U.PORTO

2º CICLO
ECONOMIA

Minimum wage growth and regional heterogeneities: An analysis of Rio de Janeiro

Fellipe Mateus Miranda da Silva

2022

MINIMUM WAGE GROWTH AND REGIONAL HETEROGENEITIES: AN ANALYSIS OF RIO DE JANEIRO

Fellipe Mateus Miranda da Silva

Dissertation

Master in Economics

Supervised by
Vitor Miguel de Sousa Ribeiro
António Neto

2022

Acknowledgements

First and foremost I would like to thank my mother, Sandra, my little brother, Ramon, and my family for the support throughout these years. I would not be able to be the first in the family to dream of becoming a master's degree holder if it was not for the sacrifices you made as I grew up, just so I could have the best life you could afford. Nor would I have the assurance that if things go awry while I am more than 8,000km away from home, I have so many people to make sure I land safely no matter how hard nor where I fall. For that alone I am forever grateful.

I would also like to thank my advisors, professor Vitor Ribeiro and Ph.D António Neto, for the extreme trust and support they gave me throughout this journey. Despite the "megalomania" of this project, they followed through and gave me guidance when I needed, as well as the belief that I could succeed in doing this. Thank you.

Lastly, I would like to thank the friends I made – those who are still with me, and also those who came and went. Especially Camila, Renato and Inês, who have gone out of their way to welcome me into the country and make sure I can call it "home" one day in the future; and Pedro Mendes, for tireless discussions on data analytics, Economics, the world, and life in general as we traveled off the beaten path to fulfil dreams of our parents, getting to places we were not supposed to reach.

Without all of you, I would not be here.

Abstract

This master thesis assesses the impact of a minimum wage policy established by Brazil's federal government from 2003 to 2010 which saw major growth in the "floor" earnings of workers in the country, and how it affected employment and economic activity at Rio de Janeiro's state capital — a very diverse metropolitan region in terms of gender and race, and also highly unequal in income and wealth. Using a difference-in-differences design inspired by Derenoncourt et al (2021)'s specification utilising strongly and weakly impacted areas given income levels pre-treatment, this research finds that minimum wage growth had positive impacts on income and hiring of workers at small and medium companies in strongly treated regions, while not having detrimental impacts in non-white people hirings. And by using principal component and factor analysis to develop measures that can capture economic activity at the microlevel, this research finds that strongly treated areas were highly influential in entrepreneurial development within the city.

Keywords: Brazil; Economic activity; Employment; Minimum wage; Regional economics; Rio de Janeiro.

JEL codes: E24; J38; R11; R23.

Resumo

Esta dissertação avalia o impacto de uma política para o salário mínimo adotada pelo governo federal do Brasil entre 2003 e 2010, responsável pelo forte crescimento do "piso" salarial de trabalhadores no país, e como esta afetou emprego e atividade econômica na capital do Rio de Janeiro – uma região metropolitana bastante diversificada em termos de gênero e raça, e também altamente desigual em renda e riqueza. Usando uma especificação "difference-in-differences" inspirada por Derenoncourt et al (2021), que determina regiões fortemente e fracamente tratadas tendo como base a renda destas regiões pré-tratamento, esta pesquisa revela que o crescimento do salário mínimo teve impacto positivo nos ganhos e na contratação de trabalhadores em pequenos e médios negócios em regiões fortemente tratadas, ao mesmo tempo que não influenciou de forma prejudicial o emprego de pessoas não-brancas. E utilizando análise de componentes e a análise fatorial para criar medidas que possam capturar atividade econômica no nível "micro", descobre-se que regiões fortemente tratadas foram altamente influenciais no desenvolvimento de empreendedorismo dentro da cidade.

Palavras-chave: Brasil; Atividade econômica; Emprego; Salário mínimo; Economia regional; Rio de Janeiro.

Códigos JEL: E24; J38; R11; R23.

Contents

List of Figures

List of Tables

1 Introduction

Brazil is a very diverse and highly unequal country which experienced major gross domestic product growth from 2000 until 2013, increasing by nearly 57% in this period (The World Bank 2021). During the same period, the federal government implemented a minimum wage growth policy that raised it from US$ 2,410 to US$ 4,883 per year in constant 2020 USD PPP prices (Organisation for Economic Co-operation and Development 2021), representing a 102.6% increase between 2000 and 2013.

But GDP growth in Brazil was not equally distributed across the country. In the 2000s, growth was stronger especially in cities in the North and Northeast regions, which are also the poorest in the country (Villela 2013). Even in its richest region, the Southeast, GDP growth was not uniform as Minas Gerais and Espírito Santo recorded stronger growth than the two leading economies in the country, São Paulo and Rio de Janeiro (Instituto Mauro Borges 2018).

That was the case with Rio de Janeiro's metropolitan area. As a very diverse and highly unequal region (Casa Fluminense 2020), the former Brazilian capital saw massive GDP growth during the 2002-2013 timeframe, increasing by 201% (Instituto Brasileiro de Geografia e Estatística 2019).

Given the positive impact of minimum wage growth on low-wage regions (Godoey and Reich 2021) and Rio de Janeiro's inequality – as well as the dynamics of wage inequality reduction in the country during the 2000s, with racial and gender heterogeneities as two explanatory factors of this mechanism (Derenoncourt et al. 2021; Komatsu and Filho 2016) –, it could be the case that one driver of the region's performance during the 2003-2013 period is similar to the minimum wage mechanism that drove wage inequality down in the rest of the country.

Thus, the main question in this research is whether the minimum wage, and so minimum wage growth policy from the federal level, can have a differential and discernible impact in local economic activity – with heterogeneity as a potential driver of such effects.

And so, this master thesis sought to find whether the minimum wage growth policy implemented in Brazil had an effect in local economic activity, and labour market outcomes of the populace. Given Rio de Janeiro's diversity, there could be differential effects of minimum wage growth depending on whether a location had more potential to be affected by the policy given its previous income level, due to its populational composition. This differential effect, caused by heterogeneities in Rio's populace in income, gender and race, could be a factor in explaining employment effects and economic performance in certain regions of the city.

With help from publicly available data for the research, this master thesis looks first at how the city of Rio de Janeiro and its surrounding metropolitan area developed in socioeconomic terms from 2000 to 2015 using Census, PNAD (*Pesquisa Nacional por Amostra de Domicílios*) and RAIS (*Relação*

Anual de Informações Sociais) data. It then assesses the effects of minimum wage increases from federal government at the microlevel using RAIS, as an employer-employee linked database, on an array of variables that can help address how impactful this policy was for Rio de Janeiro's labour market with a difference-in-differences (DiD) design based on strongly and weakly treated regions.

Minimum wage impacts are also addressed via principal component analysis (PCA) and factor analysis (FA), using RAIS and PNAD datasets. And to complete missing entries within these databases, this research utilizes predictive mean matching (PMM) combined with multivariate imputation by chained equations (MICE) to infer them.

In DiD analysis, we find that minimum wage growth had a positive impact on earnings of workers in strongly treated regions, while not having major negative impacts in hiring of both female and non-white workers. It also boosted employment at "Simples Nacional" firms, which are small and medium enterprises (SMEs) at a more favourable tax regime (Governo Federal do Brasil 2022; Receita Federal 2007).

Whereas in PCA/FA analysis, we find that strongly treated regions were highly impactful in entrepreneurship within the city. At the same time the city and its metro area also had a fair share of economic influence on the demand side from pensionists, as shown by PNAD data.

With this, this master thesis contributes to the literature of two fields: regional economics and labour economics. On the former, the dissertation identifies factors that can capture economic activity at the microlevel, to assess potential differential effects stemming from public policies, socioeconomic composition of the local economy, and other characteristics that can be important in defining the performance of a region.

On the latter, it adds to the ongoing research on minimum wage heterogenous effects, gauging how minimum wage impacts workers at the microlevel in Brazil; as well as general effects of minimum wage growth at the microlevel, with a focus on more vulnerable regions and minority groups, to identify whether there are any differential effects from such policy in populace and regions.

This master thesis proceeds as follows. Section 2 describes the literature review, covering works on minimum wage effects at national and regional levels, and economic activity indexes. Sections 3 and 4 provide context on Brazil's minimum wage policy at the federal level, as well as Rio de Janeiro's past and current economic development. Section 5 utilises several databases to describe the city's demographic and economic evolution at the domicile and individual levels. Section 6 assesses minimum wage effects on employment and economic activities. Section 7 concludes.

2 Literature review

2.1 Minimum wage and its effects

The literature on minimum wage growth effects has been recently highlighted by the Nobel prize awarded in 2021 to economist David Card, involved in the empirical study of the effects of minimum wage increase in two bordering United States' states on low-wage employment (Card and Krueger 1994). The results of this research, finding a positive effect on employment from a minimum wage hike, ran counter to general labour market theory, and gave credence to labour market models with frictions where minimum wage growth can have a positive impact on employment (Boadway and Cuff 2001; Lang and Kahn 1998).

Ever since, minimum wage effects on employment have become one of the most studied subjects in labour economics, as employment often works as a proxy for social welfare of a populace, and economic activity. A literature review by Dube (2019a), covering 55 studies on the link between minimum wage and employment in various countries – but mainly the United States and the United Kingdom – revealed a "muted effect" of the former on the latter, all while increasing earnings of low-paid workers.

This wealth in research of minimum wage effects on employment also functions to highlight not just different impact of the minimum wage in geographical terms, but also how methodology matters in assessing such effects. As per Dube, Lester, and Reich (2010), using local identification instead of broader geographical measures in assessing minimum wage impacts leads to the aforementioned effects shown in the literature review.

This can be showed by Paun et al. (2021) . By using fixed and random effects (FE & RE) estimation on country-level data, the authors identify adverse effects of minimum wage increases on employment levels of vulnerable groups in the European Union between 1999 and 2016. Meanwhile, a study by Ahlfeldt, Roth, and Seidel (2018) with difference-in-differences (DiD) methods at the regional level in Germany is in line with Dube's findings in the United States, thanks to spatial wage convergence.

But as posited by Manning (2021), this "elusive" effect of the minimum wage on employment should not be a surprise to the field at this point in time. Therefore, in this realm of research there might be need to move on from finding the signal of minimum wage on employment, toward limits of minimum wage – akin to Christl, Köppl-Turyna, and Kucsera (2019), finding a non-linear/inverted U-shaped relation between minimum wage and youth employment in 12 European Union countries between 1980 and 2011.

In Brazil, literature on minimum wage effects is also vast. Much of this research has in mind the dual labour market in Brazil, with approximately 40% of its population in informal labour market relations in 2020 (Abdala 2021).

Meghir, Narita, and Robin (2015) models and empirically tests a dual labour market with frictions, using Brazilian data from 2002 to 2007. They find that increased regulation of informal workers does not have a detrimental impact on employment, while improving workers' welfare via better wages and labour rights' coverage.

Boeri, Garibaldi, and Ribeiro (2011) uses FE and DiD specifications to show how minimum wage increases in Brazil have positive impacts on wages and educational attainment, shifting informal workers to formal labour relationships, even with the caveat of short-term adverse impact on employment. However, Jales (2018) shows via regression discontinuity design (RDD) and nonparametric kernel estimations that these minimum wage increases led to more workers moving to the informal market, which also reduces payroll revenue to the government since these workers do not pay these taxes.

Research on minimum wage effects in Brazil is also dedicated to assessing how it played an important role in diminishing the country's inequality, whose Gini index fell from 59.9 to 51.9 between 1996 and 2015 (The World Bank, Development Research Group 2021). Engbom and Moser (2021) uses a series of specifications to test their modelling on minimum wage effects in affecting inequality. They show that minimum wage hikes had spillover effects up to the 90th percentile of income in Brazil, reducing inequality while having little adverse impact on employment thanks to workers moving to companies with high productivity.

Hinojosa (2019) 's specification, using instrumental variables (IV), shows similar effects – although the author also identifies larger positive impacts on formal workers. And Brito, Foguel, and Kerstenetzky (2017), using decomposition methods, shows that nearly 65% of the reduction in intra-household inequality in Brazil between 1995 and 2014 can be attributed to the minimum wage, especially via pensions.

One important driver of this reduction in inequality in Brazil via minimum wage comes from racial dynamics. As highlighted by Derenoncourt et al. (2021), the minimum wage increases in Brazil from 2000 to 2009 had a great impact in diminishing the earnings gap between whites and non-whites in the country, with no reallocation of workers from formal to informal labour relationships nor employment effects shown during this period.

