by 6.7 percentage points. This figure is greater for boys - 8.6 percentage points -, and smaller for girls - 4.6 percentage points. Thus, PAP not only increases the adolescents' academic performance, but also their probability to attend school.

With regards to the age-grade gap, quite interesting results - that are also very different for boys and girls - can be analysed: in general, one additional year of AGG reduces the adolescents' probability of attending school by 2.02 percentage points, on average. That is, the higher the age-grade gap, the easier it is for the adolescent to drop out of school, once their school delay only increases with each year they get held back. For boys, this AGG impact is even greater, on the order of 2.6 percentage points, mainly because they face great incentive[45] from their family and from the society, in general, to enter the labor market and drop out of school. At the same time, for the girls, the age-grade gap impact is smaller when it comes to affecting the individual probability of attending school: one additional AGG year reduces the girls' probability of attending school by 1.1 percentage points, on average.

To sum it up, our count data estimations pointed out to results that were quite different for boys and girls: while for the boys the PAP impacts on AGG were mostly non-significant, being even positive for those boys who regularly attended school, for the girls, engaging on sports or PAP would reduce their age-grade gap on around 3.4 months, on average, when taking into consideration all the adolescents in our sample, not just the

[43]It is important to highlight here that on the first step of Heckman's sample selection estimation we did not use AGG as an explanatory variable, since it is the dependent variable on the second stage of estimation.

[44]MEs estimated at sample means.

[45]Boys tend to face greater incentive to enter the labor market than girls, who usually are designated to do the domestic work (VAITSMAN, 1994; KINGTON; SMITH, 1997).

Table 4.10: School Attendance Binary Estimation Results: Coefficients and Marginal Effects.

	Whole Sample		Boys		Girls	
School Attendance	**Coefficient**	**ME**	**Coefficient**	**ME**	**Coefficient**	**ME**
pap	0.3160***	0.0668***	0.3920***	0.0862***	0.2300**	0.0464**
agg	-0.0967***	-0.0202***	-0.1300***	-0.0262***	-0.0547*	-0.0114*
sex	0.0752	0.0157	(.)	(.)	(.)	(.)
chores hours	-0.0136***	-0.0028***	-0.0009	-0.0002	-0.0173***	-0.0036***
indigenous	0.8901*	0.1070***	0.3210	0.0535	1.975**	0.1280***
yellow	(.)	(.)	(.)	(.)	(.)	(.)
brown	-0.0410	-0.0085	-0.0744	-0.0150	-0.0265	-0.0055
black	0.0885	0.0177	0.1680	0.0312	0.0680	0.0137
work	0.2950*	0.0554**	0.4720**	0.0831**	0.1550	0.0301
working hours	-0.0279***	-0.0058***	-0.0304***	-0.0061***	-0.0261***	-0.0054***
mom_household	0.3340***	0.0782***	0.0191	0.0039	0.6140***	0.1550***
householder_school	0.0165*	0.0034*	0.0201	0.0040	0.0143	0.0029
dependent kids	-0.1020*	-0.0213*	-0.0265	-0.0053	-0.2080***	-0.0433***
family size	0.0682**	0.0143**	0.0371	0.0075	0.0962**	0.0200**
ln_income	0.1010***	0.0211***	0.0823*	0.0166*	0.0995***	0.0207***
north	0.1130	0.0224	0.2210	0.0402	0.0003	0.0000
northeast	0.1020	0.0209	0.2330*	0.0446*	-0.0567	-0.0120
south	0.1070	0.0214	0.2930*	0.0520**	-0.0984	-0.0213
central west	0.0553	0.0113	0.0805	0.0156	-0.0317	-0.0067
metropolitan	-0.0086	-0.0018	0.0520	0.0104	-0.1020	-0.0217
rural	0.0138	0.0028	0.0118	0.0023	0.0444	0.0091
N	3,393	3,393	1,723	1,723	1,670	1,670
Pseudo R^2	0.123	0.123	0.130	0.130	0.151	0.151

(a) Coefficients and Marginal Effects.

(b) Discrete change of dummy variable from 0 to 1.

(c) Variable *Yellow* dropped from estimation due to exact determination.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Source: Elaborate by the author.

ones regularly attending school. In addition to reducing the girls' AGG, the PAP also increased the individual probability of attending school for all adolescents, as observed from our school attendance binary estimation findings.

Thus, as stated by Esteban-Cornejo et al. (2014), these findings suggest that educational and public health policies promoting PAP programmes to all adolescents could yield important benefits to enhance adolescents' academic performance and reduce income[46] and sex inequalities, besides increasing the actual physical activity and sports practice.

[46]Since higher schooling increases future income earnings, by enhancing the Human Capital stock (MINCER, 1958; 1974; SCHULTZ, 1961; BECKER, 1962).

4.5 Concluding Remarks

Many are the benefits for both physical and mental health promoted by the practice of sports and physical activities. According to the WHO (WHO, 2018b), by practicing a minimum of 60 daily minutes of those sports or PAP, in an intensity ranging from moderate to high, adolescents can exhibit improvements in cardiorespiratory fitness, muscle strength and endurance, and bone health. It can also help reduce body fat, reduce the risk of cardiorespiratory and metabolic diseases (EKELUND et al., 2012; CHAPUT et al., 2013), and the symptoms of anxiety and depression (BABISS; GANGWISCH, 2009) - including suicidal ideation (TALIAFERRO et al., 2008) -, along with providing the experience of social and self enjoyment (GUTHOLD et al., 2019).

Along with these already very important - and very well studied - health impacts, engaging on PAP can also lead to another range of impacts on the well being of individuals, regarding the social, economic and environmental aspects of their lives. The publication of the WHO global report on PAP (WHO, 2018a) highlights the importance of a more physically active society in terms of health system costs, productivity gains/losses and environmental spillovers. In this sense, since we are interested in analysing the economic impacts of PAP on adolescents, the main goal of this research was to evaluate the effects of engaging on PAP on the academic performance of these teenagers - which could lead to future impacts on their ability to earn income -, in Brazil.

To explain how PAP and sports participation can affect the academic performance of adolescents, two theories that derive from the behaviour of these adolescents within their peers (COLEMAN, 1961; WERTS, 1967; SCHAFER; ARMER, 1968; SPADY, 1970) arise in opposite directions: on one hand, the development theory (REHBERG, 1969; SCHAFER; ARMER, 1968; REES; HOWELL; MIRACLE, 1990), which highlights the positive impacts of PAP, due to its characteristics of socialization, and character and discipline building, with regards to commitment, organizational and academic pursuits characteristics; on the other hand, the zero-sum theory (COLEMAN, 1961; SPADY, 1970), which main ideas point to a trade-off between sports practice and the academic focus of individuals, regarding their time allocation and energy expenditure - which are finite resources. In this sense, the greater the student's participation in sports, the greater the detriment to his/her academic performance.

From that, we rely on the hypothesis that PAP entails positive impacts on academic performance, not only due to that described on the development theory, but also because a higher health stock - promoted by the participation in sports and physical activities - may enhance the individuals' academic productivity (MUSHKIN, 1962; BECKER, 2007), leading to better academic results.

In order to verify if these statements check within the Brazilian reality, we estimated several linear and count data models for the existence and magnitude of age-grade gap, for the adolescents aged 15 to 17 on the 2015 PNAD dataset, taking into consideration the distributional specificities of our data - and mostly, of our dependent variable, AGG - and accounting for zero-inflation. Thus, we ended up analysing the *pap*[47] coefficients and MEs on the AGG equations. Since the greater the age-grade gap, the worst the academic performance of the adolescent, then we expected PAP coefficients to be negative - so that the fact of engaging on PAP would reduce the individuals' AGG -, for the results to match our preliminary hypothesis.

Although most of the result were not statistically significant, as discussed by Vaitsman (1994), the effects of PAP on the academic performance of adolescents were quite different for boys and girls: while, for the girls, the fact of participating on out-of-school hours PAP reduces their age-grade gap by 24.8%, or 0.281 years (around 3.4 months), on average, for the boys, the PAP was, in most cases, statistically zero in terms of affecting AGG. Even more, considering exclusively those boys regularly attending school, the NBRM MEs point out to a positive effect of participating on PAP on their age-grade gap. In this case, sports and PAP could be detrimental to the academic performance of school-going boys, resembling Coleman (1961)'s zero-sum theory discussion.

These results highlight how differently PAP is faced by boys and girls, and how it is a different custom between these two groups, so that it even has different impacts on their academic performance. Vaitsman (1994) argues that this behavior is engendered by the reproduction of other social hierarchies - in this case, sexual hierarchies - on the PAP field, once these PAP sexual differences are developed as part of the social system. Thus, the practice of sports and physical activities is something that is much more common and intrinsic to boys, and boys are way more inclined to engage on PAP than girls.

[47] Our binary variable indicating whether the individuals engaged on any type of PAP (1) or not (0).

In addition to having negative impacts on the AGG, that is, enhancing the academic performance of the girls, we also found that the PAP increases the individuals' probability of attending school, for all boys and girls. Through a binary estimation taking as dependent variable the school attendance,[48] we found that the fact of practicing sports or physical activities out of the school hours increased the probability of attending school of the adolescent by 6.7 percentage points, being this figure greater for boys - 8.6 percentage points -, and smaller for girls - 4.6 percentage points.

For further analysis, it would be interesting to analyse the effects of sports and PAP on the academic performance of adolescents - which could take the form of this same age-grade gap, or even some general grade concept - in terms of the weekly time spent on these activities, so that it is possible to observe the behavior of the adolescents' academic performance throughout each additional hour allocated on PAP. Still, it would be of main importance to consider, on these estimations, the sex differences here highlighted.

Thus, from these findings we can state that engaging on PAP may enhance the adolescents' academic performance, which, in turn, could imply greater future capacity to earn income (MINCER, 1958; 1974; SCHULTZ, 1961). As stated by the WHO (2018a) as part of the 2030 Sustainable Development goals, a more active society not only brings about benefits for the health of individuals, but these benefits overflow to the well-being of the society itself, in terms of environmental spillovers, costs reduction on health care systems, and economic development. However, without additional public effort and incentives to increase PAP, these benefits will not be realized and those Sustainable Development goals will not be met (GUTHOLD et al., 2019). Still as discussed by Guthold et al. (2019), it is not by chance that four in every five adolescents do not experience the enjoyment and social, physical, and mental health benefits of regular PAP, but it is a consequence of political choices and societal design. In this sense, PAP public policies and interventions could be used as tools not only for public health, but also for income and gender inequality reduction, both on the social and on the economic scopes.

[48] A binary variable indicating whether the adolescent was regularly attending school (1) or not (0).

Chapter 5

Final Remarks

Sedentary lifestyle still is a problematic issue both worldwide and in Brazil. According to the WHO (2018b), nearly 81% of school-going teenagers (aged 11 to 17 years old) and 23% of adults (over 18 years) were insufficiently active, worldwide, in 2010. In Brazil, the PAP reality is quite similar: data from the 2013 PNS (IBGE, 2014) show that, one in two adults does not practice the minimum level of physical activity recommended by the World Health Organization (WHO). As argued by Guthold et al. (2019), the fact that these proportions of individuals do not experience the enjoyment and social, physical, and mental health benefits of regular physical activity is not by chance, but a consequence of political choices and societal design.

Even though PAP can have multiple positive impacts on physical, mental and perceived health (WHO, 2018a; WESSEL et al., 2004; BABISS; GANGWISCH, 2009; MEYER; CASTRO-SCHILO; AGUILAR-GAXIOLA, 2014), insufficient physical activity is the fourth of the 10 leading risk factors for global mortality (WHO, 2018b). Engaging on regular PAP reduces the risk of cardiorespiratory and metabolic diseases, diabetes, breast and colon cancer, hypertension and depression, as well as acting as an important component in the individual energy balance, in order to control weight. It also improves cardiorespiratory fitness, muscle strength and endurance, reducing body fat, and reducing the symptoms of anxiety and depression (WHO, 2018b).

Taking into consideration these physical and mental health benefits from the PAP, the main objective of this research was to analyse the PAP as a socioeconomic problem, in that it can affect the individuals' productivity, leading to positive returns in terms of income earning - for the adults - and better academic performance - for

the adolescents. For the adults, we rely on Mincer (1958), Mushkin (1962), Grossman (1972), and Becker (2007) human and health capital theories. This theoretical framework considers that, once PAP enhances individuals' health stock, then it can affect their productivity by two means: (i) with a higher health stock, individuals lose less working days, and also are more productive, once they feel better (MUSHKIN, 1962; GROSSMAN, 1972); (ii) an elevated health stock also enhances individuals' life expectancy, which makes their investment in human capital more profitable. Hence, a higher stock of human capital leads to higher earnings (MINCER, 1958; SCHULTZ, 1961; BECKER, 2007).

When it comes to academic productivity regarding PAP, two theories that derive from the behaviour of the adolescents within their peers (COLEMAN, 1961; WERTS, 1967; SCHAFER; ARMER, 1968; SPADY, 1970) arise in opposite directions: on one hand, the development theory (REHBERG, 1969; SCHAFER; ARMER, 1968; REES; HOWELL; MIRACLE, 1990), which highlights the positive impacts of PAP, due to its characteristics of socialization, and character and discipline building, with regards to commitment, organizational and academic pursuits characteristics; on the other hand, the zero-sum theory (COLEMAN, 1961; SPADY, 1970), which main ideas point to a trade-off between sports practice and the academic focus of individuals, regarding their time allocation and energy expenditure - which are finite resources. In this sense, the greater the student's participation in sports, the greater the detriment to his/her academic performance.

From that, we rely on the hypothesis that PAP positively affects individuals' productivity for both adolescents and adults, entailing positive impacts on their income earnings (for the adults), and negative impacts on the adolescents' age-grade gap.

With regards to behavioural and sociodemographic characteristics that can stimulate or withdraw the PAP, King et al. (1992), Cousins (2014), and Powell, Slater and Chaloupka (2004) discuss that not only individuals' features and behaviours (such as gender, color/race, income, occupation, time spent on domestic chores, personal skills, socialization, lack of time or encouragement), but also the availability of physical activities environmental settings can affect the individual decision of practicing any kind of physical activity out of the working/school hours.

On this field, our mean differences tests using 2015 PNAD data corroborated these theoretical propositions, once the results of almost all variables analysed showed

statistical differences between individuals that practiced physical activities and those who did not. In general, as we expected, men/boys are more likely to engage in physical activities than women/girls, individuals that practice physical activities tend to earn higher income, spend less time on domestic chores, have less dependent kids and live on neighbourhoods with more physical environmental settings than those who do not. As for the teenagers' academic performance, a higher average AGG is found among those adolescents who do not participate on out of school-hours PAP, which may preliminary indicate that our hypothesis of positive PAP impacts on the adolescents academic performance is also corroborated.

Yet the results with respect to PAP impacts on income, after the estimation of an income equation through the sample selection method proposed by Heckman (1979), controlling for sociodemographic characteristics and for endogeneity[1], for the same 2015 PNAD sample, show that PAP coefficients were positive and statistically significant, as we expected *a priori*. Thus, we can remark that the practice of any kind of physical activity out of the work hours increase the individual's income earnings by 10.2% (IV-Probit IV-GMM) to 12.3% (BP IV-GMM), on average. Therefore, the higher the accumulation of health capital (MUSHKIN, 1962) through PAP and the higher the investment in human capital due to higher health stocks (BECKER, 2007), the higher is the individuals' income earnings.

By separately analysing these PAP impacts between sexes and among age groups, our results are quite different throughout these subsamples: PAP displays greater impacts on men's income earnings, and an inverted U-shaped behavior across age groups, similar to the very impacts of age and experience on income, as discussed by Mincer (1958) and Schultz (1961).

Similarly to the adults, the PAP impacts on the adolescents' academic performance were much different for boys and girls. Through the estimation of several count data models for the existence and magnitude of age-grade gap, for the adolescents aged 15 to 17 on that same 2015 PNAD dataset[2], taking into consideration the distributional specificities of our data, accounting for zero-inflation, we found that, for the girls, the fact of participating on out of-school hours PAP reduces their age-grade

[1]As suggested by Oliveira and Justus (2017), there could be endogeneity between PAP and income. Hence, we instrumentalized PAP on our estimations.

[2]In which the supplementary questionnaire regards PAP and sports practice habits.

gap by 24.8%, or 0.281 years (around 3.4 months), on average. However, for the boys, the PAP was, in most cases, statistically zero in terms of affecting AGG. Even more, considering exclusively those boys regularly attending school, the NBRM MEs point out to a positive effect of participating on PAP on their age-grade gap.

In this sense, while for the girls, engaging on PAP may enhance their academic productivity, by enhancing their health stock (MUSHKIN, 1962) and by developing organizational, commitment and pursuits characteristics (REHBERG, 1969; SCHAFER; ARMER, 1968), when it comes to the boys - particularly those boys regularly attending school -, sports and PAP could be detrimental to their academic performance, resembling the time and energy trade-off discussed on Coleman (1961)'s zero-sum theory.

Therefore, our results show that PAP can have positive impacts on the well being of individuals regarding the social, economic and environmental aspects of their lives[3], alongside with the already much discussed positive impacts on physical and mental health - for both adults and teenagers-, and hence, it should be treated as an economic problem. Failing to recognize that PAP - or PAP insufficiency - is an issue of socioeconomic matter would lead to physical inactivity rising costs, in terms of further negative impacts on health systems, the environment, economic development, community well-being and quality of life for all (WHO, 2018a).

In terms of further research, it would be interesting to analyse the effects of sports and PAP on the income earning of adults and on the academic performance of adolescents in terms of the weekly time spent by the individual on these activities, so that it is possible to observe the behavior of these variables (income and academic performance) throughout each additional hour allocated on PAP. Still, it would be of main importance to consider, on these estimations, the gender and age differences here highlighted.

Furthermore, since PAP is much differently faced by boys/men and girls/women, and taking into consideration that, as presented by the results from this research, engaging on PAP increases the adults' income earnings, and reduces the teenage girls' age-grade gap - enhancing their academic performance, which, in turn may increase their future ability to earn income (MINCER, 1958; SCHULTZ, 1961; BECKER, 1962) -, then policy interventions regarding PAP and public health should

[3]Entailing positive impacts on the individuals' productivity, either academic or labor.

address socioeconomic statuses disparities. In this sense, PAP public policies and interventions can also be used as tools for public health and income inequality reduction.

References

ALMEIDA, E. Econometria espacial. **Campinas–SP: Alınea**, 2012.