There can be a parallel drawn between this and Derenoncourt and Montialoux (2021) work. Their research shows that legislation introduced in the United States expanding minimum wage coverage to several industries with African American workers in 1966, was one of the main drivers in the reduction of the racial earnings gap at least until the early 19080s.

There is also plenty in the literature regarding minimum wage heterogeneities, with different effects found depending on geography and methodology utilized.

Callaway, Li, and Oka (2018) uses DiD with conditional quantile treatment effect on the treated to assess minimum wage effects on different gender, education and race groups in the United States. They find negative impacts on earnings for vulnerable groups such as low-educated women and African Americans. Whereas Braunstein and Seguino (2018), with research focused on Latin American countries between 1990 and 2010, finds a positive effect of higher minimum wages on female employment in the region.

Dube (2019b) finds that higher minimum wages reduce poverty rate levels in the United States, especially for vulnerable groups. And Godoey and Reich (2021) shows that minimum wage increases have higher positive impacts in low-income areas, without negative employment effects identified in the United States areas assessed during the 2005-2017 period covered by their research [1].

2.2 Economic activity indexes

It is thus somewhat clear that minimum wage effects are not uniform, nor even linear, conditional on several factors. One promising way to assess such effects might be regional analysis, utilising economic activity indexes to capture these dynamics at the microlevel.

The literature on local economic activity indexes is driven mainly by Stock and Watson (1999), using time-series data and methods to create inflation forecasts for the United States that could serve an auxiliary function to the Phillips curve. The result is an index of nearly 170 measures that combined with the Phillips curve, outperforms most traditional methods of forecasting inflation.

Their methodology was later developed by Crone and Clayton-Matthews (2005) into coincident indexes for United States' states. By combining linear transformations and calibrations with the Kalman smoother, they are able to produce proper measures of ongoing economic activity in the regions analysed even in absence of gross state product (GSP).

Vidal et al. (2017) also adapted Stock-Watson's index methodology, with Latin America in mind. They employ a three-step method to draw a coincident index for the Valle del Cauca region in Colombia, and they are able to decompose and assess which factors are most important in determining economic activity in the region.

Outside of regional economic activity, there are also its proxies. Navarro, Durán, and Bartolomé (2017) created a competitiveness index with over 60 variables, based on productive capital, human capital, and public capital, to find the best performing regions in Spain from 2001 to 2014. Whereas

[1] Support tables for the literature on minimum wage effect can be found in pages 73, 74 , 75, and 77.

Amaral et al. (2010) used market potential, defined as the sum of internal and external incomes of a locality, to check how it affects wages of a region via spatial specification, random effects and IV.

Still on assessments of regional economic activity and wages, Hacker et al. (2013) uses first-differenced vector autoregressive (DVAR) Granger causality approach to check whether concentration of population leads to economic growth, or vice-versa. Using Swedish regions between 1987 to 2006, they find a strong effect of population agglomeration leading to economic growth – assessed by average rages of a region – in urban setting, but not in rural ones.

One somewhat unexpected finding in this review, is that there is not much in the literature trying to assess the direct link between minimum wage and GDP outside of macroeconomic modelling (Cahuc and Michel 1996; Fanti and Gori 2011; Rebitzer and Taylor 1995). But there are proxies that might be able to highlight this relationship.

In Brazil, Arestis, Baltar, and Prates (2016) uses generalised method of moments (GMM) and vector error correction (VEC) to show how consumption, aided by wage income and therefore minimum wage increases, was a big driver of Brazil's economic growth before and after the 2008 financial crisis. Sectoral decomposition of gross value added (GVA) also finds that the minimum wage hikes in Brazil were the biggest determinant of individual labour income from 1996 to 2014, reigning above productivity (Katovich and Maia 2018-Jan-Apr).

Using proxies such as firm markup for economic activity, recent research using FE, IV and DiD was able to find a positive impact of minimum wages in China likely due to incentives on value-enhancing activities and efficient resource allocation (Du and Wang 2020). And in Indonesia, panel data and causality analysis unearthed a positive link from minimum wage to regional economic growth between 2003 and 2015 (Amri 2018).

For OECD countries, Askenazy (2003) models and finds via ordinary least squares (OLS) analysis of 11 nations from 1970 to 1990 a positive link between economic growth and the Kaitz index – the ratio of minimum wage to the average wage – when combined with an increase in exports. And Jung, McFarlane, and Das (2021), using autoregressive distributed lag (ARDL) methodology, estimates that a 1% increase in minimum wage led to nearly 0.5% increases in Canadian retail sales from 1991 to 2019.

Therefore, the field is rich in potential to gauge whether minimum wage has effects on economic activity. There are proxies such as firm markups, employment and consumption that could work to examine such a link, with potential to create measures that combine them proxies into a proper score for analysis with an array of techniques such as principal component analysis (PCA) and factor analysis (FA) (Dai, Xiong, and Zhou 2021; Godshalk and Timothy 1988; Lai 2003; Marques and

Neves 2001; Vyas and Kumaranayake 2006) [2].

[2] Support tables for the literature on economic activity indexes can be found in page 76

3 The minimum wage policy in Brazil

Minimum wage laws date back to 1894, starting in New Zealand and followed two years after by the Victoria state in Australia. Progress in minimum wage laws worldwide would only kickstart 15 years later, with the United Kingdom's implementation of said regulation and the start of minimum wage support targeting especially vulnerable populace – i.e., women – in the United States (International Labour Organization 2015).

In Brazil, minimum wage laws were first established in 1936, with a 1940 decree-law on Workers' Day defining and implementing wages throughout the country. Given Brazil's variance in terms of economic development across its regions, minimum wages were initially established across 50 sub-regions, with 14 different values across them (ADVFN 2018).

Minimum wage in Brazil has always been susceptible to the country's momentum, facing period of either strong growth, stagnation or strong devaluing conditioning on GDP swings and political alignment. The 1970s saw Brazil experience their "economic miracle" of high GDP growth, averaging 8.8% per year. However, minimum wage stagnated and represented a true loss of purchasing power to workers. Whereas during the 1980s, in Brazil's worst ever economic period with stagflation, the minimum wage fell to near-record lows from R$ 985 in 1982, to R$ 468 in 1990 (Instituto de Pesquisa Econômica Aplicada 2022).

3.1 From Plano Real to today

Stagflation made monetary policy the main focus of Brazil's economic guidance in the 1980s and early 1990s. The country's currency underwent several reforms during this time period, with five different currencies taking over from 1984 to 1994 to address hyperinflation.

Plano Real's implementation in 1994, transitioning from Cruzeiro Real to a dollar-parity Unidade Real de Valor, and then to a free exchange currency in the Brazilian Real, was the success required to wrest Brazil back to a path of low-inflation economic growth. Inflation fell from 2,075% in 1994 to an all-time low of 3.195% in 1998, with average GDP growth of 2.98% between 1994 and 2000.

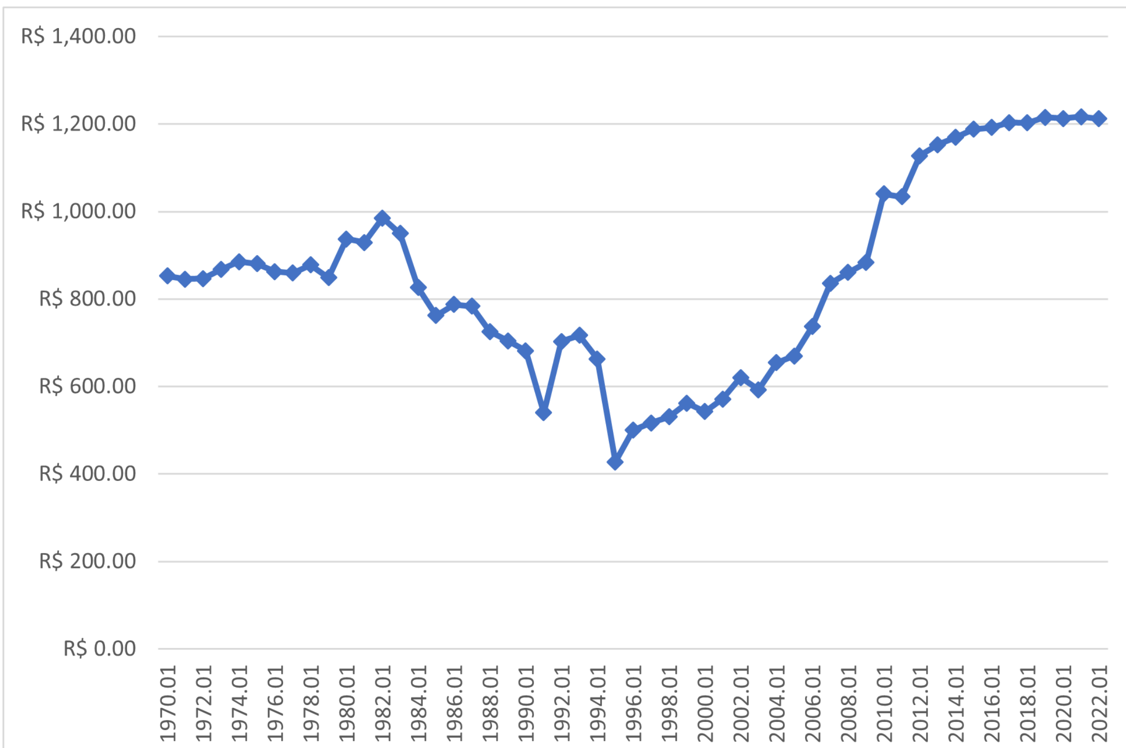
Despite strong economic growth, minimum wage lagged behind in the first years post-Plano Real implementation. The 29.8% growth between 1994 and 2002, under the presidency of Fernando Henrique Cardoso, was likely motivated by fears of a wage-price spiral that was identified as one of the main factors behind hyperinflation experienced during the 1980s and early 1990s (Lemos 2004).

Fernando Henrique Cardoso's successor, Luis Inácio Lula da Silva, had in the forefront a different distribution guidance from the federal government via minimum wage, while keeping monetary policy primacy established from his predecessor. Brazil managed to keep high GDP growth under

Lula's government, averaging 4% growth per year. Whereas minimum wage increased by nearly 80% in the eight-year period with Lula as president, between 2003 and 2010.

One of the presumed "boosts" to minimum wage law, was the establishment of an official adjustment rule in 2008 that would correct the earnings based on the previous year's inflation, and GDP growth two years prior. However, the end of Lula's government marked the end of Brazil's high economic performance, leading to a deep recession from 2014 to 2016 (Oreiro 2017-Jan-Apr). Ever since, Brazil's minimum wage is back to stagnation. Since 2011 until January 2022, minimum wage increased only 17.9% (Instituto de Pesquisa Econômica Aplicada 2022) [3].

Figure 1: Real minimum wage in Brazil from 1970 to 2022 (in January 2022 R$)



Available in: <http://www.ipeadata.gov.br/ExibeSerie.aspx?serid=37667>.

Source: IPEA. Own elaboration.

[3] Minimum wage values can be found in page 78.

4 Regional economic activity in Rio de Janeiro

4.1 A brief history of Rio de Janeiro

Rio de Janeiro was a region dominated by the Tupi populace since the beginning of the second millennium, with tamoios overtaking the location through much of its history before the arrival of Portugal to Brazil. One of the biggest “aldeias” in Rio was Carioca, which became the name given to the city’s inhabitants (Carino 2019).

Guanabara Bay was discovered by the Portuguese in a 1502 expedition. But the city of Rio would only be founded in 1565, following several battles between Portuguese and French occupants of the bay, by Estácio de Sá. It thus became Brazil’s second official city, 16 years after the foundation of São Vicente in São Paulo state.

Rio de Janeiro quickly turned into a port city and thus an economic hub, initially focused on exporting sugar from cities in its surroundings that would later become neighbourhoods or cities that are now part of its Metropolitan Area. It was also the main administrative centre of Brazil’s southern region, with Salvador being the country’s “capital” and northern regional counterpart.

The city became Brazil’s second capital in 1763, after experiencing enormous growth from being a port city to the gold extracted from the mines in Minas Gerais state. It eventually surpassed Ouro Preto to become the country’s biggest city in the late XVII century (Silva and Versiani 2015).

Rio de Janeiro became capital to the Portuguese Empire from 1808 to 1815, during the royal family’s departure from Portugal due to Napoleon Bonaparte’s threat to the kingdom. The city continued to boom after Brazil’s independence in 1822, this time thanks to coffee becoming a hot commodity throughout the world (Agência O Globo 2017).