ANGRIST, J. D.; PISCHKE, J.-S. **Mostly harmless econometrics: An empiricist's companion**. Cambridge, MA: Princeton university press, 2008.

AQUINO, E. M. L. d. Gênero, trabalho e hipertensão arterial: um estudo de trabalhadoras de enfermagem em Salvador, Bahia, 1996.

ARCAS, P. H. **Implicações da progressão continuada e do SARESP na avaliação escolar: tensões, dilemas e tendências**. 2009. PhD thesis – Universidade de São Paulo. Address: <https://teses.usp.br/teses/disponiveis/48/48134/tde-12032010-110212/en.php>.

BABISS, L. A.; GANGWISCH, J. E. Sports participation as a protective factor against depression and suicidal ideation in adolescents as mediated by self-esteem and social support. **Journal of Developmental & Behavioral Pediatrics**, LWW, vol. 30, no. 5, pp. 376–384, 2009.

BECKER, G. S. Health as human capital: synthesis and extensions. **Oxford Economic Papers**, Oxford University Press, vol. 59, no. 3, pp. 379–410, 2007.

— Investment in human capital: A theoretical analysis. **Journal of political economy**, The University of Chicago Press, vol. 70, 5, Part 2, pp. 9–49, 1962.

BERTAGNA, R. H. Progressão continuada: limites e possibilidades. [sn], 2003. Address: <http://repositorio.unicamp.br/bitstream/REPOSIP/252350/1/Bertagna_RegianeHelena_D.pdf>.

BRASIL. **LAW 9.394/1996 from December 20^{th}, 1996**. Planalto. Executive Power, Civil House - Legal Sub-Office. Brasília, DF, 1996. Estabelece as diretrizes e bases da educação nacional.

BRASIL. Constitution (1988). **Constitutional Amendment N. 20, from November 15^{th}, 1998.** Brasília, DF, 1998. Modifica o sistema de previdência social, estabelece normas de transição e dá outras providências. Address: <http://www.planalto.gov.br/ccivil_03/Constituicao/Emendas/Emc/emc20.htm#art1>.

BRASIL. Ministry of Education. **LAW 11.274/2006 from February 6^{th}, 2006**. Planalto. Executive Power, Civil House - Legal Sub-Office. Brasília, DF, 2006. Altera a redação dos arts. 29, 30, 32 e 87 da Lei no 9.394, de 20 de dezembro de 1996, que estabelece as diretrizes e bases da educação nacional, dispondo sobre a duração de 9 (nove) anos para o ensino fundamental, com matrícula obrigatória a partir dos 6 (seis) anos de idade.

CAMERON, A. C.; TRIVEDI, P. K. **Microeconometrics: methods and applications**. Cambridge: Cambridge University Press, 2005.

CAMERON, A. C.; TRIVEDI, P. K. **Microeconometrics using stata**. College Station: Stata press College Station, TX, 2010. vol. 2.

CHAPUT, J.-P. et al. Combined associations between moderate to vigorous physical activity and sedentary behaviour with cardiometabolic risk factors in children. **Applied Physiology, Nutrition, and Metabolism**, NRC Research Press, vol. 38, no. 5, pp. 477–483, 2013. Address: <https://www.nrcresearchpress.com/doi/abs/10.1139/apnm-2012-0382@apnm-vs0103#.Xa26kpJKjIU>.

CHEKROUD, S. R. et al. Association between physical exercise and mental health in 1· 2 million individuals in the USA between 2011 and 2015: a cross-sectional study. **The Lancet Psychiatry**, Elsevier, vol. 5, no. 9, pp. 739–746, 2018. Address: <https://www.thelancet.com/journals/lanpsy/article/PIIS2215-0366(18)30227-X/fulltext#seccestitle10>.

CHIBURIS, R. C.; DAS, J.; LOKSHIN, M., et al. **A practical comparison of the bivariate probit and linear IV estimators**. [**sineloco**], 2011. Address: <https://openknowledge.worldbank.org/bitstream/handle/10986/3368/WPS5601.pdf?sequence=1>.

COLEDAM, D. H. C. et al. Prática esportiva e participação nas aulas de educação física: fatores associados em estudantes de Londrina, Paraná, Brasil. **Cadernos de Saúde Pública**, SciELO Public Health, vol. 30, pp. 533–545, 2014. Address: <http://www.scielo.br/pdf/csp/v30n3/0102-311X-csp-30-3-0533.pdf>.

COLEMAN, J. **The adolescent society: the social life of the teenager and its impact on education.** Oxford: Free Press of Glencoe, 1961. Address: <http://psycnet.apa.org/record/1961-35078-000>.

COUSINS, S. O. **Exercise, aging and health: Overcoming barriers to an active old age**. New York: Taylor & Francis, 2014.

CRESPO, C. J. et al. Race/ethnicity, social class and their relation to physical inactivity during leisure time: results from the Third National Health and Nutrition Examination Survey, 1988–1994. **American journal of preventive medicine**, Elsevier, vol. 18, no. 1, pp. 46–53, 2000. Address: <https://ajph.aphapublications.org/doi/pdf/10.2105/AJPH.87.5.805>.

DAVISON, A. C. **Statistical models**. Cambridge, MA: Cambridge University Press, 2003. vol. 11.

DING, D. et al. The economic burden of physical inactivity: a global analysis of major non-communicable diseases. **The Lancet**, Elsevier, vol. 388, no. 10051, pp. 1311–1324, 2016. Address: <https://www.sciencedirect.com/science/article/pii/S014067361630383X>.

EKELUND, U. et al. Moderate to vigorous physical activity and sedentary time and cardiometabolic risk factors in children and adolescents. **Jama**, American Medical Association, vol. 307, no. 7, pp. 704–712, 2012. Address: <https://jamanetwork.com/journals/jama/articlepdf/1104986/joc25008_704_712.pdf>.

ESTEBAN-CORNEJO, I. et al. Characteristics of extracurricular physical activity and cognitive performance in adolescents. The AVENA study. **Journal of sports sciences**, Routledge, vol. 32, no. 17, pp. 1596–1603, 2014. Address: <https://shapeamerica.tandfonline.com/doi/full/10.1080/02640414.2014.910607?scroll=top&needAccess=true#.Xa3DtpJKjIU>.

FEJGIN, N. Participation in high school competitive sports: A subversion of school mission or contribution to academic goals? **Sociology of sport journal**, vol. 11, no. 3, pp. 211–230, 1994. Address: <https://journals.humankinetics.com/doi/10.1123/ssj.11.3.211>.

FLORINDO, A. A. et al. Prática de atividades fisicas e fatores associados em adultos, Brasil, 2006. **Revista de Saúde Pública**, SciELO Public Health, vol. 43, pp. 65–73, 2009. Address: <https://www.scielosp.org/scielo.php?script=sci_arttext&pid=S0034-89102009000900009>.

FREEDMAN, D. A.; SEKHON, J. S. Endogeneity in probit response models. **Political Analysis**, Cambridge University Press, vol. 18, no. 2, pp. 138–150, 2010. Address: <https://www.jstor.org/stable/25792001?read-now=1&seq=13#page_scan_tab_contents>.

GREENE, W. H. **Econometric analysis**. London: Pearson Education, 2003.

GROSSMAN, M. On the concept of health capital and the demand for health. **Journal of Political economy**, The University of Chicago Press, vol. 80, no. 2, pp. 223–255, 1972.

GUJARATI, D. N. **Basic econometrics**. Boston: Tata McGraw-Hill Education, 2009.

GUTHOLD, R. et al. Global trends in insufficient physical activity among adolescents: a pooled analysis of 298 population-based surveys with 1· 6 million participants. **The Lancet Child & Adolescent Health**, Elsevier, 2019. Address: <https://www.thelancet.com/journals/lanchi/article/PIIS2352-4642(19)30323-2/fulltext>.

HANKS, M. P.; ECKLAND, B. K. Athletics and social participation in the educational attainment process. **Sociology of Education**, JSTOR, pp. 271–294, 1976.

HAYASHI, F. Econometrics. 2000. **Princeton University Press. Section**, vol. 1, pp. 60–69, 2000.

HECKMAN, J. J. Sample selection bias as a specification error. **Econometrica: Journal of the econometric society**, JSTOR, pp. 153–161, 1979.

IBGE. Pesquisa Nacional de Saúde–PNS 2013: percepção do estado de saúde, estilos de vida e doenças crônicas. **Rio de Janeiro: IBGE**, 2014. Address: <https://biblioteca.ibge.gov.br/visualizacao/livros/liv91110.pdf>.

IBGE, P. Pesquisa Nacional por Amostra de Domicılios–2016. **Rio de Janeiro**, 2016.

KING, A. C. et al. Determinants of physical activity and interventions in adults. **Medicine & science in sports & exercise**, Lippincott Williams & Wilkins, 1992.

KINGTON, R. S.; SMITH, J. P. Socioeconomic status and racial and ethnic differences in functional status associated with chronic diseases. **American journal of public health**, American Public Health Association, vol. 87, no. 5, pp. 805–810, 1997.

KNAUS, M.; LECHNER, M.; REIMERS, A. For Better or Worse? The Effects of Physical Education on Child Development. IZA Discussion Paper, 2018. Address: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3111142>.

KNUTH, A. G. et al. Prática de atividade fısica e sedentarismo em brasileiros: resultados da Pesquisa Nacional por Amostra de Domicılios (PNAD) 2008. **Ciência & Saúde Coletiva**, SciELO Public Health, vol. 16, pp. 3697–3705, 2011. Address: <https://www.scielosp.org/scielo.php?pid=S1413-81232011001000007&script=sci_arttext&tlng=es>.

LAMBERT, D. Zero-inflated Poisson regression, with an application to defects in manufacturing. **Technometrics**, Taylor & Francis, vol. 34, no. 1, pp. 1–14, 1992. Address:
<https://www.tandfonline.com/doi/abs/10.1080/00401706.1992.10485228>.

LIN, X.; LEE, L. GMM estimation of spatial autoregressive models with unknown heteroskedasticity. **Department of Economics, Ohio State University**, 2005. Address:
<https://www.asc.ohio-state.edu/lee.1777/wp/gmm-unheter-nov-15-05.pdf>.

LONG, J. S.; FREESE, J. **Regression models for categorical dependent variables using Stata**. StataCorp, Texas: Stata press, 2006.

MACHADO, D. C.; GONZAGA, G. O impacto dos fatores familiares sobre a defasagem idade-série de crianças no Brasil. **Revista Brasileira de Economia**, SciELO Brasil, vol. 61, no. 4, pp. 449–476, 2007. Address: <http://www.scielo.br/scielo.php?pid=S0034-71402007000400002&script=sci_arttext&tlng=es>.

MARCASSA, L. P.; BUSS, P. EDUCAÇÃO FÍSICA NA ESCOLA BÁSICA: DESLOCAMENTOS, SENTIDOS E CONTROVÉRSIAS. **Revista Brasileira de Ciências do Esporte**, vol. 36, 2016. Address: <http://www.revista.cbce.org.br/index.php/RBCE/article/view/2141>.

MEYER, O. L.; CASTRO-SCHILO, L.; AGUILAR-GAXIOLA, S. Determinants of mental health and self-rated health: a model of socioeconomic status, neighborhood safety, and physical activity. **American journal of public health**, American Public Health Association, vol. 104, no. 9, pp. 1734–1741, 2014.

MINCER, J. Economic development, growth of human capital, and the dynamics of the wage structure. **Journal of Economic Growth**, Springer, vol. 1, no. 1, pp. 29–48, 1996. Address: <https://link.springer.com/article/10.1007/BF00163341>.

— Investment in human capital and personal income distribution. **Journal of political economy**, The University of Chicago Press, vol. 66, no. 4, pp. 281–302, 1958. Address: <https://www.journals.uchicago.edu/doi/abs/10.1086/258055?journalCode=jpe>.

— Schooling, Experience, and Earnings. NBER Books, 1974. Address: <https://www.nber.org/chapters/c1765.pdf>.

MUSHKIN, S. J. Health as an Investment. **Journal of political economy**, The University of Chicago Press, vol. 70, 5, Part 2, pp. 129–157, 1962.

OAXACA, R. Male-female wage differentials in urban labor markets. **International economic review**, JSTOR, pp. 693–709, 1973. Address: <http://wwwbcf.usc.edu/~ridder/Lnotes/Undeconometrics/Transparanten/Wagedecomp.pdf>.

OLIVEIRA, F. S.; JUSTUS, M. Evidências do efeito negativo da depressao nos rendimentos das mulheres no Brasil: uma abordagem de variável instrumental, 2017. Address: <https://www.anpec.org.br/encontro/2017/submissao/files_I/i13-d6c2ec83ba76aedf4351ab00560a6df4.pdf>.

PARO, V. H. Progressão continuada, supervisão escolar e avaliação externa: implicações para a qualidade do ensino. **Revista Brasileira de Educação**, Associação Nacional de Pós-Graduação e Pesquisa em Educação, vol. 16, no. 48, pp. 695–716, 2011. Address: <https://www.redalyc.org/pdf/275/27520749009.pdf>.

PETRY, N. M. A comparison of young, middle-aged, and older adult treatment-seeking pathological gamblers. **The Gerontologist**, Oxford University Press, vol. 42, no. 1, pp. 92–99, 2002. Address: <https://academic.oup.com/gerontologist/article/42/1/92/641498>.

PINTO, F. M.; VAZ, A. F. Sobre a relação entre saberes e práticas corporais: notas para a investigação empırica do fracasso em aulas de educação fısica. **Educação & Realidade**, Universidade Federal do Rio Grande do Sul, vol. 34, no. 2, pp. 261–275, 2009.

POWELL, L. M.; SLATER, S.; CHALOUPKA, F. J. The relationship between community physical activity settings and race, ethnicity and socioeconomic status. **Evidence-Based Preventive Medicine**, vol. 1, no. 2, pp. 135–44, 2004. Address: <https://pdfs.semanticscholar.org/b9e0/885ad688b8ec40e783ee417075dd463b288b.pdf>.

RANSOM, M. R.; RANSOM, T. Do High School Sports Build or Reveal Character? IZA Discussion Paper, no. 11110, 2017. Address: <https://ssrn.com/abstract=3069453>.

REES, C. R.; HOWELL, F. M.; MIRACLE, A. W. Do high school sports build character? A quasi-experiment on a national sample. **The Social Science Journal**, Elsevier, vol. 27, no. 3, pp. 303–315, 1990. Address: <https://www.sciencedirect.com/science/article/abs/pii/036233199090027H>.

REHBERG, R. A. Behavioral and attitudinal consequences of high school interscholastic sports: A speculative consideration. **Adolescence**, Libra Publishers., vol. 4, no. 13, p. 69, 1969.

ROCHA, M. A. A.; SOUZA CAMPOS, M. d. F. de. Desigualdades salariais no mercado de trabalho urbano no Paraná: uma aplicação da metodologia de Heckman. **Revista Paranaense de Desenvolvimento-RPD**, no. 112, pp. 47–71, 2011. Address: <http://www.ipardes.gov.br/ojs/index.php/revistaparanaense/article/view/37/49>.

ROODMAN, D. Estimating Fully Observed Recursive Mixed-Process Models with cmp, 2009. Address: <https://www.cgdev.org/sites/default/files/1421516_file_Roodman_cmp_FINAL.pdf>.

ROSS, C. E.; HAYES, D. Exercise and psychologic well-being in the community. **American Journal of Epidemiology**, Oxford University Press, vol. 127, no. 4, pp. 762–771, 1988.

SALLES-COSTA, R. et al. Gênero e prática de atividade fısica de lazer. **Cadernos de Saúde pública**, SciELO Public Health, vol. 19, s325–s333, 2003. Address: <https://www.scielosp.org/article/csp/2003.v19suppl2/S325-S333/#>.

SALLIS, J.; BAUMAN, A.; PRATT, M. Environmental and policy interventions to promote physical activity. **American journal of preventive medicine**, Elsevier, vol. 15, no. 4, pp. 379–397, 1998.

SANTOS, S. C.; KNIJNIK, J. D. Motivos de adesão à prática de atividade fısica na vida adulta intermediária. **Revista Mackenzie de Educação Fısica e Esporte**, vol. 5, no. 1, 2009. Address: <http://editorarevistas.mackenzie.br/index.php/remef/article/view/1299>.

SCHAFER, W. E.; ARMER, J. M. Athletes are not inferior students. **Society**, Springer, vol. 6, no. 1, pp. 21–26, 1968.

SCHULTZ, T. W. Investment in human capital. **The American economic review**, JSTOR, pp. 1–17, 1961.

SNYDER, E. E. A theoretical analysis of academic and athletic roles. **Sociology of Sport Journal**, vol. 2, no. 3, pp. 210–217, 1985. Address: <https://journals.humankinetics.com/doi/abs/10.1123/ssj.2.3.210>.

SPADY, W. G. Lament for the letterman: Effects of peer status and extracurricular activities on goals and achievement. **American Journal of Sociology**, University of Chicago Press, vol. 75, 4, Part 2, pp. 680–702, 1970. Address: <https://www.journals.uchicago.edu/doi/abs/10.1086/224896>.

STATACORP. **Stata Statistical Software: Release 14**. College Station, TX: StataCorp LP, 2015.

TALBOT, L. A. et al. Falls in young, middle-aged and older community dwelling adults: perceived cause, environmental factors and injury. **BMC public health**, BioMed Central, vol. 5, no. 1, p. 86, 2005. Address: <https://bmcpublichealth.biomedcentral.com/articles/10.1186/1471-2458-5-86>.

TALIAFERRO, L. A. et al. High school youth and suicide risk: exploring protection afforded through physical activity and sport participation. **Journal of School Health**, Wiley Online Library, vol. 78, no. 10, pp. 545–553, 2008. Address: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1746-1561.2008.00342.x>.

TEIXEIRA, A. D. **Maior o peso, menor o salário? O impacto da obesidade no mercado de trabalho**. 2016. PhD thesis – Universidade de São Paulo. Address: <http://www.teses.usp.br/teses/disponiveis/12/12138/tde-05092016-163806/en.php>.

TROIANO, R. P.; MACERA, C. A.; BALLARD-BARBASH, R. Be physically active each day. How can we know? **The Journal of nutrition**, Oxford University Press, vol. 131, no. 2, 451s–460s, 2001.

VAITSMAN, J. Hierarquia de gênero e iniqüidade em saúde. **Physis: Revista de Saúde Coletiva**, SciELO Public Health, vol. 4, pp. 7–22, 1994. Address: <https://www.scielosp.org/scielo.php?pid=S0103-73311994000100001&script=sci_arttext&tlng=pt>.