However, as Rio de Janeiro grew, its neighbouring state went the opposite way. As a federal district, it enjoyed not just the attention and funds of federal government but also attracted some of the biggest talents in Brazil from the rest of the country. Whereas the rest of the state were mostly seen as places for extractive activities, with little to no institutional foundations to be found as Rio concentrated most of it.

In the 1950s, president Juscelino Kubitchek decided to begin the construction of a new capital to Brazil – Brasília – in the Centre-West region of the country. This marked the beginning of Rio de Janeiro’s fall from grace, with Brasília becoming the capital in 1960 and most of the federal buildings and offices moving to the new federal government centre.

During Brazil’s military dictatorship, from 1964 to 1985, Rio de Janeiro suffered with economic decline due to regional and national factors. The Guanabara federal district was dismantled and

Rio de Janeiro became capital of the Rio de Janeiro state, which carried a lot of issues from its underdevelopment in previous centuries. At the same time large swaths of the population were forced out of the city centre to move to localities which would eventually develop into the "favelas" of today. These became "hotpots" of severe social problems that have marred the region ever since.

The beginning of democratic regime in 1985 had Brazil in its worst economic stage, facing hyperinflation and recession since the late stages of the military dictatorship. Rio de Janeiro suffered a lot as well, with its decline furthered by the crisis (Hermann 2014).

Recovery started in the 1990s, with the Earth Summit held in 1992 at Rio representing its turnaround. The city invested in its cultural and touristic vocations, while the rest of the state got a boon from large oil reserves found at its coastal region, generating great amounts of royalties paid directly to the state government.

Rio de Janeiro was the stage to several big events in sports, beginning with the Pan-American Games in 2007. It was also hosts to the World Cup finals in 2014, and the Olympics in 2016. However, Brazil's most recent economic crisis from 2014 to 2016 was once again a hard hit to Rio de Janeiro. That combined with severe changes in royalty distribution from oil extraction has put the state and its government back in decline, even though its population continues to grow (Agência O Globo 2017).

4.2 The capital today

Rio de Janeiro's state capital, which has the same name, had an estimated active economic population of 2.491 million in 2019. This represented 37.1% of its total population.

According to the Pesquisa Nacional por Amostra de Domicílios (PNAD) in 2019, 54% of Rio de Janeiro's 6.719 million inhabitants were women. In terms of racial composition, 53% were white and 47% non-white.

Latest unemployment numbers for the state of Rio de Janeiro, from PNAD's survey for 2021's last quarter, sets it at 14.2% - 3.1% above the national average for the same timeframe. Meanwhile average household income fell from R$ 3,480 to 2,943 year-on-year, representing a 15.4% decrease (Instituto Brasileiro de Geografia e Estatística 2022c).

In terms of economic composition, 67.13% of Rio de Janeiro's net gross regional product (GRP) of R$ 271.8 billion in 2019 was generated via services excluding public ones. Public services were the second-most important activity in the capital, representing 19.75% of net GRP, followed by industry at 13.08%. Agriculture was the least relevant sector in Rio's economy, generating only 0.03% of its net GRP.

GRP per capita in Rio de Janeiro was R$ 52,833 in 2019, from R$ 32,920 in 2010 – a 60% increase over nine years. In the same time period, Rio's GRP grew 71% (Instituto Brasileiro de Geografia e Estatística 2022b).

Another salient aspect of Brazil and especially Rio de Janeiro, is inequality. Whereas Brazil saw a 13.35% decrease in inequality during the 2000s, the capital's Gini index grew from 0.6092 in 1991 to 0.6150 in 2000, and 0.6391 in 2010 according to Census estimates for each year (Sistema Único de Saude 2011).

Table 1: Rio de Janeiro statistics from 2010 to 2021

	Time period	**Value**
Total population	2019	6.72 million
Active economic population	2019	2.49 million
Men	2019	46%
Women	2019	54%
White	2019	53%
Non-white	2019	47%
Unemployment (entire state)	2021.12	14.20%
Average household income (entire state)	2021.12	R$ 2,943
Total net GRP (in R$ bi)	2019	R$ 271.8
Inequality (Gini)	2010	0.6391

Source: IBGE. Own elaboration.

4.3 Regional divisions of Rio

Pereira Passos Institute ("Instituto Pereira Passos" – IPP, in Portuguese) is a public research organization in Rio de Janeiro responsible with aiding the capital's prefecture in urban planning for public policies. Since 1979 it also produces and archives several reports and documents related to Rio's past and present (Instituto Municipal de Urbanismo Pereira Passos 2022).

In 1981, Rio de Janeiro underwent its last major change in geographical delimitations. The city was divided by 160 "bairros", i.e., neighbourhoods, 34 administrative regions and 5 Planning Areas ("Áreas de Planejamento" – AP, in Portuguese). The latter encompassed popular city sections, which divided the locality by West, South, North, and Central Zones given environmental, historic, geographic and economic characteristics shared by the neighbourhoods in each of the Planning Areas. Jacarepaguá – located in the West Zone of Rio – was a special region, mostly due to Barra da Tijuca's late and "guided" developments as a neighbourhood (Prefeitura do Rio de Janeiro 1981).

A further subdivision of Rio de Janeiro in 2011 established Planning Regions ("Regiões de Planejamento" – RP, in Portuguese), subdividing Planning Areas by similarities in neighbourhoods. There are now 16 of them, encompassing 162 neighbourhoods in Rio (Prefeitura do Rio de Janeiro 2011).

One of IPP's most recent developments was the creation of the Social Progress Index ("Índice de Progresso Social" – IPS, in Portuguese). This index, with a 0-100 scale, evaluates differences in urban development per region by combining human development index indicators with public services provisions and human rights' ease of access divided between three major groups: Basic Human Needs, Welfare Foundations, and Opportunities ("Indice de Progresso Social Do Rio de Janeiro" 2020).

4.3.1 Planning Area 1 – Centro

AP 1, Centro, consists of only RP, also named Centro (1.1), with 16 neighbourhoods and 6 administrative regions. Its projected population in 2018 was 307,102 people, representing 4.69% of Rio de Janeiro's entire population in that year.

In 2020, average IPS in AP 1 was 53.65, with Portuária as the less developed administrative region (42.06) and Santa Teresa as the most developed counterpart (63.26). Overall, per capita earnings in Centro amounted to R$ 28,204 per year.

From 2000 to 2010, inequality as measured by the Gini index was up by 4.25% in the region, from 0.465 to 0.486. The leading contributor was São Cristóvão, with a 9.63% increase in inequality; and Santa Tereza moving in the opposite direction, with a 10.71% decrease[4].

4.3.2 Planning Area 2 – Zona Sul e Grande Tijuca

AP 2, Zona Sul and Grande Tijuca, is made by RP 2.1 – Zona Sul – and 2.2 – Tijuca. It encompasses 25 neighbourhoods and 6 administrative regions, with 1.04 million (15.84%) residents in 2018.

Average IPS in AP 2 was 73.53 in 2020, the highest in the city despite the presence of Rocinha (52.27), one of the largest "favelas" in Latin America, in the mix. Botafogo (85.03) leads the AP and the city in IPS score. It is thus no surprise that per capita earnings in the region are also the largest in the city, at R$ 80,214 per year in 2018.

Inequality in AP 2 went from 0.466 to 0.496 between 2000 and 2010, a 5.92% increase in the 10-year period. Tijuca saw the biggest increase in Gini index (10.72%), and Rocinha was the only region where inequality decreased (-3.88%)[5].

[4] A list of Centro neigbourhoods can be found in page 79.

[5] A list of Zona Sul e Grande Tijuca neighbourhoods can be found in page 80.

4.3.3 Planning Area 3 – Zona Norte

AP 3, Zona Norte, has seven different RPs: Ramos (3.1), Méier (3.2), Madureira (3.3), Inhaúma (3.4), Penha (3.5), Pavuna (3.6) and Ilha do Governador (3.7). These cover 81 neighbourhoods, in 13 administrative regions. The AP had a projected population of 2.84 million people in 2018, or 37.98% of the capital – the most in the region.

AP 3's average IPS in 2020 was 54.49 with Méier as the most developed administrative region (64.61), and Pavuna dead last (42.97). Per capita earnings in Zona Norte were at R$ 25,633 in 2018.

In AP 3, inequality as measured by the Gini index increased 0.74% between 2000 and 2010 – from 0.455 to 0.458. Ilha do Governador saw inequality increase 15.14%, while Madureira had its Gini index decreased by 10.11%[6].

4.3.4 Planning Area 4 – Baixada de Jacarepaguá

AP 4, Baixada de Jacarepaguá, has two RPs in Jacarepaguá (4.1) and Barra da Tijuca (4.2) with 19 neighbourhoods and three administrative regions. In 2018 AP 4 had 940,143 projected residents, or 14.37% of the city's total population.

2020's average IPS in AP 4 was 59.83, with Barra da Tijuca (69.7) leading the charge. Cidade de Deus, another well-known "favela" in Rio de Janeiro, boasted the lowest score (47.84) in the region. And thanks to Barra, with per capita earnings of R$ 84,849 per year – more than double of Jacarepaguá's R$ 33,373 – AP 4 registered R$ 50,036 p.c. earnings in 2018.

The Gini index in AP 4 increased from 0.445 to 0.491 between 2000 and 2010, meaning a 9.36% surge – the largest in the city. Barra da Tijuca's registered the largest inequality increase in the city – 21.41% – whereas Cidade de Deus saw the largest decrease in Rio de Janeiro (-11.44%)[7].

4.3.5 Planning Area 5 – Zona Oeste

AP 5, Zona Oeste, is made by the Bangu (5.1), Campo Grande (5.2), Santa Cruz (5.3) and Guaratiba (5.4) regions. It has 24 neighbourhoods covered by five administrative regions, with a population of 1.77 million (27.11%) according to 2018 projections.

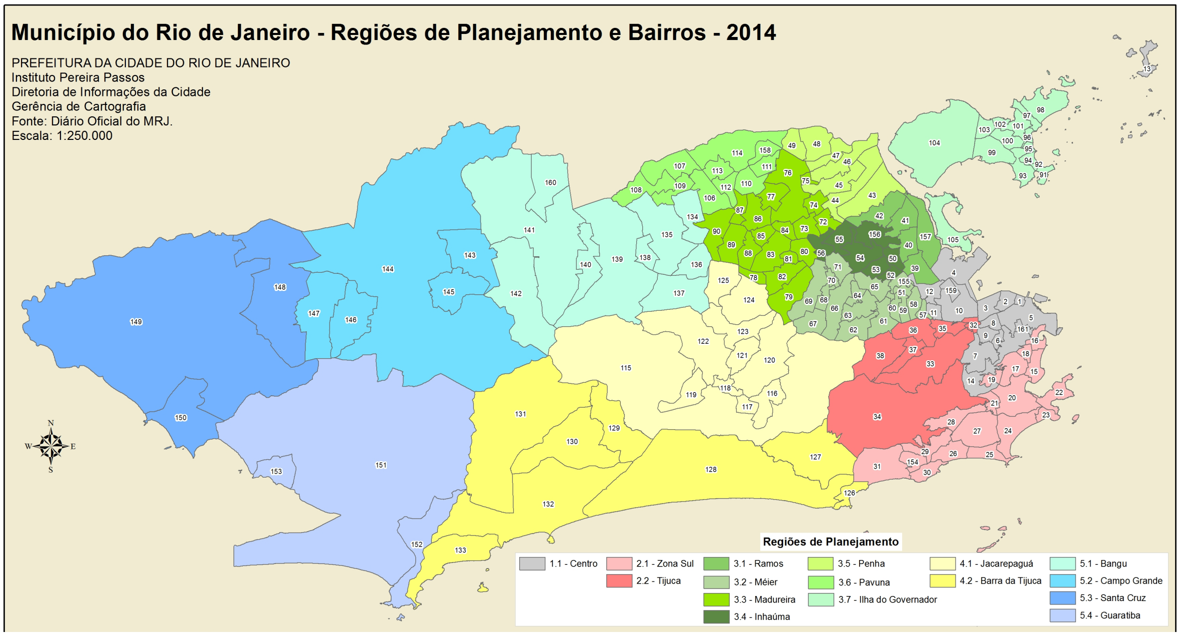
Average IPS in AP 5 was 52.13 in 2020, the lowest in the entire city. Guaratiba (43.54) was the worst performer, whereas Campo Grande (58.68) went in the opposite direction. In 2018 per capita earnings, AP 5 registered R$ 19,816.

AP 5 was the only region in Rio de Janeiro where inequality decreased overall between 2000 to 2010,

[6] Lists of Zona Norte neighbourhoods can be found in pages 81, 82, and 83.
[7] A list of Baixada de Jacarepaguá neighbourhoods can be found in page 84.

from 0.469 to 0.464 (-0.98%). The largest decrease was registered in Santa Cruz (-10.73%), and the largest increase in Guaratiba (5.61%) (Instituto Municipal de Urbanismo Pereira Passos 2019) [8] [9].

Figure 2: Map of Rio de Janeiro planning regions (RPs) and neighbourhoods



Available in <https://rioonwatch.org.br/?p=21068>.

Source: Rio On Watch.

[8] A list of Zona Oeste neighbourhoods can be found in page 85.

[9] Rio de Janeiro maps and graphs can be found from pages 86 to 89.

5 Assessing Rio de Janeiro's socioeconomic evolution from 2000 to 2015

As shown in the literature review, research on the link between minimum wage and regional economic activity is not extensive. Therefore, there are several reasons as to why research on the link between these two factors in Rio de Janeiro would be a good addition to the labour and regional economics fields.

One reason is Rio de Janeiro's position as one of Latin America and the world's biggest metropolis. With 13.2 million inhabitants in its metropolitan area region and 6.8 million in the capital, as well as its position in Brazil's economy as a leading region in sectors such as tourism, media and public services, its economy could be compared to several countries worldwide.

A second reason to motivate this research is Rio de Janeiro's inequality and diversity. The city composed by 165 neighbourhoods, 16 planning regions (APs) and 5 planning areas (RPs) with a wide range of socioeconomic conditions from access to public services to median wages, as well as 53% of non-white populace and 1.2 women/men ratio, could serve as a proper microscope to the country as a whole.

Another source of motivation for the research is the availability of data for such analysis. While there is plenty of microlevel data for developed nations such as the United States and those within the European Union, Brazil is somewhat unique as a country outside of the realms of rich nations that also has publicly available data at the individual level, with well-defined geographical delimitations for big cities such as Rio de Janeiro.

Finally, this research also goes beyond the go-to locations for this kind of research. Most research on minimum wage impacts in the past focused on United States and the European Union at the country and state levels, with lower geographical divisions taking over as time goes by (Dube 2019a). Here the objective is following the microregional analysis trend while applying it at one city, with potential of doing so in more Brazilian cities – and beyond – in the future.

5.1 Data

Brazil is in a privileged position in comparison to countries of similar size and economic dynamism when it comes to publicly available data. Much of this is thanks to its statistics institute, Instituto Brasileiro de Geografia e Estatística (IBGE), created in 1936 but with its origins dating to the 19th century during Brazil's time as an empire (Guimarães 2022).

IBGE regularly publicizes annual surveys such as Pesquisa Nacional por Amostra de Domicílios

(PNAD), which focuses on job market and educational status at the individual level and access to public utilities and home appliances at the domicile level. It also calculates Brazil's official inflation index, Índice de Preços ao Consumidor Amplo (IPCA), since 1980 using cost of living for families in the countries' biggest metropolitan areas.

IBGE's tradition of easily accessible data, including its Census – done since 1940, on somewhat regular 10-year intervals – tracks to the federal and regional level ministries and offices (Instituto Brasileiro de Geografia e Estatística 2013). One of the most utilized employer-employee databases in Brazil, Relatório Anual de Informações Sociais (RAIS), is collected and publicised yearly by Brazil's Ministry of Labour since the 1970s (Instituto Brasileiro de Geografia e Estatística 2022b).

This open data "policy" has led to more and more works in the field of applied microeconomics and econometrics on Brazilian data. While most are still done with the country, Brazil has become a choice of analysis for papers by economists in the United States and beyond.

5.1.1 Databases

For the upcoming descriptive statistics and data analysis, three databases were utilised.

Relatório Anual de Informações Sociais (RAIS):

The website Base dos Dados, maintained by a team of Brazilian researchers and lead by Ricardo Dahis, assistant professor of Economics at Pontifícia Universidade Católica do Rio de Janeiro, provides free access to RAIS and several other databases (Dahis 2022).

There are two types of RAIS databases, one called "vinculos" for workers and "estabelecimentos" for companies. At Base dos Dados, both databases were tidied according to tidy data principles popularised by statistician Hadley Wickham (Wickham 2014) with a temporal coverage from 1985 to 2020.

For this analysis, both RAIS databases were downloaded for the 2000–2015-year range. Analysis focused on the 2003-2014 interim, with most variables present and utilised.

One important addition to the databases and to the task at hand, is the presence of a categorical variable for a workers' race since 2006. For previous years in the range of analysis where race was not present in the database, a combination of predictive mean matching (PMM) and multiple imputation by chained equations (MICE) methods was used to help determine which race a worker belonged to.

Pesquisa Nacional por Amostra de Domicílios (PNAD):

Base dos Dados also provides free access to PNAD, from 1981 to 2015. PNAD also has two different databases, for individuals ("pessoa") and domiciles ("domicilio").

One difficulty from PNAD is the fact it does not have subregional delimitations similar to RAIS and Census, with the survey only informing whether observations were collected at the urban or rural area of a metropolitan area such as Rio de Janeiro. Given this research looks to draw results from minimum wage impact at the microlevel due different socioeconomic compositions of Rio de Janeiro neighbourhoods, using PNAD for such analysis is a difficult task.

Nevertheless, data was collected from 2001 to 2015. One important detail is that data for 2000 and 2010 years is non-existent due to Census being held by IGBE during these two years.

Censo Demográfico:

Once again, Base dos Dados is the data provider. It has Census data for two databases at the individual ("pessoa") and domicile ("domicilios") levels, from 1970 to 2010.

Census and PNAD data can be seen as complementary, given that PNAD surveys often carry similar questions/variables and they are not held on Census years. However, Census is much wider in its scope, and it also includes geographical subdivisions for each observation *a la* RAIS.

5.2 Descriptive statistics

One big motivation behind this research is the fact that growth in Brazil was not at all equal during its economic boom in the 2000s. At the time growth was much stronger in cities in the North and Northeast regions, which are also the poorest in the country.

From 2000 to 2009 GDP in North and Northeast regions of the country grew by 219% and 199%, respectively, leading to both regions increasing their impact in Brazil's GDP. In the richest region, Southeast – where Rio de Janeiro is located – GDP grew by 160% during this timeframe while its overall economic impact diminished (Instituto Mauro Borges 2018). That was the case with Rio de Janeiro. From 2002 to 2013 the former Brazilian capital saw massive GDP growth, increasing by 201% (Instituto Brasileiro de Geografia e Estatística 2019).

Thus given the positive impact of minimum wage growth on low-wage regions (Godoey and Reich 2021) and Rio de Janeiro's inequality – as well as the dynamics of wage inequality reduction in the country during the 2000s, with racial and gender heterogeneities as two potential explanatory factors of this mechanism (Derenoncourt et al. 2021; Komatsu and Filho 2016) –, it could be the case that one driver of the region's performance during the 2000s is similar to the minimum wage mechanism that drove wage inequality down in the rest of the country.

5.2.1 Census 2000 and 2010

In Brazil, national Census surveys are done every 10 years since 1940 – although its latest edition, which would have been done in 2020, was delayed due to the COVID-19 pandemic (Gaglioni 2022). Still, it is one of the best ways to show how much has changed in the country over the years in several aspects, from ageing trends to socioeconomic conditions such as education and income.

In both Census, most observations were taken from AP Zona Norte, with Zona Oeste following. The least populated region was Centro, which concentrates most companies and thus proper economic activity in Rio de Janeiro.

At the RP level, there was a shift in individual interviews for the Census survey with Centro increasing by 62.6% in 10 years' time. Meanwhile Bangu and Guaratiba saw decreases of 25.5% and 63.2% each, in relation to their previous impact in Rio de Janeiro's observations.

There are several ways in which the minimum wage growth policy in the 2000s might be shown to alter the populace's living standards at the lower quantiles of income and wealth in Rio de Janeiro. One illustrative way of doing so is the change of domiciles with a personal computer (PC).

In the 2000 Census, it is quite clear that access to PCs in Rio de Janeiro is scarce despite it being one of Brazil's richest cities. Back then only 22.3% of households had a PC in the city. 10 years later, nearly 60% of homes had the appliance – a 165% increase.

Figure 3: PC presence in Rio between 2000 and 2010

Source: IBGE. Own elaboration.

This also shows in income distribution across the city. In 2000, 21.3% of households were in the 0-2 minimum wage bracket of total income with most homes earning below the federal floor in all APs but Zona Sul e Grande Tijuca, where most households were in the 2-3 minimum wage bracket. However average income in households was 11.76 minimum wages, with a median of 6.00, which goes to show how unequally income was distributed within the city.

In the 2010 Census, there is again noticeable improvement for households in the lower income brackets. The bracket "hump" shifts from the 0-2 income bracket to the 1-6 counterpart, with most households – 42% of them – now in the 1-4 division.

Rio de Janeiro's socioeconomic fabric also changed markedly during the 2000s, as Census data shows. White/non-white makeup of Rio's populace went from 57.72-42.28 to 51.09-49.91 proportion between 2000 and 2010 in the entire city. Meanwhile the male population increased in the timeframe, from 47.63% to 49.34% of the city's demographic.

Census survey also captured the change in Rio de Janeiro's age distribution. In 2000, 29% of the populace was below 18-years-old. 10 years later, the curve "shifted" upwards with 19.6% of the population being underage.

Mean and median ages in the 2000-2010 interval also illustrate how the population got older. In 2000 mean and median ages were 32.44 and 30 respectively, whereas in 2010 Census it "jumped" to 36.48 and 35.

Decreases were registered in every range below 25-year-olds, with a 34.7% decrease in the 0-14 age bracket. Meanwhile every other age range increased, with the biggest ones in the 45-65 (+27.6%) and 76+ (+29.8%) brackets.

Positive shifts in working age population, along with better economic conditions in the entire country and thus more working opportunities, lead to a noticeable increase in the economically active population within Rio de Janeiro. The proportion increased from 46.1% of working population in 2000, to 59.8% in 2010.

Interestingly, there was a decrease in mean working hours per week between the Census surveys from 43.02 to 37.32, whereas the median remained at the legal limit of 40 hours as established by the country's Consolidation of Labour Laws (Consolidação das Leis do Trabalho, or CLT, in Portuguese). This mean decrease might be in part explained by an increase in internship recruitment, whose workload are limited to maximum 30 hours per week; and other assortments of working relationships, such as freelancing.

In terms of individual income, movement from 2000 to 2010 could be seen as a concentration of workers suffering from losing earnings as the 0-2 bracket increased by 54.2% - with every other range decreasing in turn. This however would ignore the fact minimum wage increased by nearly 100% between surveys in real terms.

Therefore, while mean and median wages per individual were R$ 975 and R$ 453 in 2000, respectively, these increased to R$ 2,073 and R$ 950 in 2010 – almost following minimum wage growth trends. In minimum wage terms, there was a decrease from 6.46 to 4.1 in mean earnings by using such a parameter.

Using minimum wage ranges to gauge how income distribution changed in each of Rio de Janeiro's APs, one can see the big impact it could have had in the entire location. But while the biggest increases in workers earning 0-3 minimum wages took place in Baixada de Jacarepaguá, Centro and Zona Sul e Grande Tijuca regions, every region registered similar changes.

Demographic shifts in 10 years are bound to show up in educational levels, and this is what took place in Rio de Janeiro as well. The city saw a massive incline in groups of people completing secondary and tertiary degrees, with the latter increasing 69.3% from 9.1% to 15.4% in a 10-year span. Meanwhile people with less than primary education decreased by 29% in the same time period, to 38.1% from 53.7%[10].

5.2.2 PNAD, 2001 to 2015

While the Census is held once every 10 years, PNAD is a continuous survey held every year in Brazil that is also very helpful in tracking the country's socioeconomic developments. One major caveat from PNAD to this research is the fact it does not have geographical subdivisions such as APs and RPs, which is the case for Census and RAIS.

Another issue from PNAD is that it does not even distinguish cities within a metropolitan area such as Rio de Janeiro's. Rio's metro area is quite diverse, composed of 22 cities surrounding the state capital covering 7.5 million km2 and with almost 13.2 million inhabitants according to the latest IBGE estimates (Instituto Brasileiro de Geografia e Estatística 2022a).

With this is mind, statistics were drawn from PNAD surveys only from Rio's metropolitan area and in urban areas, which means observations will be mostly taken from the state capital – which is the focus of this research.