VAZ, D. V.; HOFFMANN, R. Remuneração nos serviços no Brasil: o contraste entre funcionários públicos e privados. **Economia e Sociedade**, SciELO Brasil, vol. 16, no. 2, pp. 199–232, 2007. Address: <%3C%20http://www.scielo.br/pdf/ecos/v16n2/a04v16n2>.

VUONG, Q. H. Likelihood ratio tests for model selection and non-nested hypotheses. **Econometrica: Journal of the Econometric Society**, JSTOR, pp. 307–333, 1989. Address: <https://www.jstor.org/stable/1912557?read-now=1&seq=1#page_scan_tab_contents>.

WEINBERG, R. S.; GOULD, D. **Fundamentos da psicologia do esporte e do exercıcio**. São Paulo: Artmed editora, 2016.

WERTS, C. E. The many faces of intelligence. **Journal of Educational Psychology**, American Psychological Association, vol. 58, no. 4, p. 198, 1967.

WESSEL, T. R. et al. Relationship of physical fitness vs body mass index with coronary artery disease and cardiovascular events in women. **Jama**, American Medical Association, vol. 292, no. 10, pp. 1179–1187, 2004. Address: <https://jamanetwork.com/journals/jama/articlepdf/199393/joc41045.pdf>.

WHO, W. H. O. **Global action plan on physical activity 2018–2030: more active people for a healthier world**. Geneva: World Health Organization, 2018. Address: <https://apps.who.int/iris/bitstream/handle/10665/272722/9789241514187-eng.pdf>.

WHO, W. H. O. et al. Noncommunicable diseases country profiles 2018. World Health Organization, 2018. Address: <https://www.who.int/nmh/publications/ncd-profiles-2018/en/>.

WHO, W. H. O. WHO website. **Accessed (2018 Oct 20)**, 2018. Address: <http://www.who.int/gho/ncd/risk_factors/physical_activity_text/en/>.

WILSON, P. The misuse of the Vuong test for non-nested models to test for zero-inflation. **Economics Letters**, Elsevier, vol. 127, pp. 51–53, 2015. Address: <sciencedirect.com/science/article/abs/pii/S016517651400490X>.

WOOLDRIDGE, J. M. **Econometric analysis of cross section and panel data**. Cambridge, MA: MIT press, 2010.

Appendix A

Table 1: One-Sided t Test p-values Results

	Left-Side t Test	**Right-Side t Test**
Endogenous:		
p_i	0.000^{***}	0.999
y_i	0.000^{***}	0.999
Affect PAP:		
sex	0.999	0.000^{***}
age	0.999	0.000^{***}
young adult	0.000^{***}	0.999
middle age adult	0.999	0.000^{***}
old adult	0.999	0.000^{***}
indigenous	0.058^{*}	0.941
yellow	0.000^{***}	0.999
brown	0.999	0.000^{***}
black	0.999	0.000^{***}
householder	0.957	0.042^{**}
spouse	0.999	0.000^{***}
chores hours	0.999	0.000^{***}
working hours	0.000^{***}	0.999
working_h categories	0.229	0.770
n_jobs	0.000^{***}	0.999
dependent kids	0.999	0.000^{***}
public_neigh_space	0.000^{***}	0.999
metropolitan	0.004^{**}	0.995
rural	0.999	0.000^{***}
I_public_pap	0.000^{***}	0.999
health	0.000^{***}	0.999
Affected by PAP:		
school	0.000^{***}	0.999
health stock	0.965	0.034^{**}

* $p < 0.10$ ** $p < 0.05$, *** $p < 0.01$

(a) The number of observations N is $33,865$ for the individuals that do not practice physical activity and $21,255$ for the practitioners, except for the income variable, y_i, in which the sample size is $22,006$ and $15,057$ observations for the physical activity non practitioners and practitioners, respectively, since, for this variable, there only are observations for the individuals who are occupied.

(b) Left-side test null hypothesis: $NON - PAP < PAP$.

(c) Right-side test null hypothesis: $NON - PAP > PAP$.

Source: Elaborate by the author, from PNAD dataset.

Appendix B

Table 2: IV-Probit Occupation Equation Coefficients.

	All Ages			Young Adults			Middle-Aged			Older Adults		
p_i	All	Men	Women	All	Men	Women	All	Men	Women	All	Men	Women
pâp	-0.139***	-0.0625	-0.119***	-0.0150	-0.0129	-0.0570	-0.293***	-0.269***	-0.221***	-0.358***	0.162**	-0.337***
	(0.0264)	(0.0484)	(0.0295)	(0.0349)	(0.0592)	(0.0406)	(0.0455)	(0.0970)	(0.0477)	(0.114)	(0.0652)	(0.106)
sex	-0.326***	-	-	-0.285***	-	-	-0.241***	-	-	-0.553***	-	-
	(0.0192)	-	-	(0.0292)	-	-	(0.0306)	-	-	(0.0494)	-	-
schooling	0.0300***	0.0392***	0.0237***	0.0708***	0.0590***	0.0948***	0.0277***	0.0186*	0.0340***	0.0130	0.0362*	-0.00222
	(0.00439)	(0.00681)	(0.00589)	(0.00922)	(0.0136)	(0.0132)	(0.00650)	(0.0100)	(0.00879)	(0.0122)	(0.0185)	(0.0164)
other_income	-0.520***	-0.964***	-0.304***	-0.501***	-1.135***	-0.237***	-0.405***	-0.637***	-0.279***	-0.694***	-1.052***	-0.497***
	(0.0350)	(0.0549)	(0.0423)	(0.0565)	(0.106)	(0.0647)	(0.0513)	(0.0768)	(0.0649)	(0.0913)	(0.129)	(0.118)
metropolitan	-0.0720***	-0.0616**	-0.0750***	-0.0670***	-0.0548	-0.0654**	-0.0552**	-0.0714*	-0.0452	-0.0955**	0.00543	-0.147***
	(0.0168)	(0.0273)	(0.0214)	(0.0258)	(0.0410)	(0.0330)	(0.0262)	(0.0425)	(0.0331)	(0.0409)	(0.0627)	(0.0538)
rural	0.0796***	0.238***	-0.0668**	-0.0194	0.0814	-0.107**	0.0788**	0.208***	-0.0343	0.369***	0.772***	-0.0305
	(0.0248)	(0.0416)	(0.0332)	(0.0370)	(0.0616)	(0.0488)	(0.0367)	(0.0613)	(0.0496)	(0.0637)	(0.0973)	(0.0962)
school_9	0.0675***	0.0269**	0.0843***	0.0563***	0.0303	0.0613***	0.0327***	0.00979	0.0359**	-0.0277	-0.0713**	-0.0338
	(0.00703)	(0.0114)	(0.00910)	(0.0127)	(0.0197)	(0.0172)	(0.0110)	(0.0181)	(0.0141)	(0.0199)	(0.0305)	(0.0261)
chores hours	-0.0252***	-0.0188***	-0.0246***	-0.0271***	-0.0226***	-0.0250***	-0.0262***	-0.0206***	-0.0264***	-0.0175***	-0.0140***	-0.0178***
	(0.000638)	(0.00143)	(0.000739)	(0.00102)	(0.00232)	(0.00117)	(0.000955)	(0.00206)	(0.00112)	(0.00159)	(0.00348)	(0.00177)
indigenous	-0.150	-0.238	-0.109	-0.145	-0.314	0.0285	-0.279*	-0.123	-0.430*	0.365	0.121	0.511
	(0.124)	(0.196)	(0.164)	(0.178)	(0.282)	(0.210)	(0.166)	(0.281)	(0.236)	(0.347)	(0.355)	(0.392)
yellow	-0.179	-0.244	-0.0946	-0.281	-0.508	-0.000616	-0.151	0.120	-0.230	0.231	0.105	0.237
	(0.129)	(0.220)	(0.146)	(0.221)	(0.361)	(0.236)	(0.175)	(0.373)	(0.201)	(0.258)	(0.419)	(0.365)
brown	0.0258	-0.0585**	0.0857***	0.0638**	0.000267	0.111***	0.00606	-0.0793*	0.0711**	-0.0266	-0.105	0.0443
	(0.0178)	(0.0290)	(0.0227)	(0.0269)	(0.0433)	(0.0343)	(0.0277)	(0.0450)	(0.0353)	(0.0443)	(0.0682)	(0.0595)
black	0.0626**	-0.0221	0.122***	0.0885**	-0.00950	0.158***	0.0679	0.0101	0.116**	-0.0676	-0.132	-0.0252
	(0.0274)	(0.0435)	(0.0354)	(0.0418)	(0.0641)	(0.0550)	(0.0430)	(0.0694)	(0.0549)	(0.0658)	(0.105)	(0.0870)
experience	0.0556***	0.0463***	0.0636***	0.105***	0.112***	0.0984***	-0.0218*	-0.0293	-0.0245*	-0.189***	-0.300***	-0.151**
	(0.00202)	(0.00329)	(0.00264)	(0.00635)	(0.00961)	(0.00860)	(0.0115)	(0.0188)	(0.0147)	(0.0583)	(0.0945)	(0.0764)
exp_2	-0.000948***	-0.000725***	-0.00115***	-0.00229***	-0.00297***	-0.00160***	0.0000633	0.0000672	0.000147	0.00146**	0.00286***	0.000859
	(0.0000388)	(0.0000607)	(0.0000527)	(0.000283)	(0.000420)	(0.000395)	(0.000185)	(0.000295)	(0.000242)	(0.000619)	(0.000998)	(0.000824)
householder	0.432***	0.439***	0.301***	0.437***	0.434***	0.281***	0.401***	0.419***	0.292***	0.237***	0.218***	0.191***
	(0.0169)	(0.0282)	(0.0231)	(0.0263)	(0.0451)	(0.0353)	(0.0264)	(0.0412)	(0.0363)	(0.0471)	(0.0764)	(0.0670)
spouse	0.220***	0.414***	0.0754***	0.262***	0.499***	0.0974***	0.145***	0.309***	0.00738	0.0540	0.233***	-0.0201
	(0.0169)	(0.0290)	(0.0234)	(0.0267)	(0.0477)	(0.0357)	(0.0282)	(0.0457)	(0.0388)	(0.0446)	(0.0676)	(0.0672)
north	-0.120***	0.0417	-0.255***	-0.172***	0.0147	-0.312***	-0.0570	0.177***	-0.223***	0.0279	-0.0758	0.0320
	(0.0241)	(0.0402)	(0.0311)	(0.0351)	(0.0579)	(0.0453)	(0.0380)	(0.0651)	(0.0492)	(0.0665)	(0.105)	(0.0888)
northeast	-0.101***	-0.0829***	-0.132***	-0.135***	-0.123**	-0.150***	-0.0582*	-0.0185	-0.0947**	-0.00657	-0.0402	-0.0635
	(0.0197)	(0.0320)	(0.0252)	(0.0300)	(0.0478)	(0.0384)	(0.0305)	(0.0491)	(0.0389)	(0.0493)	(0.0806)	(0.0647)
south	0.177***	0.203***	0.155***	0.257***	0.315***	0.232***	0.0966***	0.123**	0.0755*	0.0831	-0.0454	0.138**
	(0.0234)	(0.0385)	(0.0296)	(0.0370)	(0.0605)	(0.0470)	(0.0355)	(0.0586)	(0.0451)	(0.0539)	(0.0835)	(0.0692)
central west	0.0384	0.121***	-0.0282	0.0393	0.128**	-0.0278	0.0683*	0.163**	-0.00251	0.0439	0.0717	-0.0402
	(0.0251)	(0.0419)	(0.0320)	(0.0382)	(0.0617)	(0.0490)	(0.0393)	(0.0645)	(0.0504)	(0.0623)	(0.107)	(0.0817)
dependent kids	-0.0211**	0.00468	-0.0537***	-0.0503***	-0.0265	-0.0755***	0.00318	0.0579**	-0.0493***	-0.0635**	-0.0306	-0.0956**
	(0.00883)	(0.0156)	(0.0112)	(0.0126)	(0.0222)	(0.0160)	(0.0143)	(0.0247)	(0.0189)	(0.0314)	(0.0502)	(0.0442)
retired	-1.385***	-1.773***	-1.028***	-	-	-	-	-	-	-1.077***	-1.614***	-0.669***
	(0.0360)	(0.0495)	(0.0506)	-	-	-	-	-	-	(0.0475)	(0.0644)	(0.0590)
Constant	-0.128***	-0.237***	-0.344***	-0.866***	-0.764***	-1.347***	1.295***	1.427***	1.142***	6.417***	8.277***	5.473***
	(0.0476)	(0.0748)	(0.0627)	(0.0865)	(0.131)	(0.123)	(0.183)	(0.299)	(0.232)	(1.379)	(2.249)	(1.770)
N	55,120	25,954	29,166	23,404	10,981	12,423	23,284	11,234	12,050	8,432	3,739	4,693

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

(a) Men's Older Adults model is Probit.

Source: Elaborate by the author.

Table 3: Biprobit Occupation Equation Coefficients.

	All Ages			Young Adults			Middle-Aged			Older Adults		
p_i	All	Men	Women	All	Men	Women	All	Men	Women	All	Men	Women
pâp	-0.337***	-0.373***	-0.248***	-0.238***	-0.305***	-0.133***	-0.405***	-0.451***	-0.328***	-0.584***	-0.690***	-0.510***
	(0.0218)	(0.0373)	(0.0265)	(0.0328)	(0.0546)	(0.0394)	(0.0319)	(0.0536)	(0.0394)	(0.0577)	(0.0921)	(0.0779)
sex	-0.345***	-	-	-0.329***	-	-	-0.240***	-	-	-0.542***	-	-
	(0.0189)	-	-	(0.0288)	-	-	(0.0303)	-	-	(0.0486)	-	-
schooling	0.0321***	0.0423***	0.0254***	0.0731***	0.0609***	0.0961***	0.0288***	0.0204**	0.0353***	0.0163	0.0430**	0.00239
	(0.00436)	(0.00674)	(0.00587)	(0.00927)	(0.0137)	(0.0132)	(0.00646)	(0.00990)	(0.00876)	(0.0119)	(0.0178)	(0.0162)
other_income	-0.504***	-0.925***	-0.295***	-0.483***	-1.102***	-0.231***	-0.397***	-0.617***	-0.273***	-0.654***	-0.908***	-0.469***
	(0.0348)	(0.0552)	(0.0421)	(0.0563)	(0.106)	(0.0646)	(0.0510)	(0.0764)	(0.0646)	(0.0898)	(0.135)	(0.116)
metropolitan	-0.0787***	-0.0665**	-0.0822***	-0.0730***	-0.0565	-0.0693**	-0.0592**	-0.0725*	-0.0518	-0.104***	-0.0356	-0.155***
	(0.0167)	(0.0269)	(0.0213)	(0.0257)	(0.0406)	(0.0330)	(0.0260)	(0.0419)	(0.0330)	(0.0403)	(0.0601)	(0.0533)
rural	0.0713***	0.223***	-0.0710**	-0.0197	0.0854	-0.109**	0.0713*	0.194***	-0.0394	0.336***	0.628***	-0.0407
	(0.0246)	(0.0409)	(0.0331)	(0.0367)	(0.0606)	(0.0486)	(0.0365)	(0.0605)	(0.0495)	(0.0620)	(0.0954)	(0.0955)
school_9	0.0705***	0.0313***	0.0865***	0.0601***	0.0339*	0.0628***	0.0350***	0.0142	0.0378***	-0.0224	-0.0398	-0.0339
	(0.00698)	(0.0112)	(0.00907)	(0.0127)	(0.0197)	(0.0171)	(0.0109)	(0.0179)	(0.0140)	(0.0196)	(0.0300)	(0.0259)
chores hours	-0.0246***	-0.0175***	-0.0244***	-0.0267***	-0.0214***	-0.0250***	-0.0258***	-0.0197***	-0.0261***	-0.0165***	-0.00862***	-0.0172***
	(0.000634)	(0.00141)	(0.000738)	(0.00102)	(0.00230)	(0.00117)	(0.000945)	(0.00201)	(0.00111)	(0.00152)	(0.00325)	(0.00175)
indigenous	-0.131	-0.214	-0.0991	-0.103	-0.253	0.0391	-0.275*	-0.142	-0.420*	0.355	0.291	0.479
	(0.124)	(0.197)	(0.163)	(0.180)	(0.289)	(0.209)	(0.165)	(0.271)	(0.235)	(0.341)	(0.315)	(0.384)
yellow	-0.174	-0.258	-0.0844	-0.280	-0.522	0.00626	-0.156	0.0941	-0.227	0.256	0.260	0.251
	(0.130)	(0.223)	(0.146)	(0.226)	(0.368)	(0.237)	(0.174)	(0.374)	(0.201)	(0.246)	(0.339)	(0.362)
brown	0.0223	-0.0572**	0.0813***	0.0646**	0.00693	0.111***	0.00345	-0.0795*	0.0659*	-0.0338	-0.124*	0.0369
	(0.0177)	(0.0286)	(0.0227)	(0.0267)	(0.0427)	(0.0343)	(0.0275)	(0.0444)	(0.0352)	(0.0436)	(0.0659)	(0.0589)
black	0.0614**	-0.0162	0.119***	0.0941**	0.00281	0.160***	0.0647	0.00947	0.110**	-0.0664	-0.0976	-0.0279
	(0.0272)	(0.0430)	(0.0353)	(0.0416)	(0.0635)	(0.0550)	(0.0427)	(0.0686)	(0.0547)	(0.0646)	(0.100)	(0.0859)
experience	0.0541***	0.0416***	0.0633***	0.104***	0.107***	0.0986***	-0.0225**	-0.0319*	-0.0247*	-0.179***	-0.266***	-0.148*
	(0.00201)	(0.00324)	(0.00263)	(0.00633)	(0.00958)	(0.00858)	(0.0114)	(0.0186)	(0.0147)	(0.0575)	(0.0909)	(0.0758)
exp_2	-0.000931***	-0.000675***	-0.00115***	-0.00228***	-0.00293***	-0.00160***	0.0000759	0.000106	0.000153	0.00137**	0.00252***	0.000851
	(0.0000385)	(0.0000600)	(0.0000525)	(0.000283)	(0.000419)	(0.000394)	(0.000184)	(0.000292)	(0.000241)	(0.000610)	(0.000958)	(0.000817)
householder	0.424***	0.434***	0.299***	0.429***	0.426***	0.279***	0.399***	0.419***	0.293***	0.239***	0.237***	0.191***
	(0.0168)	(0.0279)	(0.0230)	(0.0262)	(0.0447)	(0.0353)	(0.0263)	(0.0409)	(0.0362)	(0.0466)	(0.0744)	(0.0664)
spouse	0.215***	0.402***	0.0772***	0.251***	0.480***	0.0947***	0.150***	0.309***	0.0166	0.0619	0.264***	-0.0146
	(0.0168)	(0.0286)	(0.0233)	(0.0266)	(0.0472)	(0.0357)	(0.0280)	(0.0452)	(0.0387)	(0.0438)	(0.0655)	(0.0665)
north	-0.118***	0.0460	-0.257***	-0.166***	0.0250	-0.312***	-0.0586	0.174***	-0.226***	0.0229	-0.0611	0.0214
	(0.0239)	(0.0395)	(0.0310)	(0.0349)	(0.0572)	(0.0452)	(0.0379)	(0.0642)	(0.0491)	(0.0655)	(0.100)	(0.0880)
northeast	-0.0942***	-0.0741**	-0.127***	-0.127***	-0.111**	-0.149***	-0.0559*	-0.0161	-0.0912**	0.00208	-0.00814	-0.0534
	(0.0196)	(0.0315)	(0.0251)	(0.0298)	(0.0472)	(0.0384)	(0.0303)	(0.0487)	(0.0388)	(0.0486)	(0.0773)	(0.0641)
south	0.179***	0.203***	0.159***	0.259***	0.315***	0.233***	0.0983***	0.124**	0.0784*	0.0912*	-0.0243	0.149**
	(0.0232)	(0.0381)	(0.0295)	(0.0368)	(0.0600)	(0.0469)	(0.0353)	(0.0580)	(0.0449)	(0.0530)	(0.0818)	(0.0683)
central west	0.0425*	0.121***	-0.0240	0.0387	0.123**	-0.0270	0.0713*	0.166***	0.00140	0.0546	0.0777	-0.0212
	(0.0249)	(0.0414)	(0.0319)	(0.0379)	(0.0609)	(0.0489)	(0.0391)	(0.0639)	(0.0501)	(0.0613)	(0.104)	(0.0803)
dependent kids	-0.0269***	0.00101	-0.0582***	-0.0557***	-0.0307	-0.0780***	0.000409	0.0553**	-0.0523***	-0.0697**	-0.0308	-0.105**
	(0.00876)	(0.0153)	(0.0111)	(0.0125)	(0.0219)	(0.0160)	(0.0142)	(0.0243)	(0.0189)	(0.0311)	(0.0486)	(0.0438)
retired	-1.348***	-1.717***	-1.007***	-	-	-	-	-	-	-1.025***	-1.433***	-0.635***
	(0.0357)	(0.0492)	(0.0503)	-	-	-	-	-	-	(0.0442)	(0.0675)	(0.0582)
Constant	-0.0530	-0.0900	-0.320***	-0.770***	-0.613***	-1.337***	1.317***	1.486***	1.152***	6.138***	7.501***	5.375***
	(0.0469)	(0.0727)	(0.0624)	(0.0867)	(0.131)	(0.122)	(0.182)	(0.295)	(0.231)	(1.358)	(2.170)	(1.753)
athrho	0.422***	0.421***	0.387***	0.332***	0.371***	0.221***	0.501***	0.486***	0.497***	0.647***	0.784***	0.590***
	(0.0181)	(0.0279)	(0.0244)	(0.0282)	(0.0421)	(0.0373)	(0.0266)	(0.0374)	(0.0383)	(0.0501)	(0.0844)	(0.0710)
N	55,120	25,954	29,166	23,404	10,981	12,423	23,284	11,234	12,050	8,432	3,739	4,693

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

(a) *athrho* stands for the correlation between p_i and *pap*.

Source: Elaborate by the author.

Table 4: Heckman ML Occupation Equation Coefficients.

	All Ages			Young Adults			Middle-Aged			Older Adults		
p_i	All	Men	Women	All	Men	Women	All	Men	Women	All	Men	Women
pâp	0.0426***	0.0783***	0.0283	0.0393	0.102***	0.00223	0.0459*	0.0929**	0.0244	0.0850**	0.164**	0.0472
	(0.0164)	(0.0265)	(0.0212)	(0.0245)	(0.0354)	(0.0322)	(0.0260)	(0.0438)	(0.0327)	(0.0419)	(0.0650)	(0.0548)
sex	-0.301***	-	-	-0.271***	-	-	-0.225***	-	-	-0.547***	-	-
	(0.0191)	-	-	(0.0289)	-	-	(0.0307)	-	-	(0.0495)	-	-
schooling	0.0276***	0.0373***	0.0220***	0.0704***	0.0413***	0.0943***	0.0228***	0.0133	0.0308***	0.00595	0.0360*	-0.00886
	(0.00438)	(0.00679)	(0.00589)	(0.00918)	(0.0135)	(0.0133)	(0.00654)	(0.0101)	(0.00879)	(0.0122)	(0.0185)	(0.0164)
other_income	-0.530***	-0.975***	-0.312***	-0.505***	-1.010***	-0.241***	-0.424***	-0.663***	-0.290***	-0.743***	-1.041***	-0.529***
	(0.0350)	(0.0545)	(0.0421)	(0.0561)	(0.108)	(0.0647)	(0.0515)	(0.0763)	(0.0652)	(0.0896)	(0.130)	(0.120)
school_9	0.0645***	0.0248**	0.0810***	0.0543***	0.0640***	0.0588***	0.0278**	0.00493	0.0319**	-0.0371*	-0.0731**	-0.0388
	(0.00701)	(0.0113)	(0.00909)	(0.0126)	(0.0196)	(0.0171)	(0.0110)	(0.0180)	(0.0140)	(0.0201)	(0.0303)	(0.0265)
chores hours	-0.0257***	-0.0194***	-0.0249***	-0.0272***	-0.0158***	-0.0251***	-0.0273***	-0.0222***	-0.0269***	-0.0191***	-0.0146***	-0.0188***
	(0.000636)	(0.00143)	(0.000737)	(0.00102)	(0.00205)	(0.00116)	(0.000949)	(0.00206)	(0.00111)	(0.00154)	(0.00339)	(0.00175)
indigenous	-0.161	-0.242	-0.114	-0.150	-0.408*	0.0188	-0.279	-0.0902	-0.443*	0.385	0.105	0.562
	(0.124)	(0.196)	(0.164)	(0.177)	(0.246)	(0.209)	(0.170)	(0.292)	(0.238)	(0.348)	(0.358)	(0.403)
yellow	-0.184	-0.244	-0.106	-0.284	-0.447	-0.00935	-0.116	0.180	-0.219	0.146	0.0871	0.175
	(0.127)	(0.216)	(0.145)	(0.219)	(0.314)	(0.235)	(0.179)	(0.372)	(0.208)	(0.273)	(0.411)	(0.376)
brown	0.0295*	-0.0577**	0.0912***	0.0650**	-0.0109	0.113***	0.0147	-0.0777*	0.0827**	-0.0133	-0.102	0.0568
	(0.0178)	(0.0290)	(0.0228)	(0.0269)	(0.0400)	(0.0343)	(0.0278)	(0.0454)	(0.0354)	(0.0444)	(0.0681)	(0.0594)
black	0.0639**	-0.0226	0.127***	0.0883**	-0.0134	0.159***	0.0799*	0.0172	0.130**	-0.0700	-0.124	-0.0181
	(0.0274)	(0.0436)	(0.0354)	(0.0418)	(0.0594)	(0.0551)	(0.0434)	(0.0703)	(0.0551)	(0.0670)	(0.106)	(0.0884)
experience	0.0564***	0.0481***	0.0635***	0.105***	0.104***	0.0979***	-0.0192*	-0.0232	-0.0248*	-0.199***	-0.309***	-0.155**
	(0.00202)	(0.00325)	(0.00264)	(0.00635)	(0.00949)	(0.00861)	(0.0116)	(0.0190)	(0.0149)	(0.0591)	(0.0949)	(0.0772)
exp_2	-0.000954***	-0.000743***	-0.00115***	-0.00228***	-0.00260***	-0.00158***	0.0000246	-0.0000177	0.000148	0.00155**	0.00295***	0.000883
	(0.0000388)	(0.0000604)	(0.0000527)	(0.000283)	(0.000423)	(0.000395)	(0.000187)	(0.000299)	(0.000244)	(0.000627)	(0.00100)	(0.000831)
householder	0.435***	0.440***	0.301***	0.437***	0.396***	0.282***	0.400***	0.413***	0.287***	0.228***	0.219***	0.189***
	(0.0169)	(0.0282)	(0.0232)	(0.0263)	(0.0444)	(0.0354)	(0.0266)	(0.0415)	(0.0365)	(0.0474)	(0.0764)	(0.0673)
spouse	0.224***	0.417***	0.0718***	0.265***	0.443***	0.102***	0.134***	0.312***	-0.00759	0.0408	0.234***	-0.0274
	(0.0169)	(0.0290)	(0.0234)	(0.0267)	(0.0465)	(0.0357)	(0.0284)	(0.0461)	(0.0389)	(0.0447)	(0.0673)	(0.0675)
metropolitan	-0.0646***	-0.0594**	-0.0673***	-0.0652**	-0.00920	-0.0628*	-0.0402	-0.0665	-0.0306	-0.0751*	0.0115	-0.133**
	(0.0169)	(0.0273)	(0.0214)	(0.0258)	(0.0378)	(0.0330)	(0.0265)	(0.0429)	(0.0333)	(0.0416)	(0.0633)	(0.0543)
rural	0.0860***	0.243***	-0.0626*	-0.0194	0.137**	-0.106**	0.0959***	0.232***	-0.0260	0.419***	0.774***	-0.00453
	(0.0248)	(0.0418)	(0.0333)	(0.0370)	(0.0612)	(0.0488)	(0.0368)	(0.0616)	(0.0496)	(0.0629)	(0.0976)	(0.0965)
north	-0.122***	0.0389	-0.254***	-0.174***	0.0455	-0.313***	-0.0546	0.172***	-0.218***	0.0353	-0.0692	0.0461
	(0.0241)	(0.0403)	(0.0312)	(0.0352)	(0.0557)	(0.0453)	(0.0382)	(0.0658)	(0.0491)	(0.0672)	(0.105)	(0.0893)
northeast	-0.106***	-0.0849***	-0.138***	-0.137***	-0.0880*	-0.153***	-0.0663**	-0.0246	-0.105***	-0.0248	-0.0373	-0.0819
	(0.0198)	(0.0321)	(0.0252)	(0.0300)	(0.0451)	(0.0384)	(0.0306)	(0.0494)	(0.0390)	(0.0498)	(0.0806)	(0.0650)
south	0.172***	0.201***	0.150***	0.255***	0.239***	0.229***	0.0903**	0.117**	0.0695	0.0606	-0.0463	0.113
	(0.0234)	(0.0385)	(0.0297)	(0.0371)	(0.0536)	(0.0470)	(0.0358)	(0.0591)	(0.0453)	(0.0541)	(0.0836)	(0.0697)
central west	0.0333	0.119***	-0.0348	0.0375	0.133**	-0.0305	0.0572	0.150**	-0.0123	0.0207	0.0740	-0.0723
	(0.0252)	(0.0419)	(0.0321)	(0.0383)	(0.0570)	(0.0491)	(0.0395)	(0.0649)	(0.0505)	(0.0626)	(0.107)	(0.0824)
dependent kids	-0.0173*	0.00425	-0.0489***	-0.0515***	0.0141	-0.0760***	0.0101	0.0595**	-0.0414**	-0.0552*	-0.0359	-0.0784*
	(0.00884)	(0.0158)	(0.0111)	(0.0126)	(0.0204)	(0.0159)	(0.0143)	(0.0251)	(0.0188)	(0.0320)	(0.0514)	(0.0450)
retired	-1.404***	-1.787***	-1.041***	-	-	-	-	-	-	-1.136***	-1.610***	-0.712***
	(0.0359)	(0.0493)	(0.0506)	-	-	-	-	-	-	(0.0439)	(0.0647)	(0.0583)
Constant	-0.193***	-0.299***	-0.367***	-0.887***	-0.790***	-1.354***	1.203***	1.279***	1.116***	6.653***	8.506***	5.579***
	(0.0471)	(0.0724)	(0.0628)	(0.0859)	(0.126)	(0.123)	(0.185)	(0.302)	(0.234)	(1.397)	(2.256)	(1.789)
athrho	-0.117***	-0.0986**	-0.153***	-0.105***	1.102***	-0.149***	-0.151***	-0.107***	-0.188***	-0.156**	-0.201**	-0.131
	(0.0213)	(0.0388)	(0.0264)	(0.0302)	(0.0655)	(0.0348)	(0.0267)	(0.0350)	(0.0381)	(0.0652)	(0.0895)	(0.0896)
lnsigma	-0.346***	-0.368***	-0.321***	-0.418***	-0.340***	-0.365***	-0.328***	-0.342***	-0.316***	-0.195***	-0.208***	-0.194***
	(0.00959)	(0.0122)	(0.0152)	(0.0168)	(0.0261)	(0.0255)	(0.0130)	(0.0157)	(0.0218)	(0.0254)	(0.0350)	(0.0367)
Mills Lambda	-0.0823	-0.0681	-0.110	-0.0687	0.570	-0.103	-0.108	-0.0756	-0.136	-0.127	-0.161	-0.107
N	55,120	25,954	29,166	23,404	10,981	12,423	23,284	11,234	12,050	8,432	3,739	4,693

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

(a) Variables *dependent kids* and *chores hours* used only on the occupation equation, due to identification reasons.

(b) *athrho* stands for the inverse hyperbolic tangent of ρ (the correlation between the error terms of the equations); and *lnsigma* stands for the natural logarithm of σ (the standard error of the residuals of the income equation).

(c) $\lambda = \rho\sigma$. Mills Lambda is not estimated, so that it does not present statistical significance tests.

Source: Elaborate by the author.

Table 5: Heckman Two-Step Occupation Equation Coefficients.