Similar to Census, PNAD has data at the domicile and individual levels. Domicile data is quite helpful here to show how income has evolved over the years in Rio de Janeiro overall. In 2001, 60.4% of homes in Rio's metro area had a total income below R$ 1,000. 14 years later, 16.2% of domiciles were at this income level, meaning a 73.17% relative decrease in the interim. At the same time every other income bracket grew strongly. Most notable are increases at the R$ 16,000-17,000 and beyond, with increases that surpass the 1,000% mark. Mean income at the domicile level in 2001 was R$ 1,392, with a median of R$ 800. It grew to R$ 3,935 in 2015, at a R$ 2,300 median – increases of 182% and 187%, respectively.

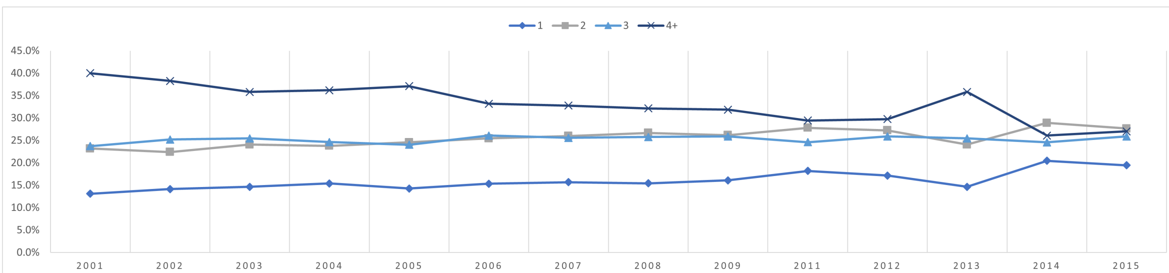
Another insight that PNAD can give into Rio's development, is the number of people living at a domicile. Year-on-year changes to occupation of a home can be indicative of how a country's

[10] Descriptive statistics for Censo 2000 vs. Censo 2010 can be found in page 90.

quality of life has evolved, with fewer people living at a domicile helping denote not just better living conditions but also how the population has been able to find and maintain a roof above their head with little to no need of cost-sharing (Chimed-Ochir et al. 2021; Zainal et al. 2012).

Data here is quite telling in this regard, with 40% of Rio's population living at homes with 4 or more people in 2001. Whereas in 2015 this percentage fell to 27%, indicating a 32.5% decrease in 2015. Meanwhile, there was strong growth in homes with 3 or fewer individuals, with the most significant increase taking place in single-occupation domiciles – from 13.1% to 19.5%, a 48.55% increase between 2001 and 2015.

Figure 4: People per domicile in Rio between 2001 and 2015



Source: IBGE. Own elaboration.

When looking at individual data from PNAD, it needs to be highlighted again that Rio de Janeiro's metro area is quite diverse. Therefore, numbers can be skewed given the sample pool from which observations are being drawn, which will include not just the capital but also cities with very different socioeconomic conditions such as Niterói, which has the state's best socioeconomic indicators; and Japeri, which is in the exact opposite spectrum (Casa Fluminense 2020).

In terms of racial composition, Rio's metropolitan area followed the capital trend. The white population in the region fell 25.5% percent between 2001 and 2015, from 59.5% to 44.3%. Which means the non-white population grew proportionally, from 40.5% to 55.7%, thanks to an increase in black and brown people living in the metro area. Gender composition, however, was stable throughout.

Educational level in Rio de Janeiro's metropolis also changed remarkably. In terms of last degree attended, which include students and graduates of said degree, only secondary, tertiary and post-tertiary education frequencies grew between 2001 and 2015 to 60.6% from 34.5% - a 75.61% increase. Counting only tertiary and post-tertiary education, numbers are just as impressive with 20.9% studying or graduating at these levels in 2015, from 13.1% in 2001.

Years of studying complement the picture, showing how many have graduated from secondary degrees and beyond. In this regard, 26% of survey respondents had studied 11 years or more back in 2001. 14 years later, 43.6% had done so, meaning a 67.7% increase in this time period.

Analysing mean and median statistics of years of studying help completing what has been described here. Mean years of studying in 2015 was 8 years, the median being the same. Back in 2001 the mean was 6.2 years, with a 6-year median.

In terms of age, Rio de Janeiro's metro area got older between 2001 and 2015. Population below 34 years of age fell to 47% from 55.8%, a 15.8% decline. On the opposite spectrum there was 20% growth of people aged 34+, with strongest growth in elders from 87 and above (+62.5% from 2001 to 2015).

Back in 2001, PNAD registered mean and median ages of 32.32 and 30 years, respectively, among interviewees for the survey. In 2015 mean age was 37 years, and the median, 36 years. Interestingly, in Rio de Janeiro's metropolitan area working population was kept mostly stable according to PNAD data, which goes in the opposite direction of trends shown by Census when looking only at the state capital. Working populace did grow to 50.4% in 2015 from 48% in 2001.

Among workers, there were some interesting changes in terms of worked hours per week. There was a 169.3% increase in piecemeal work, also colloquially known as "bicos", where workers can only find casual jobs at 10 hours of occupation or less – from 2.4% in 2001, to 6.4% in 2015. At the same time, there were increases in worked hours at the 20-40 range, which are within the legal limits set by CLT for internships and full-time jobs – to 53.3% from 41.8%, between 2001 and 2015.

A significant number of workers back in 2001 worked between 40 and 60 hours per week back in 2001, making up 44.1% of the working population in Rio's metro area. In 2015 this section of the populace fell to 32.5%, meaning a 26.4% decrease.

Mean worked hours in 2001 were 42.7, with a 40-hour median. While the median stayed the same in 2015, mean worked hours fell to 38.32.

Before moving to the last variable of interest in PNAD, changes in Social Security contribution should also be highlighted. Brazil's labour market is stratified by formal and informal contractual relationships, with the former under CLT laws which come with a plethora of obligations to the employer. While hiring workers without a formal contract is an illegal practice, loose regulation makes it commonplace in the entire country.

Most papers analysing Brazil's labour market evolution during the 2000s indicate there has not been a loss of welfare to workers in informal relationships, when measured by wages obtained in their jobs and/or moving from formal to informal work spots (Boeri, Garibaldi, and Ribeiro 2011; Derenoncourt et al. 2021; Engbom and Moser 2021; Komatsu and Filho 2016).

PNAD helps backing these findings. In 2001, 67.2% of the working population in Rio's metro area were under CLT labour laws – also known as "carteira assinada", in Portuguese. The number increased by 13.1% in 2015, to 75.9%.

It is only natural that contributions to Social Security, which is tied to having a "carteira assinada", would increase in turn from 60.4% to 68.6% between 2001 and 2015, with a 13.7% growth. The 7.3% absolute gap between CLT workers and INSS contributions could be explained by workers in internships, who do not contribute to Social Security; and those in alternative labour relationships which are still legal under labour law such as independent/freelance workers who are companies themselves.

Last but not least, changes in income at the individual level pretty much track the domicile-level trends explained above. People earning less than R$ 1,000 were 81% of Rio's metro area population in 2001, and 39.5% of the PNAD sampling in 2015 – a 51.2% decrease in 14 years. Individuals earning R$ 1,000 to R$ 6,000 became 54.6% of the populace in 2015, compared to 18% 14 years before.

On average, PNAD interviewees in Rio de Janeiro's metropolis earned R$ 817.65 with a R$ 420 median in 2001. 14 years later, mean income grew to R$ 2280, with median income also increasing, to R$ 1,270[11].

5.2.3 RAIS, 2003 to 2014

RAIS is an employer-employee linked database, whose data is filled every year by employers of millions of companies all around Brazil. It contains detailed information on several aspects of business such as the company's number of contracted employees, geographical location and market; and the same for workers, including level of education, time of employment, race and gender.

Plenty of research using RAIS databases focuses on state-by-state or city-by-city comparisons. Here the focus is on effects of the minimum wage in Rio de Janeiro's state capital only, which is facilitated by the fact RAIS uses neighbourhood classifications which are similar to Census' data. Therefore, the city can be divided by APs and RPs, and the presence of heterogeneous effects on city-wide treatment effects such as minimum wage growth can be assessed as well.

In 2003 the AP with most companies was Zona Sul e Grande Tijuca, with 28.1% of the city's enterprises. Zona Oeste was on the opposite spectrum with only 9.2% of companies opened and registered to Brazil's official systems.

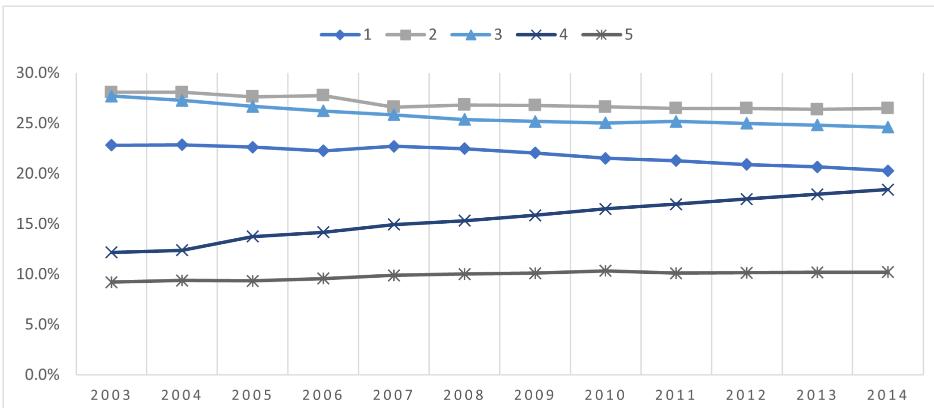
While Zona Sul e Grande Tijuca still lead the chart in this regard 11 years later, the city's outlook did change a bit. Baixada de Jacarepaguá saw strong growth since 2003, jumping 48.3% from 12.2% to 18.4% of Rio de Janeiro companies located in the region. Zona Oeste also grew in number of

[11] Descriptive statistics for PNAD 2001 vs. PNAD 2015 can be found in page 91.

companies installed in the AP, to 10.2% - meaning 13.2% growth between 2003 and 2014.

Looking at the RP level, things get a bit more interesting – and telling. One can see that the reason behind Baixada de Jacarepaguá's growth was both Barra da Tijuca and Jacarepaguá RPs doubling the number of companies installed in these locations. With that Barra da Tijuca became the city's third-largest RP in companies installed with 11.4% representation in 2014, overtaking Tijuca RP.

Figure 5: Companies per AP in Rio between 2003 and 2014



Source: Ministério do Trabalho e Previdência. Own elaboration.

In terms of legal nature, most companies in Rio de Janeiro are limited companies ("sociedade limitada" in Portuguese) which grew from 141,726 to 150,777 enterprises between 2003 and 2014 – a 6.3% increase.

It is also interesting looking at companies opting for the "Simples Nacional" tax regime. This is limited to micro and small enterprises which employ up to 49 people, providing incentives to those who opt into the program (Governo Federal do Brasil 2022; Receita Federal 2007).

While 41.9% of companies in Rio de Janeiro were "Simples Nacional" opt-ins in 2014, in comparison to 42.3% in 2003, the number of companies in the regime jumped to 115,496 from 90,959 in this time frame. Most of them are in the city's poorest regions, Zona Norte and Zona Oeste, which are the two APS that have over 50% of companies choosing to opt into "Simples Nacional" in 2003 and 2014.

Micro companies and SMEs are the ones listed the most in Rio de Janeiro's data from RAIS over the years. They represent at least 97% of companies in the city between 2003 and 2014 – although several of them are those which employ zero people, meaning they are either freelancing/professional liberal office operations and/or companies whose operations were over during the year. These companies hiring no people whatsoever but still active in the economy, represent at least 51.6% of the RAIS "estabelecimentos" sample between 2003 and 2014. However, the kind of enterprise which grew the most in Rio are those which employed 1,000+ people, by 90% from 2003 (135 companies) to 2014 (257).

When looking at workers in the payroll, the statistic is also heavily slanted towards micro and SME companies. More than 93% of firms in Rio employ fewer than 50 people, something that does not change between 2003 and 2014.

The aforementioned trends mostly track to CLT employees. But such is not really the case with statutory workers. These are public service employees which are a significant part of Rio de Janeiro's economy, due to the city's legacy as a former federal capital. As exposed previously, public administration work represents 19.75% of Rio's GRP in 2019 (Instituto Brasileiro de Geografia e Estatística 2019).

This kind of employment has seen a lot of change in terms of firm size. Back in 2003 firms with 0 to 10 statutory workers were the largest group in the city, with 46.1% of representation throughout. This fell to 11.7% in 2014, meaning a 74.6% decrease.