	All Ages			Young Adults			Middle-Aged			Older Adults		
p_i	All	Men	Women	All	Men	Women	All	Men	Women	All	Men	Women
pâp	0.0494***	0.0769***	0.0431**	0.0524***	0.0631**	0.0325	0.0665***	0.113***	0.0389	0.0679**	0.130**	0.0402
	(0.0134)	(0.0217)	(0.0174)	(0.0199)	(0.0312)	(0.0264)	(0.0213)	(0.0352)	(0.0272)	(0.0346)	(0.0558)	(0.0446)
sex	-0.346***	-	-	-0.295***	-	-	-0.284***	-	-	-0.549***	-	-
	(0.0151)	-	-	(0.0231)	-	-	(0.0241)	-	-	(0.0384)	-	-
schooling	0.0257***	0.0341***	0.0219***	0.0674***	0.0536***	0.0868***	0.0209***	0.0104	0.0297***	-0.00455	0.0201	-0.0173
	(0.00350)	(0.00544)	(0.00469)	(0.00746)	(0.0110)	(0.0108)	(0.00530)	(0.00815)	(0.00714)	(0.00981)	(0.0150)	(0.0133)
other_income	-0.710***	-1.195***	-0.473***	-0.718***	-1.418***	-0.445***	-0.605***	-0.887***	-0.440***	-0.833***	-1.132***	-0.609***
	(0.0256)	(0.0438)	(0.0315)	(0.0397)	(0.0761)	(0.0471)	(0.0383)	(0.0623)	(0.0484)	(0.0736)	(0.111)	(0.0979)
school_9	0.0621***	0.0206**	0.0770***	0.0507***	0.0227	0.0580***	0.0242***	-0.00248	0.0294***	-0.00596	-0.0324	-0.00628
	(0.00562)	(0.00905)	(0.00734)	(0.0102)	(0.0158)	(0.0141)	(0.00879)	(0.0142)	(0.0114)	(0.0161)	(0.0253)	(0.0214)
chores hours	-0.0250***	-0.0178***	-0.0245***	-0.0266***	-0.0201***	-0.0249***	-0.0267***	-0.0204***	-0.0269***	-0.0191***	-0.0164***	-0.0182***
	(0.000483)	(0.00112)	(0.000558)	(0.000773)	(0.00184)	(0.000892)	(0.000736)	(0.00165)	(0.000851)	(0.00117)	(0.00257)	(0.00136)
indigenous	-0.0770	-0.140	-0.0192	-0.117	-0.180	-0.0482	-0.191	-0.0374	-0.294	0.423*	-0.0957	0.705**
	(0.0946)	(0.157)	(0.120)	(0.138)	(0.222)	(0.179)	(0.152)	(0.246)	(0.201)	(0.239)	(0.433)	(0.289)
yellow	-0.116	-0.111	-0.0775	-0.0317	-0.171	0.134	-0.142	0.192	-0.225	-0.0720	-0.130	-0.0128
	(0.0980)	(0.169)	(0.122)	(0.156)	(0.250)	(0.205)	(0.157)	(0.329)	(0.186)	(0.214)	(0.357)	(0.271)
brown	0.0168	-0.0406*	0.0547***	0.0390*	0.00960	0.0572**	-0.00357	-0.0786**	0.0494*	0.0103	-0.0310	0.0423
	(0.0146)	(0.0236)	(0.0188)	(0.0220)	(0.0355)	(0.0285)	(0.0229)	(0.0369)	(0.0294)	(0.0365)	(0.0563)	(0.0489)
black	0.0659***	-0.000961	0.110***	0.0677**	-0.0179	0.124***	0.0762**	0.0487	0.0979**	-0.0315	-0.127	0.0342
	(0.0223)	(0.0351)	(0.0292)	(0.0337)	(0.0519)	(0.0449)	(0.0347)	(0.0555)	(0.0449)	(0.0568)	(0.0874)	(0.0756)
experience	0.0574***	0.0515***	0.0620***	0.112***	0.120***	0.105***	-0.0162*	-0.0270*	-0.0168	-0.126**	-0.175**	-0.109*
	(0.00161)	(0.00257)	(0.00211)	(0.00517)	(0.00797)	(0.00697)	(0.00954)	(0.0157)	(0.0122)	(0.0497)	(0.0799)	(0.0655)
exp_2	-0.000986***	-0.000817***	-0.00114***	-0.00252***	-0.00323***	-0.00193***	-0.0000535	0.0000223	-0.0000166	0.000739	0.00149*	0.000423
	(0.0000309)	(0.0000476)	(0.0000419)	(0.000231)	(0.000351)	(0.000319)	(0.000154)	(0.000248)	(0.000200)	(0.000528)	(0.000837)	(0.000704)
householder	0.491***	0.531***	0.345***	0.529***	0.568***	0.355***	0.402***	0.429***	0.299***	0.264***	0.220***	0.244***
	(0.0138)	(0.0234)	(0.0190)	(0.0212)	(0.0372)	(0.0287)	(0.0218)	(0.0354)	(0.0301)	(0.0395)	(0.0648)	(0.0561)
spouse	0.135***	0.322***	-0.00977	0.171***	0.438***	-0.0104	0.0219	0.176***	-0.0929***	-0.00668	0.128**	-0.0534
	(0.0137)	(0.0236)	(0.0192)	(0.0216)	(0.0398)	(0.0289)	(0.0222)	(0.0361)	(0.0315)	(0.0362)	(0.0556)	(0.0554)
metropolitan	-0.0488***	-0.0369*	-0.0516***	-0.0613***	-0.0519	-0.0570**	-0.00952	-0.00529	-0.0121	-0.0712**	-0.0360	-0.0905**
	(0.0136)	(0.0221)	(0.0174)	(0.0206)	(0.0331)	(0.0267)	(0.0213)	(0.0347)	(0.0272)	(0.0338)	(0.0541)	(0.0438)
rural	0.106***	0.299***	-0.0590**	-0.00994	0.130**	-0.101**	0.0845***	0.222***	-0.0393	0.467***	0.795***	0.0789
	(0.0207)	(0.0331)	(0.0280)	(0.0312)	(0.0508)	(0.0414)	(0.0318)	(0.0499)	(0.0432)	(0.0529)	(0.0769)	(0.0806)
north	-0.129***	0.0235	-0.233***	-0.190***	-0.0156	-0.295***	-0.0540*	0.160***	-0.189***	-0.0183	-0.0952	-0.0107
	(0.0204)	(0.0330)	(0.0264)	(0.0298)	(0.0481)	(0.0388)	(0.0327)	(0.0540)	(0.0421)	(0.0544)	(0.0830)	(0.0739)
northeast	-0.107***	-0.109***	-0.118***	-0.113***	-0.119***	-0.120***	-0.0956***	-0.0785*	-0.110***	-0.0613	-0.111*	-0.0676
	(0.0168)	(0.0270)	(0.0217)	(0.0255)	(0.0405)	(0.0332)	(0.0261)	(0.0418)	(0.0337)	(0.0425)	(0.0672)	(0.0559)
south	0.129***	0.130***	0.121***	0.212***	0.267***	0.190***	0.0460	0.0525	0.0361	0.0164	-0.116	0.0779
	(0.0202)	(0.0330)	(0.0257)	(0.0319)	(0.0524)	(0.0407)	(0.0311)	(0.0505)	(0.0397)	(0.0460)	(0.0726)	(0.0600)
central west	0.0146	0.0937**	-0.0397	0.0311	0.122**	-0.0273	0.0308	0.0987*	-0.0160	-0.0134	0.0501	-0.0899
	(0.0226)	(0.0370)	(0.0289)	(0.0344)	(0.0555)	(0.0445)	(0.0353)	(0.0572)	(0.0450)	(0.0562)	(0.0901)	(0.0738)
dependent kids	-0.0142**	-0.00443	-0.0406***	-0.0520***	-0.0455**	-0.0670***	0.0142	0.0527***	-0.0298**	-0.0489*	-0.0263	-0.0758**
	(0.00706)	(0.0124)	(0.00898)	(0.00999)	(0.0178)	(0.0128)	(0.0115)	(0.0195)	(0.0150)	(0.0255)	(0.0393)	(0.0354)
retired	-1.446***	-1.795***	-1.165***	-	-	-	-	-	-	-1.173***	-1.570***	-0.858***
	(0.0272)	(0.0397)	(0.0380)	-	-	-	-	-	-	(0.0351)	(0.0547)	(0.0470)
Constant	-0.0525	-0.204***	-0.243***	-0.788***	-0.733***	-1.180***	1.369***	1.543***	1.141***	5.084***	5.700***	4.506***
	(0.0377)	(0.0585)	(0.0506)	(0.0704)	(0.106)	(0.1000)	(0.152)	(0.254)	(0.192)	(1.175)	(1.912)	(1.523)
Mills Lambda	-0.174***	-0.275***	-0.189***	-0.152***	-0.155*	-0.193***	-0.234***	-0.126***	-0.203***	-0.227**	-0.366*	-0.271**
	(0.0251)	(0.0720)	(0.0312)	(0.0346)	(0.0898)	(0.0454)	(0.0378)	(0.119)	(0.0450)	(0.0894)	(0.187)	(0.122)
N	55,120	25,954	29,166	23,404	10,981	12,423	23,284	11,234	12,050	8,432	3,739	4,693

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Source: Elaborate by the author.

Table 6: IV-Probit IV-GMM Income Equation Coefficients.

	All Ages			Young Adults			Middle-Aged			Older Adults		
y_i	All	Men	Women	All	Men	Women	All	Men	Women	All	Men	Women
pâp	0.102***	0.101***	0.0950***	0.0469***	0.0444*	0.0434	0.170***	0.172***	0.149***	0.153*	0.111	0.0774
	(0.0142)	(0.0197)	(0.0189)	(0.0182)	(0.0242)	(0.0266)	(0.0241)	(0.0409)	(0.0283)	(0.0813)	(0.0886)	(0.0936)
sex	-0.183***	-	-	-0.140***	-	-	-0.213***	-	-	-0.0542	-	-
	(0.0150)	-	-	(0.0207)	-	-	(0.0217)	-	-	(0.0909)	-	-
schooling	0.0197***	0.0286***	0.00481	0.0150***	0.0200***	0.00461	0.0173***	0.0258***	0.00439	0.0105	0.0231*	-0.0112
	(0.00287)	(0.00374)	(0.00483)	(0.00573)	(0.00674)	(0.0114)	(0.00427)	(0.00538)	(0.00688)	(0.00973)	(0.0119)	(0.0180)
other_income	0.0892***	0.170***	0.0398	-0.00175	0.100	-0.0302	0.131***	0.207***	0.0774*	0.206*	0.230	0.0920
	(0.0245)	(0.0538)	(0.0287)	(0.0363)	(0.0884)	(0.0396)	(0.0330)	(0.0618)	(0.0420)	(0.117)	(0.211)	(0.138)
metropolitan	0.161***	0.117***	0.221***	0.138***	0.0914***	0.196***	0.167***	0.135***	0.214***	0.224***	0.134***	0.304***
	(0.0102)	(0.0133)	(0.0156)	(0.0147)	(0.0184)	(0.0234)	(0.0146)	(0.0196)	(0.0221)	(0.0359)	(0.0460)	(0.0543)
rural	-0.161***	-0.159***	-0.156***	-0.151***	-0.136***	-0.169***	-0.155***	-0.136***	-0.178***	-0.257***	-0.369***	0.0291
	(0.0162)	(0.0218)	(0.0264)	(0.0218)	(0.0268)	(0.0381)	(0.0250)	(0.0339)	(0.0381)	(0.0719)	(0.117)	(0.112)
school_9	0.0788***	0.0704***	0.0942***	0.0846***	0.0859***	0.0890***	0.0762***	0.0662***	0.0954***	0.128***	0.107***	0.169***
	(0.00494)	(0.00603)	(0.00833)	(0.00808)	(0.00982)	(0.0144)	(0.00733)	(0.00879)	(0.0122)	(0.0181)	(0.0239)	(0.0296)
indigenous	-0.154**	-0.186**	-0.110	-0.0175	-0.0283	-0.00808	-0.340***	-0.394***	-0.267	0.0577	0.200	-0.0121
	(0.0675)	(0.0831)	(0.116)	(0.0800)	(0.0918)	(0.152)	(0.101)	(0.118)	(0.173)	(0.175)	(0.263)	(0.267)
yellow	0.230***	0.228***	0.243*	0.116	0.142	0.0871	0.374***	0.310**	0.424*	-0.0532	0.0403	-0.0547
	(0.0798)	(0.0873)	(0.137)	(0.0807)	(0.108)	(0.122)	(0.141)	(0.145)	(0.252)	(0.146)	(0.186)	(0.232)
brown	-0.112***	-0.112***	-0.111***	-0.109***	-0.115***	-0.0990***	-0.116***	-0.111***	-0.122***	-0.115***	-0.0982**	-0.130**
	(0.0102)	(0.0132)	(0.0161)	(0.0147)	(0.0184)	(0.0237)	(0.0148)	(0.0202)	(0.0227)	(0.0363)	(0.0488)	(0.0565)
black	-0.109***	-0.0885***	-0.141***	-0.0932***	-0.0716**	-0.129***	-0.115***	-0.0844**	-0.163***	-0.133**	-0.137*	-0.149*
	(0.0163)	(0.0214)	(0.0248)	(0.0229)	(0.0281)	(0.0377)	(0.0246)	(0.0339)	(0.0347)	(0.0587)	(0.0767)	(0.0846)
experience	0.0200***	0.0268***	0.0113***	0.0323***	0.0455***	0.0208***	0.0246***	0.0271***	0.0240**	0.0322	0.0241	0.0431
	(0.00153)	(0.00227)	(0.00246)	(0.00457)	(0.00703)	(0.00706)	(0.00670)	(0.00915)	(0.00983)	(0.0584)	(0.0876)	(0.0903)
exp_2	-0.000235***	-0.000345***	-0.0000910*	-0.000687***	-0.00114***	-0.000287	-0.000299***	-0.000336**	-0.000285*	-0.000231	-0.000178	-0.000382
	(0.0000300)	(0.0000410)	(0.0000508)	(0.000180)	(0.000253)	(0.000299)	(0.000111)	(0.000150)	(0.000166)	(0.000613)	(0.000901)	(0.00101)
union	0.0907***	0.102***	0.0824***	0.0938***	0.117***	0.0735***	0.0878***	0.103***	0.0757***	0.0988***	0.0900*	0.132**
	(0.0112)	(0.0147)	(0.0173)	(0.0168)	(0.0211)	(0.0267)	(0.0160)	(0.0211)	(0.0242)	(0.0383)	(0.0480)	(0.0589)
CEO	0.530***	0.591***	0.485***	0.514***	0.544***	0.479***	0.546***	0.660***	0.488***	0.497***	0.580***	0.499***
	(0.0253)	(0.0333)	(0.0391)	(0.0391)	(0.0553)	(0.0538)	(0.0350)	(0.0451)	(0.0561)	(0.0857)	(0.105)	(0.147)
sciencces/art	0.535***	0.620***	0.478***	0.497***	0.505***	0.486***	0.571***	0.700***	0.509***	0.473***	0.735***	0.268**
	(0.0228)	(0.0325)	(0.0330)	(0.0315)	(0.0454)	(0.0451)	(0.0336)	(0.0465)	(0.0496)	(0.0890)	(0.129)	(0.116)
technical/secondary	0.323***	0.386***	0.281***	0.305***	0.285***	0.321***	0.353***	0.493***	0.276***	0.247***	0.431***	0.101
	(0.0207)	(0.0299)	(0.0282)	(0.0291)	(0.0386)	(0.0431)	(0.0302)	(0.0462)	(0.0384)	(0.0832)	(0.121)	(0.100)
administrative	0.0854***	0.142***	0.0409*	0.0777***	0.0753**	0.0641*	0.0813***	0.225***	0.0249	0.115	0.190*	0.131
	(0.0174)	(0.0275)	(0.0234)	(0.0241)	(0.0361)	(0.0334)	(0.0269)	(0.0447)	(0.0354)	(0.0749)	(0.112)	(0.104)
commerce services	-0.0232	0.0388	-0.0633**	0.00827	0.0207	-0.00229	-0.0562*	0.0878**	-0.143***	-0.0994	-0.0345	-0.0954
	(0.0179)	(0.0244)	(0.0261)	(0.0235)	(0.0319)	(0.0342)	(0.0289)	(0.0368)	(0.0432)	(0.0683)	(0.0970)	(0.102)
agriculture	-0.237***	-0.178***	-0.291***	-0.159***	-0.144***	-0.231***	-0.272***	-0.174***	-0.266***	-0.305***	-0.128	-0.547***
	(0.0228)	(0.0273)	(0.0544)	(0.0324)	(0.0376)	(0.0853)	(0.0332)	(0.0390)	(0.0770)	(0.0786)	(0.0976)	(0.150)
product industry	0.0579***	0.134***	-0.0864***	0.0613***	0.0882***	-0.0608*	0.0506***	0.174***	-0.0788**	0.0526	0.223***	-0.215***
	(0.0131)	(0.0176)	(0.0240)	(0.0194)	(0.0247)	(0.0352)	(0.0189)	(0.0239)	(0.0368)	(0.0464)	(0.0702)	(0.0674)
army	0.741***	0.798***	0.738***	0.620***	0.621***	0.801***	0.869***	0.980***	0.710***	0.556***	0.697***	-
	(0.0387)	(0.0413)	(0.121)	(0.0553)	(0.0583)	(0.174)	(0.0516)	(0.0554)	(0.154)	(0.158)	(0.172)	-
householder	0.0464***	0.0697***	0.0177	0.0448***	0.0986***	0.00105	0.0392**	0.0438	0.0297	0.0594	0.127**	0.0133
	(0.0112)	(0.0180)	(0.0170)	(0.0160)	(0.0233)	(0.0247)	(0.0170)	(0.0288)	(0.0246)	(0.0441)	(0.0515)	(0.0686)
spouse	0.0959***	0.0695***	0.106***	0.0525***	0.0344	0.0701***	0.112***	0.0803**	0.124***	0.156***	0.115**	0.155**
	(0.00972)	(0.0212)	(0.0156)	(0.0140)	(0.0305)	(0.0213)	(0.0152)	(0.0331)	(0.0247)	(0.0362)	(0.0576)	(0.0622)
north	-0.121***	-0.157***	-0.0590**	-0.129***	-0.135***	-0.118***	-0.126***	-0.200***	-0.0189	-0.102*	-0.120	-0.0189
	(0.0146)	(0.0188)	(0.0231)	(0.0211)	(0.0263)	(0.0344)	(0.0213)	(0.0293)	(0.0331)	(0.0563)	(0.0737)	(0.0847)
northeast	-0.345***	-0.356***	-0.318***	-0.340***	-0.338***	-0.343***	-0.352***	-0.388***	-0.293***	-0.342***	-0.324***	-0.313***
	(0.0116)	(0.0152)	(0.0182)	(0.0164)	(0.0219)	(0.0255)	(0.0172)	(0.0224)	(0.0266)	(0.0413)	(0.0509)	(0.0695)
south	0.00652	0.00543	0.00698	0.00435	0.0260	-0.0137	0.0250	0.00144	0.0545**	-0.0462	0.00208	-0.0923
	(0.0120)	(0.0165)	(0.0181)	(0.0165)	(0.0234)	(0.0254)	(0.0184)	(0.0251)	(0.0275)	(0.0414)	(0.0529)	(0.0634)
central west	0.102***	0.0999***	0.104***	0.0616***	0.0756***	0.0459	0.129***	0.102***	0.163***	0.120**	0.174***	0.0521
	(0.0138)	(0.0181)	(0.0217)	(0.0193)	(0.0250)	(0.0312)	(0.0210)	(0.0285)	(0.0324)	(0.0480)	(0.0617)	(0.0756)
Mills Lambda	0.265***	0.199	0.329***	0.208***	0.00761	0.314***	0.386***	0.414	0.369***	0.518*	0.498	0.230
	(0.0566)	(0.151)	(0.0669)	(0.0779)	(0.184)	(0.0974)	(0.0862)	(0.256)	(0.0950)	(0.326)	(0.510)	(0.346)
retired	0.110***	0.101	0.130**	-	-	-	-	-	-	0.173	0.237	0.0350
	(0.0407)	(0.0923)	(0.0532)	-	-	-	-	-	-	(0.126)	(0.277)	(0.121)
Constant	2.553***	2.416***	2.564***	2.610***	2.617***	2.585***	2.378***	2.187***	2.283***	1.738	1.819	1.863
	(0.0424)	(0.0903)	(0.0512)	(0.0585)	(0.100)	(0.0956)	(0.129)	(0.260)	(0.163)	(1.501)	(2.321)	(2.089)
N	37,063	20,912	16,151	15,734	8,781	6,953	17,356	9,728	7,628	3,973	2403	1,570
Adj. R^2	0.391	0.411	0.367	0.351	0.367	0.335	0.418	0.430	0.399	0.384	0.415	0.338

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

(a) Variable *army* dropped on the *older women* subsample due to exact identification reasons.

(b) Men's Older Adults estimation method is Probit IV-GMM.

Source: Elaborate by the author.

Table 7: Biprobit IV-GMM Income Equation Coefficients.

	All Ages			Young Adults			Middle-Aged			Older Adults		
y_i	All	Men	Women	All	Men	Women	All	Men	Women	All	Men	Women
pâp	0.123***	0.133***	0.111***	0.0634***	0.0537	0.0516*	0.186***	0.193***	0.163***	0.211*	0.489*	0.0791
	(0.0166)	(0.0317)	(0.0202)	(0.0202)	(0.0343)	(0.0273)	(0.0268)	(0.0557)	(0.0302)	(0.124)	(0.268)	(0.131)
sex	-0.170***	-	-	-0.127***	-	-	-0.208***	-	-	-0.0360	-	-
	(0.0162)	-	-	(0.0221)	-	-	(0.0225)	-	-	(0.115)	-	-
schooling	0.0190***	0.0266***	0.00441	0.0139**	0.0182***	0.00411	0.0170***	0.0256***	0.00398	0.00945	0.00630	-0.0124
	(0.00289)	(0.00399)	(0.00484)	(0.00577)	(0.00702)	(0.0114)	(0.00429)	(0.00544)	(0.00690)	(0.00994)	(0.0175)	(0.0179)
other_income	0.0994***	0.221***	0.0423	0.00772	0.141	-0.0292	0.135***	0.215***	0.0792*	0.219*	0.599*	0.0640
	(0.0248)	(0.0603)	(0.0287)	(0.0363)	(0.0928)	(0.0396)	(0.0332)	(0.0672)	(0.0421)	(0.129)	(0.319)	(0.136)
metropolitan	0.163***	0.120***	0.222***	0.140***	0.0932***	0.197***	0.168***	0.136***	0.215***	0.228***	0.151***	0.303***
	(0.0102)	(0.0135)	(0.0156)	(0.0147)	(0.0186)	(0.0234)	(0.0146)	(0.0198)	(0.0221)	(0.0371)	(0.0469)	(0.0570)
rural	-0.161***	-0.169***	-0.155***	-0.150***	-0.139***	-0.168***	-0.155***	-0.138***	-0.177***	-0.262***	-0.580***	0.0345
	(0.0162)	(0.0225)	(0.0264)	(0.0218)	(0.0269)	(0.0381)	(0.0250)	(0.0345)	(0.0381)	(0.0759)	(0.184)	(0.111)
school_9	0.0772***	0.0688***	0.0932***	0.0831***	0.0848***	0.0885***	0.0756***	0.0656***	0.0949***	0.128***	0.116***	0.171***
	(0.00497)	(0.00609)	(0.00834)	(0.00814)	(0.00989)	(0.0144)	(0.00733)	(0.00884)	(0.0122)	(0.0181)	(0.0243)	(0.0299)
indigenous	-0.152**	-0.175**	-0.110	-0.0169	-0.0168	-0.00951	-0.338***	-0.391***	-0.266	0.0498	0.0750	0.00445
	(0.0675)	(0.0836)	(0.116)	(0.0800)	(0.0920)	(0.152)	(0.101)	(0.119)	(0.173)	(0.174)	(0.240)	(0.270)
yellow	0.232***	0.239***	0.243*	0.121	0.162	0.0865	0.376***	0.310**	0.425*	-0.0681	-0.118	-0.0659
	(0.0798)	(0.0876)	(0.137)	(0.0806)	(0.109)	(0.122)	(0.141)	(0.145)	(0.252)	(0.150)	(0.230)	(0.231)
brown	-0.111***	-0.109***	-0.111***	-0.109***	-0.115***	-0.0992***	-0.115***	-0.110***	-0.122***	-0.112***	-0.0561	-0.130**
	(0.0102)	(0.0133)	(0.0161)	(0.0148)	(0.0184)	(0.0237)	(0.0148)	(0.0206)	(0.0227)	(0.0365)	(0.0526)	(0.0565)
black	-0.109***	-0.0871***	-0.141***	-0.0944***	-0.0709**	-0.129***	-0.115***	-0.0839**	-0.163***	-0.131**	-0.107	-0.145*
	(0.0164)	(0.0214)	(0.0249)	(0.0229)	(0.0281)	(0.0378)	(0.0246)	(0.0339)	(0.0347)	(0.0590)	(0.0788)	(0.0851)
experience	0.0191***	0.0248***	0.0109***	0.0308***	0.0419***	0.0205***	0.0249***	0.0278***	0.0241**	0.0347	0.0961	0.0427
	(0.00154)	(0.00245)	(0.00247)	(0.00460)	(0.00749)	(0.00707)	(0.00670)	(0.00922)	(0.00983)	(0.0590)	(0.0922)	(0.0913)
exp_2	-0.000220***	-0.000311***	-0.0000842*	-0.000651***	-0.00104***	-0.000282	-0.000301***	-0.000340**	-0.000286*	-0.000246	-0.000838	-0.000402
	(0.0000303)	(0.0000446)	(0.0000509)	(0.000181)	(0.000264)	(0.000299)	(0.000111)	(0.000150)	(0.000166)	(0.000614)	(0.000926)	(0.00101)
union	0.0904***	0.102***	0.0822***	0.0938***	0.117***	0.0734***	0.0875***	0.103***	0.0755***	0.0985**	0.0889*	0.133**
	(0.0112)	(0.0146)	(0.0173)	(0.0168)	(0.0210)	(0.0267)	(0.0160)	(0.0211)	(0.0242)	(0.0384)	(0.0483)	(0.0589)
CEO	0.528***	0.589***	0.485***	0.513***	0.544***	0.479***	0.545***	0.660***	0.487***	0.496***	0.557***	0.504***
	(0.0253)	(0.0334)	(0.0391)	(0.0391)	(0.0553)	(0.0538)	(0.0351)	(0.0453)	(0.0561)	(0.0861)	(0.107)	(0.148)
sciences/art	0.534***	0.618***	0.478***	0.495***	0.504***	0.486***	0.570***	0.700***	0.509***	0.472***	0.719***	0.269**
	(0.0228)	(0.0325)	(0.0330)	(0.0316)	(0.0454)	(0.0451)	(0.0337)	(0.0466)	(0.0496)	(0.0892)	(0.130)	(0.116)
technical/secondary	0.322***	0.385***	0.281***	0.304***	0.285***	0.321***	0.353***	0.493***	0.275***	0.246***	0.420***	0.102
	(0.0207)	(0.0299)	(0.0282)	(0.0291)	(0.0386)	(0.0432)	(0.0302)	(0.0462)	(0.0384)	(0.0836)	(0.123)	(0.101)
administrative	0.0837***	0.141***	0.0405*	0.0761***	0.0754**	0.0638*	0.0807***	0.224***	0.0247	0.114	0.174	0.133
	(0.0174)	(0.0275)	(0.0234)	(0.0244)	(0.0362)	(0.0334)	(0.0269)	(0.0447)	(0.0354)	(0.0752)	(0.115)	(0.104)
commerce services	-0.0235	0.0386	-0.0632**	0.00753	0.0209	-0.00252	-0.0559*	0.0876**	-0.143***	-0.0990	-0.0341	-0.0964
	(0.0179)	(0.0244)	(0.0261)	(0.0235)	(0.0319)	(0.0342)	(0.0289)	(0.0369)	(0.0432)	(0.0683)	(0.0964)	(0.102)
agriculture	-0.239***	-0.179***	-0.291***	-0.161***	-0.145***	-0.232***	-0.272***	-0.175***	-0.265***	-0.306***	-0.143	-0.546***
	(0.0228)	(0.0272)	(0.0544)	(0.0324)	(0.0376)	(0.0853)	(0.0332)	(0.0390)	(0.0770)	(0.0786)	(0.0972)	(0.150)
product industry	0.0567***	0.132***	-0.0865***	0.0598***	0.0881***	-0.0611*	0.0503***	0.174***	-0.0788**	0.0520	0.211***	-0.217***
	(0.0132)	(0.0176)	(0.0240)	(0.0194)	(0.0247)	(0.0352)	(0.0189)	(0.0239)	(0.0368)	(0.0464)	(0.0694)	(0.0672)
army	0.739***	0.797***	0.736***	0.619***	0.622***	0.800***	0.868***	0.978***	0.709***	0.547***	0.633***	-
	(0.0386)	(0.0412)	(0.121)	(0.0553)	(0.0581)	(0.174)	(0.0516)	(0.0556)	(0.154)	(0.159)	(0.177)	-
householder	0.0399***	0.0511**	0.0160	0.0393**	0.0885***	0.000346	0.0358**	0.0384	0.0281	0.0524	0.0430	0.0230
	(0.0114)	(0.0213)	(0.0171)	(0.0160)	(0.0246)	(0.0247)	(0.0174)	(0.0328)	(0.0247)	(0.0497)	(0.0835)	(0.0687)
spouse	0.0941***	0.0469*	0.107***	0.0517***	0.0198	0.0709***	0.111***	0.0746**	0.124***	0.153***	0.00755	0.150**
	(0.00973)	(0.0246)	(0.0156)	(0.0139)	(0.0315)	(0.0213)	(0.0152)	(0.0375)	(0.0247)	(0.0369)	(0.0955)	(0.0618)
north	-0.119***	-0.160***	-0.0567**	-0.127***	-0.135***	-0.116***	-0.125***	-0.203***	-0.0173	-0.101*	-0.0952	-0.0215
	(0.0146)	(0.0188)	(0.0232)	(0.0211)	(0.0263)	(0.0345)	(0.0213)	(0.0300)	(0.0332)	(0.0563)	(0.0752)	(0.0843)
northeast	-0.344***	-0.353***	-0.317***	-0.338***	-0.334***	-0.343***	-0.352***	-0.388***	-0.292***	-0.343***	-0.322***	-0.316***
	(0.0116)	(0.0153)	(0.0182)	(0.0164)	(0.0220)	(0.0255)	(0.0172)	(0.0224)	(0.0267)	(0.0412)	(0.0513)	(0.0696)
south	0.00331	-0.00296	0.00541	0.000386	0.0179	-0.0118	0.0240	-0.0000787	0.0538*	-0.0503	0.00888	-0.0911
	(0.0121)	(0.0174)	(0.0181)	(0.0167)	(0.0248)	(0.0255)	(0.0184)	(0.0254)	(0.0275)	(0.0429)	(0.0523)	(0.0675)
central west	0.101***	0.0942***	0.104***	0.0609***	0.0722***	0.0459	0.128***	0.0998***	0.162***	0.116**	0.144**	0.0486
	(0.0138)	(0.0184)	(0.0217)	(0.0193)	(0.0251)	(0.0312)	(0.0210)	(0.0293)	(0.0324)	(0.0484)	(0.0643)	(0.0754)
retired	0.135***	0.214*	0.136**	-	-	-	-	-	-	0.194	0.866*	-0.000519
	(0.0418)	(0.113)	(0.0534)	-	-	-	-	-	-	(0.157)	(0.520)	(0.115)
Mills Lambda	0.332***	0.417**	0.354***	0.265***	0.137	0.326***	0.422***	0.473	0.389***	0.641	1.933*	0.134
	(0.0627)	(0.199)	(0.0692)	(0.0816)	(0.213)	(0.0983)	(0.0930)	(0.315)	(0.0986)	(0.441)	(1.133)	(0.418)
Constant	2.514***	2.291***	2.555***	2.588***	2.566***	2.584***	2.345***	2.127***	2.269***	1.585	-1.088	1.979
	(0.0460)	(0.121)	(0.0517)	(0.0603)	(0.115)	(0.0956)	(0.134)	(0.311)	(0.164)	(1.587)	(2.832)	(2.167)
N	37,063	20,912	16,151	15,734	8,781	6,953	17,356	9,728	7,628	3,973	2,403	1,570
Adj. R^2	0.391	0.411	0.368	0.351	0.367	0.335	0.418	0.430	0.399	0.384	0.411	0.337

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

(a) Variable *army* dropped on the *older women* subsample due to exact identification reasons.

Source: Elaborate by the author.

Table 8: IV-Probit OLS Income Equation Coefficients.

	All Ages	Young Adults		Middle-Aged	Older Adults		
y_i	**Women**	**Men**	**Women**	**Women**	**All**	**Men**	**Women**
pâp	0.0865***	0.0347**	0.0299	0.144***	0.109**	0.0172	0.102
	(0.0154)	(0.0164)	(0.0218)	(0.0236)	(0.0457)	(0.0502)	(0.0649)
sex	-	-	-	-	-0.0857	-	-
	-	-	-	-	(0.0708)	-	-
schooling	0.00585	0.0243***	0.00859	0.00370	0.0149	0.0258**	-0.00926
	(0.00483)	(0.00680)	(0.0116)	(0.00687)	(0.00978)	(0.0123)	(0.0182)
other_income	0.0371	0.0704	-0.0309	0.0739*	0.174	0.317	0.0999
	(0.0287)	(0.0891)	(0.0397)	(0.0421)	(0.108)	(0.195)	(0.133)
metropolitan	0.224***	0.102***	0.197***	0.220***	0.235***	0.150***	0.326***
	(0.0158)	(0.0189)	(0.0236)	(0.0228)	(0.0369)	(0.0482)	(0.0564)
rural	-0.152***	-0.131***	-0.168***	-0.175***	-0.262***	-0.447***	0.0331
	(0.0264)	(0.0268)	(0.0381)	(0.0383)	(0.0696)	(0.109)	(0.113)
school_9	0.0961***	0.0828***	0.0867***	0.0997***	0.123***	0.108***	0.165***
	(0.00840)	(0.00983)	(0.0145)	(0.0125)	(0.0183)	(0.0240)	(0.0300)
indigenous	-0.113	-0.0341	-0.00833	-0.273	0.0693	0.219	-0.0218
	(0.115)	(0.0915)	(0.152)	(0.174)	(0.179)	(0.288)	(0.272)
yellow	0.307**	0.121	0.0905	0.564**	-0.0585	0.0267	-0.0884
	(0.138)	(0.109)	(0.123)	(0.252)	(0.145)	(0.187)	(0.244)
brown	-0.113***	-0.111***	-0.0924***	-0.134***	-0.102***	-0.0843*	-0.117**
	(0.0161)	(0.0185)	(0.0242)	(0.0232)	(0.0372)	(0.0490)	(0.0578)
black	-0.146***	-0.0752***	-0.127***	-0.167***	-0.148**	-0.138*	-0.163*
	(0.0249)	(0.0283)	(0.0383)	(0.0351)	(0.0590)	(0.0768)	(0.0864)
experience	0.0120***	0.0481***	0.0203***	0.0211**	0.0329	0.0424	0.0628
	(0.00247)	(0.00713)	(0.00708)	(0.00990)	(0.0583)	(0.0869)	(0.0920)
exp_2	-0.000104**	-0.00117***	-0.000245	-0.000250	-0.000236	-0.000326	-0.000567
	(0.0000510)	(0.000255)	(0.000301)	(0.000167)	(0.000616)	(0.000895)	(0.00103)
union	0.0775***	0.109***	0.0637**	0.0739***	0.0963**	0.0922*	0.131**
	(0.0174)	(0.0214)	(0.0275)	(0.0242)	(0.0385)	(0.0485)	(0.0596)
CEO	0.472***	0.506***	0.476***	0.461***	0.494***	0.551***	0.545***
	(0.0393)	(0.0580)	(0.0542)	(0.0571)	(0.0870)	(0.110)	(0.150)
sciences/art	0.463***	0.490***	0.482***	0.478***	0.477***	0.730***	0.296**
	(0.0333)	(0.0457)	(0.0457)	(0.0508)	(0.0906)	(0.137)	(0.118)
technical/secondary	0.270***	0.271***	0.314***	0.257***	0.253***	0.424***	0.112
	(0.0284)	(0.0391)	(0.0440)	(0.0391)	(0.0839)	(0.126)	(0.101)
administrative	0.0273	0.0765**	0.0539	0.00421	0.113	0.159	0.141
	(0.0239)	(0.0361)	(0.0342)	(0.0364)	(0.0756)	(0.116)	(0.104)
commerce services	-0.0506*	-0.00146	-0.00348	-0.106**	-0.104	-0.0564	-0.0709
	(0.0264)	(0.0323)	(0.0344)	(0.0448)	(0.0690)	(0.100)	(0.103)
agriculture	-0.291***	-0.161***	-0.231***	-0.277***	-0.302***	-0.147	-0.531***
	(0.0544)	(0.0379)	(0.0853)	(0.0775)	(0.0791)	(0.0985)	(0.152)
product industry	-0.0867***	0.0768***	-0.0639*	-0.0819**	0.0406	0.187**	-0.202***
	(0.0240)	(0.0249)	(0.0354)	(0.0369)	(0.0476)	(0.0742)	(0.0687)
army	0.728***	0.600***	0.802***	0.705***	0.547***	0.687***	-
	(0.121)	(0.0584)	(0.174)	(0.159)	(0.157)	(0.173)	-
householder	0.0204	0.101***	0.00233	0.0355	0.0786**	0.126**	0.0120
	(0.0171)	(0.0233)	(0.0248)	(0.0248)	(0.0399)	(0.0529)	(0.0659)
spouse	0.108***	0.0414	0.0721***	0.128***	0.163***	0.0995*	0.158**
	(0.0157)	(0.0307)	(0.0215)	(0.0248)	(0.0361)	(0.0565)	(0.0634)
north	-0.0663***	-0.126***	-0.125***	-0.0258	-0.116**	-0.127*	-0.0424
	(0.0232)	(0.0264)	(0.0347)	(0.0333)	(0.0569)	(0.0741)	(0.0860)
northeast	-0.323***	-0.344***	-0.347***	-0.305***	-0.358***	-0.340***	-0.327***
	(0.0183)	(0.0220)	(0.0258)	(0.0270)	(0.0419)	(0.0527)	(0.0712)
south	0.00404	0.0325	-0.0148	0.0450	-0.0539	0.000235	-0.115*
	(0.0181)	(0.0234)	(0.0255)	(0.0278)	(0.0408)	(0.0534)	(0.0642)
central west	0.0998***	0.0780***	0.0429	0.152***	0.117**	0.162***	0.0468
	(0.0218)	(0.0251)	(0.0314)	(0.0326)	(0.0482)	(0.0621)	(0.0765)
retired	0.122**	-	-	-	0.134	0.385	0.0596
	(0.0531)	-	-	-	(0.0977)	(0.253)	(0.112)
Mills Lambda	0.318***	-0.0880	0.299***	0.364***	0.444*	0.778*	0.344
	(0.0663)	(0.184)	(0.0976)	(0.0928)	(0.238)	(0.464)	(0.308)
Constant	2.559***	2.644***	2.563***	2.350***	1.778	1.106	1.275
	(0.0511)	(0.0993)	(0.0969)	(0.164)	(1.453)	(2.273)	(2.112)
N	16,151	8,781	6,953	7,628	3,973	2,403	1,570
Adj. R^2	0.368	0.368	0.335	0.400	0.384	0.417	0.338

Standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

(a) Variable *army* dropped on the *older women* subsample due to exact identification reasons.

(b) IV-Probit OLS model estimated only for those subsamples in which *pap* was not endogenous to y_i.

(c) Men's Older Adults estimation method is Probit OLS.

Source: Elaborate by the author.

Table 9: Heckman ML Income Equation Coefficients.