In 2014, the most representative group are firms with 100 to 200 statutory employees, with 15.9% of the city's share. Meanwhile there are two "companies" – the state and the municipal governments – employing 70,000 to 100,000 people since 2003.

Another thing of note is how the city's business kept open throughout the years. From 2005, which is the first year where the active business indicator is available, 82.5% of companies listed in RAIS "estabelecimentos" were active during that period. This fell to 79.7% in 2014.

Companies in Baixada de Jacarepaguá were the most successful ones in this regard during the timeframe, with 82% of them kept open in 2014. The least favourable regions were Zona Oeste (77.2% in 2014) and Centro (77.6% in 2014).

On average, companies in 2003 had 8.23 active workers in their payroll. In 2014 the mean jumped to 9.63. CLT links increased from 6.35 to 7.89; and statutory employment decreased, although slightly, to 1.74 from 1.88 between 2003 and 2014.

At the AP level, Centro is the region with most employees per enterprise with an 18.1 average in 2014, from 14.2 in 2003. On the opposite side of the spectrum is Zona Sul e Grande Tijuca, with a 6.6 average which sets them behind Baixada de Jacarepaguá and Zona Oeste – both of which overtook the former in terms of employees on payroll per company[12].

The last findings to be discussed in this section are about the research's main database, RAIS "vínculos" – or RAIS at the workers' level. This is a database with over 38 million observations across

[12] Descriptive statistics for RAIS "estabelecimentos" 2003 vs. 2014 can be found in page 92.

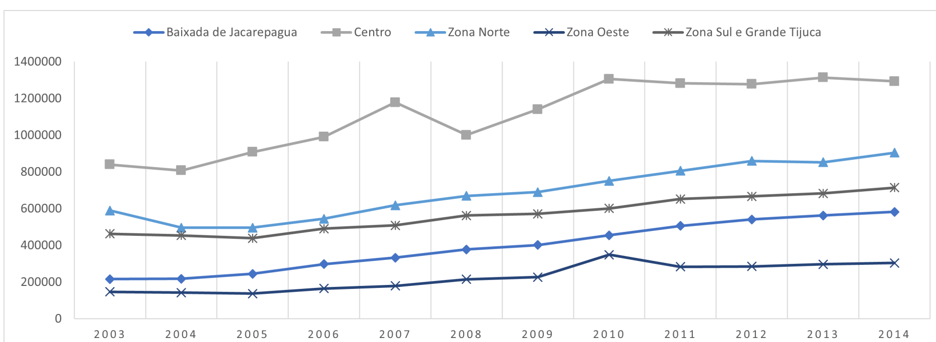
12 years, for workers in Rio de Janeiro's state capital only.

As the main economic hub in Rio de Janeiro, the Centro AP had the biggest slice of work contracts in the city throughout the period assessed in this research. The share declined however, from 37.3% to 34% between 2003 and 2014, as Baixada de Jacarepaguá (15.3% in 2014, from 9.6% in 2003) and Zona Oeste (8%, from 6.5% in 2003) saw massive growth in their participation in Rio's labour market.

Baixada de Jacarepaguá and Zona Oeste's growth is even more impressive in absolute numbers. The former went from 219,002 to 582,676 workers, meaning a 170% increase in 11 years. Meanwhile Zona Oeste got to 304,340 workers in the region from 146,754 in 2003, with a 107.6% rate of growth ever since.

At the RP level Guaratiba's growth was a stalwart, as the least economically relevant region in the city grew 382.3% in terms of workers at the location, from 3,182 in 2003 to 15,346 in 2014. Just as impressive were the increases in Jacarepaguá, Santa Cruz and Barra da Tijuca RPs, with up to 34% increases in the state capital's labour market participation; and more than 140% growth in number of workers at these regions.

Figure 6: Work contracts per AP in Rio, between 2003 and 2014



Source: Ministério do Trabalho e Previdência. Own elaboration.

The racial outlook in Rio de Janeiro's labour market did replicate trends shown in Census and PNAD. From the first available year in which race is coded, 2006, white workers represented 58.7% of the city's share. This decreased by 15.2% until 2014, when the share fell to 49.8%.

The AP outlook in racial terms shows how changes were drastic, specially in the Zona Norte and Zona Oeste regions. White workers represented 54% of the labour force in both regions back in 2003. Their representation fell to 47.5% in Zona Oeste, and 43.9% in Zona Norte over the next 11 years. Meanwhile regions such as Zona Sul e Grande Tijuca, with 62.2% of its workers being white in 2003, saw this share fall by 15.2% in 2014 to 52.7%.

Such demographic changes could be explained by big increases in non-white labour market presence throughout the city when looked at the AP level. Non-white worker's participation increased by 112.8% between 2003 and 2014 in Zona Oeste, and by 116.5% in Baixada de Jacarepaguá in this time frame.

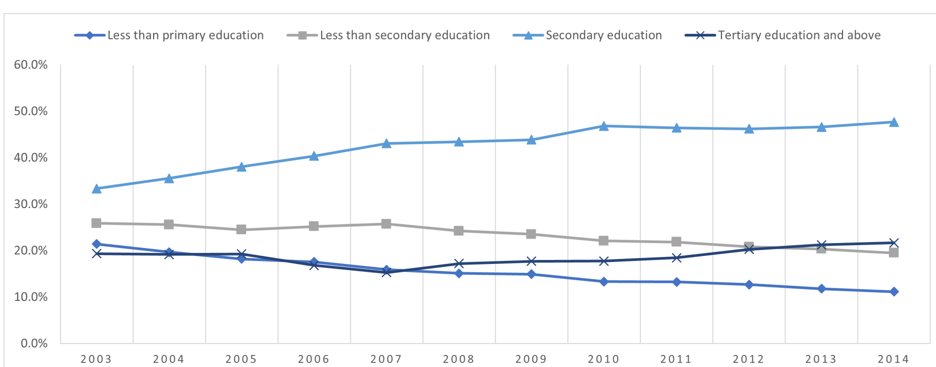
The gender outlook, however, did not change much throughout the years. Female participation did increase by 5.6% between 2003 and 2014, from 39.7% to 41.9% citywide. At the AP level, every region but Zona Norte saw growth in female labour market participation between 40.4% in Baixada de Jacarepaguá, and 44.5% in Zona Sul e Grande Tijuca. In Zona Norte female labour market presence fell to 39.1% from 40.5% in the 2003-2014 interval.

Until 2005, RAIS did not distinguish levels of tertiary education between bachelor's, master's and doctorate degrees. These were all bunched into tertiary degrees, which were the second-most representative group in the city at 19.3% in 2005, only behind secondary degree holders at 38%.

From 2006, when master and doctorate degrees began being distinguished, the share of tertiary degree holders fell to 16.8%. But it increased again to 21.7% in 2014, representing a 28.9% increase ever since at the city level. However secondary degree holders still lead the charts with 47.7% of the labour market's share.

When it comes to AP, trends are quite similar. Secondary, tertiary and post-tertiary degrees all increased in labour market presence, both in relative and absolute numbers. In all APs holders of secondary degrees in the respective labour markets grew by at least 24% with strong increases in Zona Norte (+62.1%) and Zona Oeste (+50.5%) from 2003 to 2014. In the meantime Baixada de Jacarepaguá saw the strongest relative increase in tertiary and post-tertiary workers, from 11.9% to 14.9% – a 25% "jump".

Figure 7: Education of workers in Rio between 2003 and 2014



Source: Ministério do Trabalho e Previdência. Own elaboration.

Age in Rio's labour market shows some interesting trends. Groups with the strongest increases in participation and absolute numbers over the 2003-2014 period were workers between 15 and 17 years of age, and 50+ workers. Every other age group but 30-39 years olds saw decreases in labour market share, although this group is still the most representative in the entire city.

However mean and median ages in the city's labour environment remained largely stable. Mean age in the labour market increased to 36.7 from 36.1 years in 2014, whereas median age remained at 35 years in the 2003-2014 interim.

Most workers in Rio work either in limited or closed capital companies, which hired between 61.4% and 65.7% of workers in the city. As for company size, most workers in Rio de Janeiro between 2003 and 2014 were at enterprises hiring more than 1,000 people with 25 to 29.9% of the city's labour market share. Employment at this level increased by 83.8% from 2003 to 2014, with hiring in companies between 20 and 99 workers growing by at least 81.5%.

One interesting trend in type of hiring in Rio's capital is the fall in first employment hires in terms of labour market participation. This kind of hiring represented 4.4% of hires in the city back in 2003, and it fell to 3.4% in 2014. In the meantime, transfers without costs of workers between companies grew from 2% to 3.3%; and rehires also increased, to 27.3% in 2014 from 21.7% in 2003.

The most common employment relationship in RAIS falls under CLT, representing at least 76% of hires in the city since 2003. The second-most representative group are statutory workers, although its share fell to 12.6% in 2014 from 17.4% 11 years before – a 27.5% decrease.

As mentioned in previous sections for the Census and PNAD surveys, most workers in Rio are hired at the maximum law limit of 44 hours of work per week. Their participation in the labour market remained largely stable, from 63.9% to 68.1% between 2003 and 2014 in the state capital.

However, the two contracts which grew the most in a 11-year span were 16-20 and 31-40 weekly hour hirings. The former increased by 198% city-wide, and the latter by 101.7% in absolute numbers over the 2003-2014 time frame. One thing of note in the AP-level analysis is that Centro has more workers under 31-40 hours per week than 41-44 hours. This is largely explained by statutory workers whose hours are limited to 40 every week.

There have been some interesting trends in hirings and firings in Rio de Janeiro in monthly terms over the years. Peak hiring and firing months changed markedly over the years, with the only consistent trend being December as the month with the least hires between 2003 and 2014.

Average income for workers in Rio de Janeiro, when assessed by minimum wage earnings, show large relative increases in the 0-2 minimum wage groups. Every other group saw a decrease between 2003 and 2014, which might be explained by new entrants in the labour market and also strong minimum wage growth itself in the assessed time period outpacing wage adjustment for workers.

In every AP region but Centro, workers earning between 1 and 3 minimum wages represent over 71.1% of the labour market in 2014. Zona Oeste has the largest share of such group at 82.4%, and Centro has the lowest at 53.2%.

In 2003 mean wages were R$ 1,188 in absolute terms, and 5.2 minimum wages at a median of R$ 593 and 2.6 MWs. 11 years later mean wages increased to R$ 2,688 city-wide, with a median of R$ 1,378. However, earnings measured by the minimum wage fell to 3.7 on average, at a 1.9 median.

And looking only at December earnings, the scenario is not too different. There were similar trends across city-wide and AP-level workers, and also in mean and median wages earned during the 2003-2014 interim.

Back in 2003 18.6% of workers were at their jobs for more than 10 years in Rio. This fell to 13.3% in 2014, which is a 28.4% decline in the time frame. Meanwhile workers with less than 3 months at a job became the most significant "slice" of the labour market, representing 16.4% of workers in 2014 from 12.7% in 2003.

Mean employment time in 2003 was 66.2 months, with a 25.8 median city-wide. 11 years later the mean fell to 53.2 months, at a 17.9 median.

At the city level in 2003, 73.8% of workers kept their job at the end of the year. This fell by 10% over the years, to 66.5% in 2014. At the AP level, Centro was the "safest" region for workers at 73.3% of them keeping their job by the end of the year in 2014, in comparison to 75.9% in 2003 – something that could be explained by the number of statutory/public service workers in the region[13] [14].

[13] Descriptive statistics for RAIS "vínculos" 2003 vs. 2014 can be found in page 93.

[14] Further descriptive analysis visualizations can be found from pages 94 to 101.

6 Econometric analysis of minimum wage effects in regional economic activity

Richness in data, as shown by the previous chapter, will also allow a deep analysis into the economic trends of Rio de Janeiro's labour market and economy from the early 2000s until the mid-2010s. This is in track with ongoing developments in labour economics that entwined with regional economics, turns research outlooks from macro- to micro-level/granular analysis thanks to public access data throughout the world (Spanache 2020; New Zealand Government 2022; Dahis 2022).

For this research, there is a wide range of econometric tools that can be employed to assess effects of minimum wage growth at each of Rio de Janeiro's main regions over the years. A combination of difference-in-differences (DiD) methodology, which has been the go-to "instrument" for the field in the last few years, is the main candidate for any assessments.

However, gaps in data from earlier years in the main databases for this assessment demand some kind of arrangement that could help avoid biased and "noisy" estimates paint a divergent picture form what this data really tells. This is where the aforementioned PMM-MICE (Predictive mean matching/Multivariate imputation by chained equations) methodology comes in.