	All Ages			Young Adults			Middle-Aged			Older Adults		
y_i	All	Men	Women	All	Men	Women	All	Men	Women	All	Men	Women
pâp	0.0718***	0.0787***	0.0723***	0.0302**	0.0438**	0.0245	0.118***	0.120***	0.117***	0.0536	0.0330	0.0659
	(0.00953)	(0.0121)	(0.0153)	(0.0130)	(0.0181)	(0.0217)	(0.0147)	(0.0190)	(0.0231)	(0.0364)	(0.0464)	(0.0566)
sex	-0.207***	-	-	-0.161***	-	-	-0.245***	-	-	-0.141***	-	-
	(0.0111)	-	-	(0.0157)	-	-	(0.0166)	-	-	(0.0439)	-	-
schooling	0.0209***	0.0294***	0.00692	0.0190***	0.0337***	0.0124	0.0192***	0.0274***	0.00592	0.0161*	0.0309***	-0.00977
	(0.00285)	(0.00355)	(0.00483)	(0.00563)	(0.00744)	(0.0113)	(0.00426)	(0.00537)	(0.00692)	(0.00973)	(0.0116)	(0.0178)
school_9	0.0828***	0.0710***	0.101***	0.0848***	0.0912***	0.0908***	0.0799***	0.0663***	0.102***	0.123***	0.0995***	0.168***
	(0.00487)	(0.00601)	(0.00814)	(0.00796)	(0.0109)	(0.0144)	(0.00740)	(0.00887)	(0.0124)	(0.0182)	(0.0230)	(0.0296)
other_income	0.0681***	0.152***	0.0185	-0.0187	-0.236***	-0.0459	0.105***	0.164***	0.0554	0.125	0.162	0.0637
	(0.0229)	(0.0396)	(0.0274)	(0.0341)	(0.0801)	(0.0381)	(0.0316)	(0.0484)	(0.0407)	(0.0931)	(0.132)	(0.116)
experience	0.0216***	0.0274***	0.0144***	0.0344***	0.0730***	0.0243***	0.0224***	0.0252***	0.0202**	0.0234	0.0127	0.0540
	(0.00135)	(0.00173)	(0.00217)	(0.00405)	(0.00579)	(0.00648)	(0.00671)	(0.00913)	(0.00991)	(0.0569)	(0.0777)	(0.0902)
exp_2	-0.000265***	-0.000357***	-0.000146***	-0.000706***	-0.00187***	-0.000312	-0.000287***	-0.000336**	-0.000246	-0.000171	-0.0000427	-0.000500
	(0.0000271)	(0.0000337)	(0.0000459)	(0.000174)	(0.000239)	(0.000296)	(0.000111)	(0.000150)	(0.000167)	(0.000607)	(0.000811)	(0.00101)
householder	0.0594***	0.0756***	0.0317*	0.0522***	0.164***	0.0116	0.0614***	0.0672***	0.0477**	0.0948**	0.150***	0.0273
	(0.00969)	(0.0130)	(0.0163)	(0.0141)	(0.0214)	(0.0238)	(0.0148)	(0.0197)	(0.0239)	(0.0369)	(0.0475)	(0.0621)
spouse	0.101***	0.0786***	0.104***	0.0539***	0.135***	0.0671***	0.118***	0.107***	0.123***	0.167***	0.137***	0.152**
	(0.00953)	(0.0136)	(0.0156)	(0.0138)	(0.0225)	(0.0215)	(0.0152)	(0.0209)	(0.0246)	(0.0355)	(0.0442)	(0.0617)
indigenous	-0.159**	-0.188**	-0.116	-0.0179	-0.106	-0.00553	-0.359***	-0.404***	-0.291*	0.0764	0.219	-0.00403
	(0.0670)	(0.0829)	(0.114)	(0.0791)	(0.119)	(0.151)	(0.100)	(0.118)	(0.173)	(0.185)	(0.281)	(0.275)
yellow	0.253***	0.217**	0.305**	0.107	0.0161	0.0924	0.432***	0.315**	0.560**	-0.0341	0.0414	-0.0989
	(0.0802)	(0.0880)	(0.138)	(0.0812)	(0.136)	(0.122)	(0.142)	(0.146)	(0.251)	(0.147)	(0.192)	(0.245)
brown	-0.112***	-0.113***	-0.110***	-0.102***	-0.115***	-0.0888***	-0.122***	-0.118***	-0.132***	-0.106***	-0.0955**	-0.118**
	(0.0102)	(0.0131)	(0.0161)	(0.0148)	(0.0206)	(0.0239)	(0.0150)	(0.0195)	(0.0232)	(0.0370)	(0.0466)	(0.0572)
black	-0.111***	-0.0918***	-0.141***	-0.0921***	-0.0892***	-0.122***	-0.118***	-0.0895***	-0.161***	-0.153***	-0.154**	-0.159*
	(0.0164)	(0.0215)	(0.0248)	(0.0228)	(0.0309)	(0.0380)	(0.0247)	(0.0340)	(0.0350)	(0.0585)	(0.0752)	(0.0856)
union	0.0897***	0.103***	0.0781***	0.0873***	0.120***	0.0635**	0.0879***	0.104***	0.0748***	0.0978**	0.0940*	0.133**
	(0.0112)	(0.0147)	(0.0174)	(0.0169)	(0.0201)	(0.0274)	(0.0159)	(0.0212)	(0.0242)	(0.0384)	(0.0482)	(0.0593)
CEO	0.529***	0.589***	0.475***	0.497***	0.446***	0.479***	0.544***	0.666***	0.463***	0.500***	0.553***	0.551***
	(0.0252)	(0.0335)	(0.0393)	(0.0398)	(0.0555)	(0.0541)	(0.0349)	(0.0446)	(0.0571)	(0.0861)	(0.109)	(0.148)
sciences/art	0.531***	0.619***	0.463***	0.491***	0.440***	0.484***	0.565***	0.706***	0.478***	0.479***	0.734***	0.294**
	(0.0227)	(0.0325)	(0.0333)	(0.0316)	(0.0461)	(0.0456)	(0.0336)	(0.0465)	(0.0508)	(0.0904)	(0.136)	(0.117)
technical/secondary	0.320***	0.388***	0.271***	0.292***	0.237***	0.317***	0.353***	0.502***	0.257***	0.256***	0.427***	0.114
	(0.0207)	(0.0300)	(0.0284)	(0.0294)	(0.0368)	(0.0438)	(0.0302)	(0.0463)	(0.0391)	(0.0833)	(0.126)	(0.100)
administrative	0.0822***	0.145***	0.0296	0.0730***	0.0761**	0.0575*	0.0789***	0.228***	0.00534	0.116	0.160	0.141
	(0.0175)	(0.0275)	(0.0239)	(0.0245)	(0.0324)	(0.0341)	(0.0269)	(0.0445)	(0.0364)	(0.0752)	(0.115)	(0.104)
commerce services	-0.0170	0.0355	-0.0504*	0.000396	-0.0107	-0.00132	-0.0374	0.0922**	-0.107**	-0.104	-0.0567	-0.0724
	(0.0179)	(0.0215)	(0.0264)	(0.0236)	(0.0307)	(0.0343)	(0.0294)	(0.0369)	(0.0447)	(0.0687)	(0.0999)	(0.101)
agriculture	-0.233***	-0.177***	-0.288***	-0.165***	-0.210***	-0.230***	-0.264***	-0.168***	-0.278***	-0.299***	-0.144	-0.527***
	(0.0228)	(0.0273)	(0.0544)	(0.0324)	(0.0420)	(0.0851)	(0.0332)	(0.0390)	(0.0773)	(0.0789)	(0.0980)	(0.151)
product industry	0.0612***	0.135***	-0.0864***	0.0573***	0.0675***	-0.0616*	0.0578***	0.179***	-0.0828**	0.0436	0.189**	-0.204***
	(0.0131)	(0.0176)	(0.0240)	(0.0194)	(0.0237)	(0.0353)	(0.0189)	(0.0240)	(0.0368)	(0.0474)	(0.0737)	(0.0680)
army	0.744***	0.802***	0.734***	0.607***	0.566***	0.809***	0.876***	0.996***	0.711***	0.559***	0.688***	-
	(0.0387)	(0.0412)	(0.121)	(0.0554)	(0.0623)	(0.174)	(0.0516)	(0.0552)	(0.158)	(0.157)	(0.173)	-
metropolitan	0.164***	0.120***	0.222***	0.144***	0.0828***	0.195***	0.168***	0.131***	0.219***	0.228***	0.150***	0.324***
	(0.0102)	(0.0133)	(0.0157)	(0.0148)	(0.0200)	(0.0235)	(0.0148)	(0.0194)	(0.0228)	(0.0362)	(0.0480)	(0.0549)
rural	-0.157***	-0.154***	-0.158***	-0.149***	-0.102***	-0.176***	-0.150***	-0.124***	-0.179***	-0.241***	-0.358***	0.0392
	(0.0162)	(0.0206)	(0.0263)	(0.0217)	(0.0312)	(0.0373)	(0.0250)	(0.0325)	(0.0384)	(0.0648)	(0.0785)	(0.111)
north	-0.127***	-0.155***	-0.0773***	-0.133***	-0.136***	-0.139***	-0.128***	-0.190***	-0.0345	-0.115**	-0.138*	-0.0475
	(0.0145)	(0.0188)	(0.0227)	(0.0207)	(0.0290)	(0.0332)	(0.0213)	(0.0279)	(0.0330)	(0.0568)	(0.0733)	(0.0849)
northeast	-0.353***	-0.359***	-0.331***	-0.349***	-0.376***	-0.356***	-0.360***	-0.390***	-0.310***	-0.358***	-0.345***	-0.330***
	(0.0115)	(0.0150)	(0.0179)	(0.0162)	(0.0239)	(0.0252)	(0.0172)	(0.0225)	(0.0267)	(0.0417)	(0.0526)	(0.0700)
south	0.00962	0.00831	0.0113	0.00941	0.0830***	-0.00506	0.0286	0.0102	0.0495*	-0.0477	-0.00532	-0.110*
	(0.0118)	(0.0156)	(0.0179)	(0.0160)	(0.0229)	(0.0247)	(0.0183)	(0.0244)	(0.0278)	(0.0407)	(0.0529)	(0.0631)
central west	0.102***	0.102***	0.100***	0.0602***	0.0986***	0.0427	0.132***	0.113***	0.154***	0.122**	0.170***	0.0455
	(0.0138)	(0.0178)	(0.0218)	(0.0194)	(0.0270)	(0.0314)	(0.0209)	(0.0270)	(0.0327)	(0.0480)	(0.0617)	(0.0757)
retired	0.0653*	0.0630	0.0921*	-	-	-	-	-	-	0.0612	0.120	0.0178
	(0.0344)	(0.0491)	(0.0494)	-	-	-	-	-	-	(0.0560)	(0.0832)	(0.0764)
Constant	2.764***	2.589***	2.780***	2.748***	2.045***	2.726***	2.763***	2.580***	2.670***	2.459*	2.488	1.808
	(0.0316)	(0.0424)	(0.0496)	(0.0562)	(0.0771)	(0.105)	(0.103)	(0.144)	(0.146)	(1.340)	(1.861)	(2.004)
athrho	-0.117***	-0.0986**	-0.153***	-0.105***	1.102***	-0.149***	-0.151***	-0.107***	-0.188***	-0.156**	-0.201**	-0.131
	(0.0213)	(0.0388)	(0.0264)	(0.0302)	(0.0655)	(0.0348)	(0.0267)	(0.0350)	(0.0381)	(0.0652)	(0.0895)	(0.0896)
lnsigma	-0.346***	-0.368***	-0.321***	-0.418***	-0.340***	-0.365***	-0.328***	-0.342***	-0.316***	-0.195***	-0.208***	-0.194***
	(0.00959)	(0.0122)	(0.0152)	(0.0168)	(0.0261)	(0.0255)	(0.0130)	(0.0157)	(0.0218)	(0.0254)	(0.0350)	(0.0367)
Mills Lambda	-0.0823	-0.0681	-0.110	-0.0687	0.570	-0.103	-0.108	-0.0756	-0.136	-0.127	-0.161	-0.107
N	55,120	25,954	29,166	23,404	10,981	12,423	23,284	11,234	12,050	8,432	3,739	4,693
N_censored	18,057	5,042	13,015	7,670	2,200	5,470	5,928	1,506	4,422	4,459	1,336	3,123

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

(a) Variable *army* dropped on the *older women* subsample due to exact identification reasons.

(b) *athrho* stands for the inverse hyperbolic tangent of ρ (the correlation between the error terms of the equations); and *lnsigma* stands for the natural logarithm of σ (the standard error of the residuals of the income equation).

(c) $\lambda = \rho\sigma$. Mills Lambda is not estimated, so that it does not present statistical significance tests.

Source: Elaborate by the author.

Table 10: Heckman Two-Step Income Equation Coefficients.

	All Ages			Young Adults			Middle-Aged			Older Adults		
y_i	**All**	Men	Women	**All**	Men	Women	**All**	Men	Women	**All**	Men	Women
pâp	0.0670***	0.0698***	0.0653***	0.0229**	0.0292**	0.0130	0.112***	0.0997***	0.123***	0.0466	0.0448	0.0289
	(0.00793)	(0.0106)	(0.0123)	(0.0112)	(0.0143)	(0.0179)	(0.0120)	(0.0168)	(0.0181)	(0.0292)	(0.0392)	(0.0454)
sex	-0.180***	-	-	-0.160***	-	-	-0.193***	-	-	-0.124**	-	-
	(0.0113)	-	-	(0.0156)	-	-	(0.0166)	-	-	(0.0500)	-	-
schooling	0.0189***	0.0263***	0.00421	0.0147***	0.0208***	-0.00134	0.0164***	0.0258***	0.00125	0.0264***	0.0328***	0.0187
	(0.00231)	(0.00293)	(0.00397)	(0.00485)	(0.00571)	(0.00942)	(0.00339)	(0.00427)	(0.00576)	(0.00869)	(0.0108)	(0.0152)
school_9	0.0869***	0.0769***	0.104***	0.0911***	0.0875***	0.102***	0.0861***	0.0740***	0.109***	0.117***	0.104***	0.137***
	(0.00381)	(0.00484)	(0.00622)	(0.00655)	(0.00804)	(0.0114)	(0.00559)	(0.00732)	(0.00893)	(0.0145)	(0.0188)	(0.0236)
other_income	0.105***	0.244***	0.0646**	0.0280	0.159*	0.0130	0.142***	0.278***	0.0906**	0.198**	0.319**	0.150
	(0.0212)	(0.0473)	(0.0265)	(0.0324)	(0.0826)	(0.0378)	(0.0290)	(0.0584)	(0.0379)	(0.0835)	(0.137)	(0.121)
experience	0.0206***	0.0229***	0.0156***	0.0306***	0.0360***	0.0240***	0.0232***	0.0268***	0.0208**	0.0214	0.0185	0.0309
	(0.00124)	(0.00197)	(0.00195)	(0.00372)	(0.00613)	(0.00568)	(0.00532)	(0.00731)	(0.00811)	(0.0420)	(0.0550)	(0.0688)
exp_2	-0.000246***	-0.000286***	-0.000159***	-0.000615***	-0.000822***	-0.000378	-0.000268***	-0.000290**	-0.000242*	-0.0000957	-0.0000662	-0.000103
	(0.0000238)	(0.0000349)	(0.0000392)	(0.000153)	(0.000221)	(0.000253)	(0.0000894)	(0.000121)	(0.000140)	(0.000451)	(0.000581)	(0.000760)
householder	0.0396***	0.0411**	0.00958	0.0389***	0.0721***	-0.00506	0.0431***	0.0307	0.0255	0.0151	0.0679	-0.0643
	(0.00946)	(0.0161)	(0.0144)	(0.0132)	(0.0217)	(0.0202)	(0.0142)	(0.0228)	(0.0217)	(0.0369)	(0.0483)	(0.0624)
spouse	0.0926***	0.0591***	0.0930***	0.0547***	0.0188	0.0679***	0.107***	0.0716***	0.103***	0.139***	0.0951**	0.144**
	(0.00799)	(0.0140)	(0.0138)	(0.0115)	(0.0204)	(0.0194)	(0.0125)	(0.0204)	(0.0216)	(0.0304)	(0.0408)	(0.0588)
indigenous	-0.128**	-0.180**	-0.0556	-0.0882	-0.111	-0.0419	-0.222**	-0.291**	-0.135	0.0590	0.149	-0.000706
	(0.0620)	(0.0809)	(0.0970)	(0.0872)	(0.114)	(0.136)	(0.0969)	(0.123)	(0.162)	(0.196)	(0.292)	(0.270)
yellow	0.182***	0.176**	0.197**	0.0262	0.135	-0.0521	0.376***	0.250*	0.458***	-0.0452	0.0134	-0.0523
	(0.0581)	(0.0852)	(0.0804)	(0.0850)	(0.123)	(0.119)	(0.0866)	(0.131)	(0.118)	(0.190)	(0.273)	(0.268)
brown	-0.115***	-0.112***	-0.117***	-0.103***	-0.103***	-0.103***	-0.124***	-0.115***	-0.129***	-0.112***	-0.112***	-0.109**
	(0.00880)	(0.0116)	(0.0137)	(0.0126)	(0.0163)	(0.0197)	(0.0132)	(0.0181)	(0.0202)	(0.0308)	(0.0390)	(0.0506)
black	-0.112***	-0.0951***	-0.137***	-0.0751***	-0.0362	-0.133***	-0.140***	-0.137***	-0.146***	-0.140***	-0.148**	-0.126
	(0.0133)	(0.0174)	(0.0210)	(0.0191)	(0.0242)	(0.0307)	(0.0198)	(0.0266)	(0.0304)	(0.0489)	(0.0638)	(0.0783)
union	0.0974***	0.105***	0.0965***	0.0837***	0.101***	0.0671***	0.101***	0.110***	0.0990***	0.116***	0.109***	0.148**
	(0.00949)	(0.0122)	(0.0151)	(0.0144)	(0.0186)	(0.0227)	(0.0136)	(0.0177)	(0.0215)	(0.0325)	(0.0388)	(0.0581)
CEO	0.559***	0.607***	0.530***	0.518***	0.532***	0.500***	0.583***	0.664***	0.545***	0.532***	0.606***	0.578***
	(0.0192)	(0.0257)	(0.0298)	(0.0298)	(0.0403)	(0.0449)	(0.0269)	(0.0363)	(0.0421)	(0.0680)	(0.0877)	(0.111)
sciences/art	0.570***	0.651***	0.518***	0.551***	0.536***	0.564***	0.597***	0.705***	0.528***	0.492***	0.827***	0.279***
	(0.0168)	(0.0256)	(0.0232)	(0.0247)	(0.0366)	(0.0346)	(0.0243)	(0.0380)	(0.0335)	(0.0625)	(0.0938)	(0.0855)
technical/secondary	0.318***	0.387***	0.274***	0.295***	0.287***	0.310***	0.337***	0.463***	0.256***	0.304***	0.475***	0.199**
	(0.0164)	(0.0230)	(0.0240)	(0.0230)	(0.0316)	(0.0343)	(0.0246)	(0.0354)	(0.0353)	(0.0614)	(0.0837)	(0.0927)
administrative	0.0952***	0.142***	0.0640***	0.0694***	0.0592*	0.0726***	0.109***	0.211***	0.0592*	0.199***	0.251**	0.225**
	(0.0152)	(0.0247)	(0.0200)	(0.0201)	(0.0321)	(0.0267)	(0.0246)	(0.0404)	(0.0320)	(0.0658)	(0.0974)	(0.0908)
commerce services	-0.0358**	0.0169	-0.0632***	-0.00960	-0.0239	0.00514	-0.0569***	0.0622*	-0.128***	-0.130***	-0.0185	-0.155**
	(0.0141)	(0.0212)	(0.0193)	(0.0196)	(0.0292)	(0.0269)	(0.0218)	(0.0329)	(0.0298)	(0.0495)	(0.0709)	(0.0706)
agriculture	-0.230***	-0.176***	-0.292***	-0.187***	-0.164***	-0.309***	-0.250***	-0.174***	-0.282***	-0.280***	-0.146**	-0.345***
	(0.0183)	(0.0222)	(0.0403)	(0.0283)	(0.0331)	(0.0653)	(0.0265)	(0.0330)	(0.0565)	(0.0575)	(0.0690)	(0.127)
product industry	0.0610***	0.134***	-0.0747***	0.0507***	0.0708***	-0.0406	0.0639***	0.170***	-0.0831**	0.0508	0.225***	-0.173**
	(0.0119)	(0.0157)	(0.0238)	(0.0174)	(0.0221)	(0.0363)	(0.0176)	(0.0238)	(0.0350)	(0.0402)	(0.0535)	(0.0724)
army	0.773***	0.817***	0.873***	0.655***	0.624***	0.979***	0.899***	0.996***	0.813***	0.495**	0.643***	-
	(0.0364)	(0.0388)	(0.118)	(0.0502)	(0.0527)	(0.165)	(0.0522)	(0.0565)	(0.162)	(0.251)	(0.250)	-
metropolitan	0.147***	0.107***	0.198***	0.120***	0.0825***	0.164***	0.156***	0.122***	0.203***	0.210***	0.142***	0.292***
	(0.00822)	(0.0111)	(0.0124)	(0.0118)	(0.0156)	(0.0181)	(0.0122)	(0.0169)	(0.0182)	(0.0294)	(0.0390)	(0.0454)
rural	-0.168***	-0.167***	-0.178***	-0.154***	-0.144***	-0.170***	-0.171***	-0.157***	-0.208***	-0.218***	-0.314***	-0.0698
	(0.0143)	(0.0188)	(0.0242)	(0.0204)	(0.0254)	(0.0348)	(0.0213)	(0.0283)	(0.0354)	(0.0539)	(0.0820)	(0.0997)
north	-0.106***	-0.138***	-0.0566***	-0.120***	-0.122***	-0.113***	-0.102***	-0.173***	-0.0221	-0.0844*	-0.106*	-0.00645
	(0.0124)	(0.0160)	(0.0199)	(0.0177)	(0.0224)	(0.0288)	(0.0185)	(0.0253)	(0.0291)	(0.0446)	(0.0561)	(0.0760)
northeast	-0.334***	-0.340***	-0.309***	-0.331***	-0.327***	-0.325***	-0.341***	-0.362***	-0.301***	-0.331***	-0.328***	-0.297***
	(0.0104)	(0.0140)	(0.0161)	(0.0150)	(0.0198)	(0.0233)	(0.0155)	(0.0211)	(0.0236)	(0.0370)	(0.0481)	(0.0605)
south	-0.00997	-0.00331	-0.0205	-0.00539	0.0223	-0.0400	0.00741	-0.00927	0.0268	-0.0740*	-0.00364	-0.145**
	(0.0117)	(0.0157)	(0.0179)	(0.0171)	(0.0231)	(0.0265)	(0.0174)	(0.0237)	(0.0262)	(0.0393)	(0.0523)	(0.0616)
central west	0.123***	0.120***	0.121***	0.0811***	0.0916***	0.0674**	0.150***	0.128***	0.168***	0.151***	0.192***	0.0993
	(0.0131)	(0.0174)	(0.0202)	(0.0189)	(0.0247)	(0.0296)	(0.0194)	(0.0266)	(0.0293)	(0.0461)	(0.0586)	(0.0760)
retired	0.145***	0.269***	0.134***	-	-	-	-	-	-	0.147**	0.306*	0.134
	(0.0322)	(0.0725)	(0.0453)	-	-	-	-	-	-	(0.0717)	(0.158)	(0.0915)
Constant	2.817***	2.739***	2.830***	2.846***	2.749***	2.915***	2.769***	2.636***	2.716***	2.389**	2.328*	2.002
	(0.0327)	(0.0636)	(0.0534)	(0.0604)	(0.103)	(0.112)	(0.0828)	(0.120)	(0.122)	(0.982)	(1.297)	(1.564)
Mills Lambda	-0.174***	-0.275***	-0.189***	-0.152***	-0.155*	-0.193***	-0.234***	-0.426***	-0.203***	-0.227**	-0.366*	-0.271**
	(0.0251)	(0.0720)	(0.0312)	(0.0346)	(0.0898)	(0.0454)	(0.0378)	(0.119)	(0.0450)	(0.0894)	(0.187)	(0.122)
N	55,120	25,954	29,166	23,404	10,981	12,423	23,284	11,234	12,050	8,432	3,739	4,693
N_censored	18,057	5,042	13,015	7,670	2,200	5,470	5,928	1,506	4,422	4,459	1,336	3,123