There is also a need to check not just how robust findings are from the combination of the two aforementioned methods, but also to help assess which components in the data influence economic activity the most – or the least – in Rio de Janeiro. For this purpose, principal component analysis (PCA) and factor analysis (FA) can be of great aid to this research (Dai, Xiong, and Zhou 2021; Godshalk and Timothy 1988; Lai 2003; Marques and Neves 2001; Vyas and Kumaranayake 2006).

6.1 Methodologies

Here is more detail on the econometric methodologies to be used in this research.

Difference-in-differences (DiD):

The major intuition behind DiD is comparing the change in outcome of a certain dependent variable in a treatment group, to the change in outcome of a control group. This way the method can help identify what would have happened to treated groups, had the intervention not occurred in the first place, thanks to a general, or common, time trend.

It also means we assume that in absence of intervention, treatment and control groups would have evolved similarly – something that can be assessed by using previous time periods to control for differences in environment that are not captured by randomisation (Columbia Mailman School of Public Health 2013; McKenzie 2020).

In DiD, the identification setup work by defining expected outcomes given treatment and its absence. To draw out the changes in outcomes of treatment and control groups and estimate treatment effects, time and group-specific fixed effects are estimated as a way to best "isolate" the impact of a treatment in the sample. This gives a path to getting group-specific linear trends as well, from the DiD method (Callaway and Sant'Anna 2022) [15].

Treatment in this research is inspired by Derenoncourt et al. (2021) which utilises the concept of strongly and weakly treated regions to assess effects of public policy at the micro-level. Common measures to define said regions are average and median wages, as well as a conjunction of socioeconomic statistics that can help determine how "vulnerable" a region and its populace are to policies such as minimum wage growth over an extensive period of time.

For this research, the chosen statistics to define strongly and weakly treated Rio de Janeiro APs and RPs are average, median and standard deviation of wages in the year 2000, using the RAIS "vínculos" database. We thus define strongly and weakly treated regions in Rio by looking at information pre-treatment – i.e., strong minimum wage growth due to federal government policy.

Table 2: Average and median income of APs in Rio in 2000 (in R$)

	City	**Baixada de Jacarepagua**	**Centro**	**Zona Norte**	**Zona Oeste**	**Zona Sul e Grande Tijuca**
Mean	897.72	703.32	1058.53	699.72	601.89	864.47
Std.Dev	1248.47	1031.34	1407.95	870.33	747.86	1272.35
Min	0	0	0	0	0	0
Median	483.95	390.2	598.16	423.46	376.34	430.81
Max	18020.07	17325	18020.07	17910.54	17538.83	18000

Source: Ministério do Trabalho e Previdência. Own elaboration.

[15] More details can be found in Appendix A1 – Difference-in-differences (DiD)

Table 3: Average and median income of RPs in Rio in 2000 (in R$)

	City	Bangu	Barra da Tijuca	Campo Grande	Centro	Guaratiba	Ilha do Governador
Mean	897.72	552.36	732.77	591.09	1058.53	714.62	1028.41
Std.Dev	1248.47	694.25	1112.2	699.64	1407.95	1057.14	1221.7
Min	0	0	0	0	0	0	0
Median	483.95	360.49	384.25	377.47	598.16	320.31	576.57
Max	18020.07	16099.95	17015.61	17538.83	18020.07	12187.18	16719.49

	Inhauma	Jacarepagua	Madureira	Meier	Pavuna	Penha
Mean	562.53	661.81	749.2	598.48	696.29	667.64
Std.Dev	642.53	903.52	832.04	798.17	948.94	865.28
Min	0	0	0	0	0	0
Median	400.33	397.9	483.23	375.18	426.39	413.51
Max	15062.73	17325	16925.9	15653.91	15456.8	17910.54

	Ramos	Santa Cruz	Tijuca	Zona Sul
Mean	604.84	730	756.93	918.57
Std.Dev	719.17	909.42	1057.32	1364.61
Min	0	0	0	0
Median	406.66	414.34	393.58	455.2
Max	14622.81	16665.26	17785.13	18000

Source: Ministério do Trabalho e Previdência. Own elaboration.

Therefore, these are the strongly and weakly treated APs in Rio:

- **Strongly treated**: Zona Norte (AP 3), Baixada de Jacarepaguá (AP 4) & Zona Oeste (AP 5);
- **Weakly treated**: Centro (AP 1) & Zona Sul e Grande Tijuca (AP 2).

And these are the strongly and treated RPs in Rio, according to the aforementioned measures:

- **Strongly treated**: Bangu (RP 5.1), Campo Grande (RP 5.2), Guaratiba (R.P 5.4), Inhaúma (RP 3.4), Jacarepaguá (RP 4.1), Méier (RP 3.2), Pavuna (RP 3.6), Penha (RP 3.5), Ramos (RP 3.1), Santa Cruz (RP 5.3);
- **Weakly treated**: Barra da Tijuca (RP 4.2), Centro (RP 1.1), Ilha do Governador (RP 3.7), Madureira (RP 3.3), Tijuca (RP 2.2), Zona Sul (RP 2.1).

Predictive mean matching (PMM):

Although RAIS is a rich database, some of its years have gaps in variables that could be very important to this research. One of which is race, with every year of RAIS "vínculos" before 2006 missing data for observations in the dataset; as well as AP and RP, which are missing from RAIS "estabelecimentos" dataset in 2002.

There are however several statistical techniques in which one can infer which value an observation can take based on forward or backward data, most of which based on maximum likelihood methods – PMM being one of them.

With PMM the intention is to impute missing entries of a database by drawing subsets of observations, which work as “candidates” to fill these values. “Candidates” are selected by checking whether the predicted value (of the missing value) are the closest in the entire database to the predicted value of the NA that will be imputed. Then a “donor” out of these candidates is selected at random, which will fill the missing value in the dataset (Buuren 2018) [16].

There are several alternative imputation methods that could be used over PMM. Considering how the aforementioned variables could be treated as categorical/discrete, one could use logistic regressions to evaluate and fill the missing values. These however proved to be too computational- – and thus, time- – demanding especially with AP and RP as variables with 5 and 16 values, respectively. There was also the potential and risk of running into several issues in case model parametrization was done incorrectly (e.g., in case the relationship between the missing value/estimate and the regressors were non-linear) (Nooraee et al. 2018).

Another alternative was the utilization of “random forests”, a well-known data mining and machine learning technique which helps selecting “decision trees”, created in a process of estimating and predicting values either in discrete (classification) or continuous (regression) form, to reduce variance of these “trees” to improve their overall performance (Ho 1995) . While the random forest method does not run into the parametrization issues of a logistic regression technique, it was also very computational- and time-consuming. At the same time there is no consensus on whether it is a better performing technique than either logistic regressions or PMM (Young 2017; Hapfelmeier and Ulm 2014).

Thus, PMM was selected thanks to being roughly efficient, not running into parametrization issues and being the best in terms of computational processing and time spent on assessing values. The latter helped in fully utilizing the tools made available by the mice (“Multivariate Imputation by Chained Equations”) package for the programming software R, which combines imputation techniques with multiple imputations and iterations of this technique to infer a missing value in the database (Buuren and Groothuis-Oudshoorn 2011).

Principal component analysis (PCA):

While there are several variables in the databases utilised in this research that can be good indicators of economic activity measure on their own, there are several methodologies that allow investigators to combine these variables and a set of controls to draw a general and easily interpretable picture of whatever analysis is being made.

In economics, indexes are the most common way to draw this. However, there are limitations to

[16] More details can be found in Appendix A2 – Predictive mean matching (PMM)

applying indexing methods to the databases at hand starting with the fact they do not have monthly or quarterly data. While there are ways in which these can be inferred and transformed into time series data to produce indexes, this can produce too much noise especially if variables are highly correlated, and the methodology utilised does not account for endogeneity and similar issues.

A good alternative to generating economic activity measures with the kind of data available, is using PCA. Here these measures can be produced for city-wide activity over the years assessed, combining several variables to produce "scores" which can help identify better performing years and/or regions with the databases at hand.

One issue is the potential loss of accuracy in doing this process, compared to traditional estimation methods for indexing. However, the dimensionality issues which can be solved by the method as highlighted above, and the ease of interpretation from the output which can in turn provide better information to researchers, can certainly compensate any accuracy loss from applying PCA to a dataset [17].

Thus, with PCA, the aforementioned "indexes" can be turned into "summary scores" for a number of purposes in this research given the wealth of data available, as well as potentially valuable variables that could paint this picture (Dai, Xiong, and Zhou 2021; Lai 2003; Marques and Neves 2001; Vyas and Kumaranayake 2006) . An alternative that proved valuable in easing statistical interpretation further was factor analysis (FA), which uses PCA principles to help seeing the "link" between principal components and the variables that compose them over the years for RAIS and PNAD databases (Godshalk and Timothy 1988; Blackman, Seligman, and Sogliero 1973; Chateau et al. 2012).

6.2 Sample

To assess economic effects of our variables of interest via DiD, specifically with the RAIS "vínculos" database, sample will be restricted to workers between 23 and 65 years of age – i.e., those who have just left college to workers about to enter retirement. For PNAD, the sample was restricted to respondents over the age of 23. In terms of racial breakdown, workers have been grouped between white and non-white as per Gerard et al. (2021) .

As mentioned before, treatment will be based on strongly and weakly treated regions based on average, median and standard deviation of wages in the year 2000, using the RAIS "vínculos" database. DiD will be applied on RAIS data between 2001 and 2013, alongside PMM-MICE to help infer race of workers pre-2006 and beyond.

RAIS will also be helpful in PCA determination for economic activity in each Rio region. In this,

[17] More details can be found in Appendix A3 – Principal component analysis (PCA)

we use data not just from RAIS "vínculos" but also from RAIS "estabelecimentos", which has a wealth of information on companies at the city over the years.

PNAD is also useful for PCA, as a "check" on economic effects of the minimum wage across not just the Rio de Janeiro state capital, but the metropolitan area in its entirety. Given it also covers people out of the labour market who still earn income via pensions, it can be helpful in revealing what might be lost in an outlook that focus just on companies and workers in the city – as is the case with RAIS.

6.3 Variables of interest and controls

For the RAIS database, which is the main database for this research, there are many potential variables of interest for the purposes of this work. But given time and space constraints, it is best to focus on those who could be the most telling in terms of minimum wage economic impact at the regional level, using Rio de Janeiro's labour market as a laboratory for such. Among those, we will assess:

- "**vinculo_ativo_3112**", which indicates whether a worker has kept their job at the end of the year. This could function as a measure of how much minimum wage growth influenced in entrance/exit flows in Rio's labour market over the assessed time period;
- "**indicador_simples**", to check whether minimum wage increases affected how many workers have been employed by SME's under the "Simples Nacional" tax regime;
- "**avg_income_defl**", in log form since it is an income measure, to gauge minimum wage growth impact on the salaries of Rio de Janeiro workers;
- "**gender_id**" and "**race_id**", which identify whether a worker is female or non-white, to measure if minimum wage growth affected their employment levels;
- And "**q1_id**", a dummy variable that indicates someone earning below two minimum wages, which can help assess if there were any changes in the hiring of workers at the aforementioned income level.

The equation form for treatment effect estimations of "**vinculo_ativo_3112**", "**indicador_simples**", and other binary dependent variables, will be

$$1\left\{ \text{vinculo\_ativo\_3112}_{\,ist}\right\} = \delta_k + \sum_k \beta_k \text{ Strongly}_{\,s} \times \delta_{t+k} + X'_{ist}\Gamma + \delta_s + \varepsilon_{ist} \tag{1}$$

Where:

- δ_k is the accumulated time fixed effect;
- $\sum_k \beta_k \text{ Strongly}_{\,s} \times \delta_{t+k}$ is the accumulated effect of minimum wage increases over time in

strongly treated regions of Rio – with β_k as our coefficient of interest;

- $X'_{ist}\Gamma$ are a set of control variables;
- δ_s are regional fixed effects;
- And ε_{ist} is the residual.

Whereas the equation for "**avg_income_defl**", given the log transformation of the variable, will be

$$\log\ \text{avg\_income\_defl}_{ist} = \delta_k + \sum_k \beta_k \text{ Strongly }_s \times \delta_{t+k} + X'_{ist}\Gamma + \delta_s + \varepsilon_{ist} \tag{2}$$

For the above, the R package deflateBR, which allows deflating of values based on Brazil's inflation indexes such as IPCA, was used (Meireles 2018). The chosen index was IBGE's IPCA, the main inflation measure utilised for wage adjustments, with January 2022 as the reference month [18].