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

(a) Variable *army* dropped on the *older women* subsample due to exact identification reasons.

Source: Elaborate by the author.

Appendix C

Table 11: Hausman *"by-hand"* Test Results: η significance test

		Whole Sample			Attending School		
		All	Boys	Girls	All	Boys	Girls
NBRM	η	-0.1065	-0.0425	-0.1651	-0.1011*	-0.0750*	-0.1243
ZINB	η - NB	-0.0334	-0.0103	-0.0624	-0.0319	-0.0774	-0.0112
	η - Logit	0.2570	0.1838	0.6657	0.4858	0.0460	0.8253

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

η is the error term of the PAP binary equation estimation.

Source: Elaborate by the author, from PNAD dataset.

Table 12: Full Sample AGG Estimations: Coefficients and Marginal Effects of Selected Models.

	Whole Sample					
AGG	**LR**	**NBRM**		**ZINB**		
	Coeff/ME	**Coefficient**	**ME**	**Coeff - NB**	**Coeff - inflate**	**ME**
pap	-0.0495	-0.0496	-0.0562	-0.146**	-0.621**	-0.0901
sex	-0.457***	-0.328***	-0.371***	-0.276***	0.393	-0.412***
chores hours	0.0002	0.0007	0.0008	0.0008	0.002	0.0006
indigenous	0.903*	0.325	0.434	0.215	-36.66***	0.550
yellow	-0.233	-0.331	-0.319	-0.217	0.181	-0.273
brown	0.178**	0.169**	0.192**	0.111	-0.206	0.173**
black	0.439***	0.315***	0.409***	0.138	-1.149	0.321**
working hours	0.00761**	0.00548***	0.00620***	0.00552**	0.00118	0.00689***
mom_household	-0.572***	-0.391***	-0.504***	-0.247***	1.170**	-0.498***
householder_school	-0.0831***	-0.0687***	-0.0777***	-0.0480***	0.0935**	-0.0755***
dependent kids	0.131**	0.0968**	0.109**	0.115***	0.204	0.117**
ln_income	-0.201***	-0.147***	-0.166***	-0.0720***	0.697***	-0.197***
north	0.449***	0.318***	0.410***	0.217**	-0.602	0.394***
northeast	0.192*	0.166**	0.194**	0.0845	-0.432	0.173*
south	0.0492	0.0562	0.0649	-0.00717	-0.308	0.0331
central west	0.0907	0.104	0.123	-0.147	-1.667***	-0.0427
metropolitan	0.0424	0.0492	0.0562	0.0854	0.156	0.0864
rural	0.138	0.0723	0.0837	0.0128	-0.715	0.107
N	3,402	3,402	3,402	3,402	3,402	3,402

(a) Coefficients and Marginal effects.

(b) Discrete change of dummy variable from 0 to 1.

(c) All MEs estimated at sample means.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Source: Elaborate by the author, from PNAD dataset.

Table 13: Attending School Subsample AGG Estimations: Coefficients and Marginal Effects of Selected Models.

AGG	Attending School					
	LR	NBRM		ZINB		
	Coeff/ME	Coefficient	ME	Coeff - NB	Coeff - inflate	ME
pap	0.0229	0.0111	0.0109	-0.0545	-0.475	0.0024
sex	-0.324***	-0.267***	-0.262***	-0.256***	0.168	-0.299***
chores hours	-0.0023	-0.0019	-0.0019	-0.0016	0.0001	-0.0017
indigenous	1.361***	0.526**	0.676*	0.476**	-18.80***	0.904**
yellow	-0.246	-0.243	-0.212	0.141	0.893	-0.0145
brown	0.182**	0.188**	0.185**	0.146	-0.114	0.174**
black	0.340***	0.304***	0.340***	0.117	-1.393	0.259*
working hours	0.0041	0.0031	0.0030	0.0030	-0.0002	0.0033
mom_household	-0.357***	-0.305***	-0.334***	-0.185*	1.060	-0.330***
householder_school	-0.0732***	-0.0684***	-0.0671***	-0.0562***	0.0437	-0.0666***
dependent kids	0.118**	0.108**	0.106**	0.132***	0.274	0.109**
ln_income	-0.141***	-0.128***	-0.125***	-0.0368	0.840**	-0.146***
north	0.354***	0.332***	0.375***	0.201*	-0.818	0.331***
northeast	0.263***	0.263***	0.272***	0.149	-0.654	0.246**
south	0.1000	0.151	0.157	0.0927	-0.298	0.141
central west	0.107	0.175*	0.186*	-0.0982	-1.971*	0.0253
metropolitan	0.103	0.125*	0.126*	0.162*	0.199	0.154*
rural	0.213*	0.143*	0.147*	0.105	-0.542	0.182*
N	2,865	2,865	2,865	2,865	2,865	2,865

(a) Coefficients and Marginal effects.

(b) Discrete change of dummy variable from 0 to 1.

(c) All MEs estimated at sample means.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Source: Elaborate by the author, from PNAD dataset.

Table 14: Boys Full Sample AGG Estimations: Coefficients and Marginal Effects of Selected Models.

Boys - Full Sample						
AGG	**LR**	**NBRM**		**ZINB**		
	Coeff/ME	**Coefficient**	**ME**	**Coeff - NB**	**Coeff - inflate**	**ME**
pap	0.0512	-0.0119	-0.0155	0.0328	-0.360	0.124
chores hours	-0.0031	-0.0035	-0.0045	-0.0073	-0.0305	-0.0039
indigenous	0.0596	-0.132	-0.160	-0.0657	0.385	-0.175
yellow	-0.572	-1.096	-0.866**	0.973***	3.775***	-0.891*
brown	0.0592	0.0732	0.0948	-0.0416	-0.517	0.0507
black	0.413**	0.236**	0.337*	0.166	-0.417	0.339*
working hours	0.0134***	0.0122	0.0157	0.00264	-0.0182	0.00753
mom_household	-0.477***	-0.325***	-0.471***	-0.154	1.162	-0.432***
householder_school	-0.0999***	-0.0748***	-0.0969***	-0.0499***	0.0959**	-0.0902***
dependent kids	0.0122	-0.0007	-0.0009	0.0185	0.0391	0.0178
ln_income	-0.275***	-0.230***	-0.297***	-0.163*	0.151	-0.260***
north	0.744***	0.474***	0.747***	0.377***	-1.155	0.853***
northeast	0.272**	0.170	0.228	0.192	-0.0807	0.298*
south	0.110	0.0631	0.0836	0.141	0.111	0.183
central west	0.212	0.189*	0.265*	0.0185	-1.370	0.222
metropolitan	0.157	0.138*	0.183*	0.250**	0.564	0.227**
rural	0.306*	0.144	0.196	0.130	-0.210	0.236
mills_lambda	-	0.635	0.822	-0.581	-	-0.816
N	1,729	1,729	1,729	1,729	1,729	1,729

(a) Coefficients and Marginal effects.

(b) Discrete change of dummy variable from 0 to 1.

(c) All MEs estimated at sample means.

(d) Heckman's sample selection method applied to Boys NBRM and ZINB estimations.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Source: Elaborate by the author, from PNAD dataset.

Table 15: Boys Attending School Subsample AGG Estimations: Coefficients and Marginal Effects of Selected Models.

Boys - Attending School Subsample						
AGG	**LR**	**NBRM**		**ZINB**		
	Coeff/ME	**Coefficient**	**ME**	**Coeff - NB**	**Coeff - inflate**	**ME**
pap	0.155	0.128	0.138*	0.117	-0.0662	0.159
chores hours	-0.0063	-0.0059	-0.0066	-0.0154***	-0.0919	-0.0192***
indigenous	0.610	0.148	0.177	0.139	2.198	0.0181
yellow	-0.595*	-0.985	-0.703*	1.013***	3.995***	0.630
brown	0.109	0.106	0.118	0.00306	-0.576	0.0180
black	0.227	0.199	0.242	0.0801	-0.516	0.126
working hours	0.0085*	0.0064**	0.0072*	0.0043	-0.0134	0.0063
mom_household	-0.291**	-0.241***	-0.293**	-0.101	1.462	-0.169
householder_school	-0.0771***	-0.0619***	-0.0691***	-0.0517***	0.0523	-0.0729***
dependent kids	0.0227	0.0156	0.0175	0.0613	0.294	0.0780
ln_income	-0.236***	-0.218***	-0.244***	-0.141**	0.467	-0.207**
north	0.554***	0.466***	0.632***	0.264**	-23.17***	0.623***
northeast	0.367***	0.325***	0.389***	0.281**	-0.148	0.417**
south	0.171	0.220*	0.267	0.250	0.115	0.376
central west	0.212*	0.278**	0.349**	0.0711	-1.484	0.124
metropolitan	0.164*	0.188**	0.217**	0.257***	0.486	0.361***
rural	0.384**	0.208**	0.249*	0.231**	0.204	0.340**
N	1,484	1,484	1,484	1,484	1,484	1,484

(a) Coefficients and Marginal effects.

(b) Discrete change of dummy variable from 0 to 1.

(c) All MEs estimated at sample means.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Source: Elaborate by the author, from PNAD dataset.

Table 16: Girls Full Sample AGG Estimations: Coefficients and Marginal Effects of Selected Models.

Girls - Full Sample						
AGG	**LR**	**NBRM**		**ZINB**		
	Coeff/ME	**Coefficient**	**ME**	**Coeff - NB**	**Coeff - inflate**	**ME**
pap	-0.192*	-0.168*	-0.155*	-0.371***	-1.521**	-0.281**
chores hours	0.0033	0.0032	0.0030	0.0046	0.0082	0.0044
indigenous	2.039***	0.916***	1.418**	0.768***	-29.48***	1.529**
yellow	0.512	0.367	0.421	-0.0210	-1.802	0.0566
brown	0.299***	0.305***	0.290***	0.215**	-0.278*	0.269**
black	0.420***	0.402***	0.457**	0.120	-2.326**	0.256
working hours	-0.0076*	-0.0056	-0.0053	-0.00011	0.028	-0.0027
mom_household	-0.669***	-0.501***	-0.559***	-0.265**	2.430*	-0.480***
householder_school	-0.0663***	-0.0699***	-0.0664***	-0.0490***	0.0931	-0.0639***
dependent kids	0.246***	0.213***	0.203***	0.248***	0.370	0.247***
ln_income	-0.133***	-0.0845***	-0.0803***	-0.0268	0.832*	-0.106
north	0.148	0.112	0.111	-0.0645	-1.102*	-0.00400
northeast	0.108	0.129	0.126	-0.0785	-1.245	0.00647
south	0.0294	0.0549	0.0532	-0.0949	-0.645	-0.0585
central west	0.0809	0.106	0.105	-0.245	-2.229**	-0.166
metropolitan	-0.0824	-0.0523	-0.0492	-0.124	-0.572	-0.0908
rural	-0.159	-0.132	-0.120	-0.263**	-1.787	-0.180
N	1,673	1,673	1,673	1,673	1,673	1,673

(a) Coefficients and Marginal effects.

(b) Discrete change of dummy variable from 0 to 1.

(c) All MEs estimated at sample means.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Source: Elaborate by the author, from PNAD dataset.

Table 17: Girls Attending School Subsample AGG Estimations: Coefficients and Marginal Effects of Selected Models.

Girls - Attending School Subsample						
AGG	**LR**	**NBRM**		**ZINB**		
	Coeff/ME	**Coefficient**	**ME**	**Coeff - NB**	**Coeff - inflate**	**ME**
pap	-0.156	-0.145	-0.118	-0.313	-1.921	-0.302
chores hours	0.00238	0.00179	0.00148	0.00470	0.0164	0.00469
indigenous	2.304***	1.086***	1.617***	1.023***	-43.05***	1.777*
yellow	0.471	0.426	0.440	0.191	-1.434	0.212
brown	0.241**	0.275**	0.229**	0.201	-0.222	0.202
black	0.432***	0.450***	0.456**	0.248	-1.930	0.277
working hours	-0.0086*	-0.0091	-0.0075	-0.0069	0.0064	-0.0070
mom_household	-0.428**	-0.386***	-0.368***	-0.228	2.629	-0.249
householder_school	-0.0684***	-0.0774***	-0.0641***	-0.0619	0.110	-0.0620**
dependent kids	0.204**	0.205***	0.170***	0.243*	0.523	0.243***
ln_income	-0.0493	-0.0349	-0.0289	0.0220	0.965	0.0211
north	0.143	0.172	0.153	-0.0165	-1.520	-0.0156
northeast	0.159	0.203	0.175	-0.0288	-3.543	-0.0261
south	0.0523	0.110	0.0950	0.0201	-0.337	0.0205
central west	0.0862	0.164	0.145	-0.101	-1.971*	-0.0959
metropolitan	0.0127	0.0467	0.0390	-0.0207	-0.714	-0.0200
rural	-0.0737	-0.0539	-0.0438	-0.147	-22.18***	-0.103
N	1,381	1,381	1,381	1,381	1,381	1,381

(a) Coefficients and Marginal effects.

(b) Discrete change of dummy variable from 0 to 1.

(c) All MEs estimated at sample means.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Source: Elaborate by the author, from PNAD dataset.