Datasets from RAIS "vínculos" pre-2006 do not have one of the potential controls that could be used to this research, which is race. There is also missing data for AP and RP variables in both RAIS "estabelecimentos" and "vínculos" databases.

To solve this issue, we use PMM – as described above – to infer the racial composition of city workers, as well as fill AP and RP missing values from the databases at hand.

Programming software R has a package called "mice", which stands for "Multivariate Imputation by Chained Equations", which combines an array of methods and Markov chain Monte Carlo-like algorithms for imputation of missing values (Buuren and Groothuis-Oudshoorn 2011). The package also selects the best variables for the process, guaranteeing an acceptable success rate in filling NAs for the variables of interest in both RAIS databases.

PCA has the advantage of being flexible enough that it can be applied across a range of databases not just to create the aforementioned economic activity scores/factors, but also to gauge how impactful some of the variables available in each dataset is to paint the socioeconomic "picture" of Rio de Janeiro over the 2000s and 2010s. Still the main objective here is to use PCA to draw measures of economic activity in Rio over the years, using RAIS "vínculos" and "estabelecimentos", as well as PNAD, for that.

A PCA factor decomposition using RAIS "vínculos" to assess economic activity in each Rio de Janeiro AP and RP will include:

- "**avg_income_defl**";
- "**vinculo_ativo_3112**";

[18] Deflators can be found in page 79

- "**tempo_emprego**", which is time at employment of a certain worker;
- "**quantidade_horas_contratadas**", meaning how many hours a worker has been hired for;
- "**idade**", which is the age of a worker;
- "**tamanho_estabelecimento**", for the company's size in which the worker is currently employed;
- "**race_id**";
- "**educ_id**", which denotes how further in education a worker has gone into before taking his job;
- "**gender_id**";
- "**treat_AP**" and "**treat_RP**", to identify workers at regions that were strongly treated by the minimum wage hikes from 2003 to 2010
- "**public_work**", a dummy that denotes whether a worker is under a statutory regime rather than CLT;
- And "**q1_id**", an indicator for workers earning below two minimum wages in that year.

Whereas for RAIS "estabelecimentos", factor analysis will include:

- "**quantidade_vinculos_clt**" and "**quantidade_vinculos_estatutarios**", which shows how many people have been hired by a company either under CLT or as a statutory worker;
- "**indicador_simples**", for companies that are under the "Simples Nacional" tax regime;
- "**indicador_atividade_ano**", indicating whether the company was active that year;
- And "**treat_AP**" and "**treat_RP**", to identify companies at regions that were strongly treated by the minimum wage hikes from 2003 to 2010.

And for PNAD, the variables included in PCA will be:

- "**renda_mensal_ocupacao_principal**", "**renda_mensal_todas_fontes**" and "**renda_aposentadoria_pensao**", meaning income from work, all sources, and from pensions for an individual, respectively;
- "**idade**", "**anos_estudo**", "**gender_id**" and "**race_id**" as controls for age, education, gender and race, respectively;
- "**trabalhou_semana**", "**possui_carteira_assinada**" and "**horas_trabalhadas_todos_trabalhos**", which indicates if an individual worked that week, if he is a formal/CLT worker and how many hours he worked at all jobs he had;
- "**q1_id**";

- And "**entrep**", indicating whether the individual is an entrepreneur either owning a company, or working on his own.

The number of controls that can be used in each database is quite extensive. For RAIS "vínculos", and the assessment of variables of interest using DiD, these include:

- "**gender_id**" and "**race_id**";
- "**educ_id**";
- "**under_40**", an indicator for workers under 40 years of age;
- "**tamanho_estabelecimento**";
- And "**public_work**".

6.4 Results

Henceforth are the results from the econometric analysis utilising both RAIS "vínculos" and "estabelecimentos", as well as PNAD, to assess economic impact of minimum wage growth on Rio de Janeiro APs and RPs between 2000 and 2015.

6.4.1 AP, RP and "indicador_atividade_ano" inference using PMM-MICE in RAIS "estabelecimentos"

During assessment of descriptive statistics for the RAIS "estabelecimentos" database, it was noticed that there were a number of observations missing for both AP and RP variables, which help determine the strongly and weakly treated regions in this research. To address this issue, the PMM-MICE methodology was applied.

Given recommendations from the academic paper outlining the R package's methodology and implementation (Buuren and Groothuis-Oudshoorn 2011), as well as other sources in the literature (Morris, White, and Royston 2014; Nooraee et al. 2018) , inferences were made by using five imputations per ten iterations of the databases.

As shown in the annex [19], PMM-MICE kept distribution of AP and RP variables roughly in line with the original.

One problem with treating RAIS "estabelecimentos" databases was the break in the series during the 2002 year in geographical delimitations, in which there were no observations for neighbourhoods from which AP and RP could be inferred. Attempts at bridging the gap using databases from years after and before the year did not prove fruitful since the years 2000 and 2001 had old neighbourhood

[19] Results can be found in Annex D, subsection PMM-MICE results in the RAIS "estabelecimentos" database

codes whose reference could not be found as of August 2022.

For that reason it was decided it was best to proceed with inferring only the years 2003-2015 for RAIS "estabelecimentos" for AP, RP, and also the active company indicator whose series started in 2005. Below are the results for the two years in which values were missing: 2003 and 2014.

Table 4: PMM-MICE estimated values for "indicador_atividade_ano" in 2003 and 2004

	2003	2004
0	16.2%	18.6%
1	83.8%	81.4%
NA	0.0%	0.0%

Source: Ministério do Trabalho e Previdência. Own elaboration.

6.4.2 AP, RP and race inference using PMM-MICE in RAIS "vínculos"

Another noticeable event in the descriptive statistics/analysis stage, was that the race variable from RAIS "vínculos" was missing from data before 2006 in the open repository. Whereas for AP and RP variables, there were also some years where the number of missing entries surpassed 10%, which is above the acceptable values for such in the literature (Dong and Peng 2013; Madley-Dowd et al. 2019; Montelpare et al. 2020).

PMM-MICE was thus run first on AP and RP to fill missing values in these variables. And as shown in the annex [20], PMM-MICE kept distribution of AP and RP variables roughly in line with the original.

As for RAIS "vínculos" pre-2006, datasets of previous years were utilised to infer missing data for the "race_id" variable. Here the PMM-MICE inference showed that Rio de Janeiro's job market was still largely white in racial composition, with the "gap" closing in later years

Table 5: Comparison of white/non-white workers (0/1) in the RAIS "vínculos" database (original vs. treated)

	2000	2001	2002	2003	2004	2005	2006	2007
0	60.7%	60.5%	60.3%	60.8%	60.4%	60.6%	13.4%	14.2%
1	39.3%	39.5%	39.7%	39.2%	39.6%	39.4%	6.0%	7.7%
NA	-100.0%	-100.0%	-100.0%	-100.0%	-100.0%	-100.0%	-19.4%	-21.8%

	2008	2009	2010	2011	2012	2013	2014	2015
0	12.1%	12.5%	4.2%	11.2%	11.1%	11.0%	11.1%	11.3%
1	6.1%	6.3%	3.2%	6.2%	6.0%	6.3%	6.5%	7.0%
NA	-18.2%	-18.8%	-7.4%	-17.4%	-17.1%	-17.3%	-17.6%	-18.3%

Source: Ministério do Trabalho e Previdência. Own elaboration.

[20] Results can be found in Annex D, subsection PMM-MICE results in the RAIS "vínculos" database

There was, however, one year in the dataset that proved difficult to fill in missing "race_id" entries: 2010. Using alternative methods such as removing collinear variables – which had already been done with other databases –, and increasing number of imputations and iterations, did not help solve the issue.

6.4.3 DiD estimates for RAIS "vínculos", 2001 to 2013

The reason for a shorter time frame in this analysis is 1) the fact one of the independent variables being assessed, "indicador_simples", has no entries before 2001; and 2) Brazil's major economic crisis in 2014 from which the country only started recovering in 2017 (Brinca and Costa-Filho 2021; Vartanian and Garbe 2018; European Central Bank 2016), which could add unwarranted noise to an analysis of minimum wage growth impact – a policy that took place from 2003 to 2010.

Another thing of note is that since the policy – and thus the treatment – takes place over a range of years, it would be problematic to estimate average effect of the treatment on the treated on a static model as is the case in simple DiD estimations. Therefore this research chose a dynamic effect estimation approach, which can also be viewed as an event studies analysis for periods before and beyond the treatment's first impact on Rio de Janeiro's labor market.

As the literature reveals in recent developments (Borusyak, Jaravel, and Spiess 2022; Callaway, Goodman-Bacon, and Sant'Anna 2021; de Chaisemartin and D'Haultfoeuille 2020; Goodman-Bacon 2021; Sun and Abraham 2021), ordinary least squares estimation of two-way fixed effects (TWFE) DiD models either on static or dynamic forms can be problematic. An alternative, provided by Butts and Gardner (2022) and Borusyak, Jaravel, and Spiess (2022) and implemented by Butts (2021) via R package "did2s" is to first estimate fixed effects (and potential regressors) on untreated/not-yet-treated units against the dependent variable, taking its residuals and estimating them against all observations to get group-time average treatment effect(s) [21].

Taking the above into consideration, here are the treatment effects results for each of the dependent variables assessed in the research at the AP level [22], based on second-stage FE estimations of treated units.

[21] First-stage results can be found in Annex D, subsection DiD first-stage results - AP & RP level

[22] RP level results can be found in Annex D, subsection DiD second-stage results - RP level

Table 6: Dependent variable: Log Average Income (deflated) – second-stage estimation, AP level

	Model 1	Model 2	Model 3
fixest = rel_year = -2	0.008	0.007	0.006
	(0.006)	(0.007)	(0.006)
fixest = rel_year = 0	-0.038**	-0.030**	0.018***
	(0.006)	(0.005)	(0.002)
fixest = rel_year = 1	-0.084	-0.069	0.028***
	(0.040)	(0.034)	(0.002)
fixest = rel_year = 2	-0.047	-0.036	0.048**
	(0.043)	(0.037)	(0.007)
fixest = rel_year = 3	-0.040	-0.025	0.022*
	(0.045)	(0.038)	(0.007)
fixest = rel_year = 4	-0.031	-0.027	-0.020*
	(0.039)	(0.032)	(0.006)
fixest = rel_year = 5	-0.046	-0.040	-0.003
	(0.049)	(0.043)	(0.010)
fixest = rel_year = 6	-0.018	-0.008	0.029
	(0.050)	(0.042)	(0.014)
fixest = rel_year = 7	0.088	0.064	0.043
	(0.081)	(0.060)	(0.022)
fixest = rel_year = 8	-0.010	-0.007	0.037**
	(0.043)	(0.034)	(0.005)
fixest = rel_year = 9	-0.041	-0.033	0.038**
	(0.046)	(0.036)	(0.005)
fixest = rel_year = 10	-0.075	-0.070	0.001
	(0.047)	(0.036)	(0.008)
Num.Obs.	33830503	33339651	33339651
R2	0.001	0.001	0.0006
R2 Adj.	0.001	0.001	0.0006

Notes: standard errors clustered by AP

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Source: Own elaboration.

Table 7: Dependent variable: Worker at "Simples Nacional" company (Y/n) – second-stage estimation, AP level

	Model 1	Model 2	Model 3
fixest = rel_year = -2	0.021***	0.021***	0.021***
	(0.001)	(0.001)	(0.0009)
fixest = rel_year = 0	0.057***	0.057***	0.054***
	(0.001)	(0.001)	(0.002)
fixest = rel_year = 1	0.071**	0.070**	0.057***
	(0.010)	(0.009)	(0.0007)
fixest = rel_year = 2	0.086***	0.086***	0.072***
	(0.010)	(0.010)	(0.004)
fixest = rel_year = 3	0.086**	0.086**	0.075***
	(0.017)	(0.017)	(0.006)
fixest = rel_year = 4	0.086**	0.087**	0.079***
	(0.015)	(0.015)	(0.008)
fixest = rel_year = 5	0.072**	0.072**	0.073***
	(0.014)	(0.014)	(0.004)
fixest = rel_year = 6	0.074+	0.075+	0.074*
	(0.028)	(0.028)	(0.019)
fixest = rel_year = 7	0.070	0.075+	0.085*
	(0.036)	(0.030)	(0.019)
fixest = rel_year = 8	0.094*	0.095*	0.092**
	(0.022)	(0.022)	(0.012)
fixest = rel_year = 9	0.070**	0.071**	0.070***
	(0.012)	(0.012)	(0.002)
fixest = rel_year = 10	0.070**	0.071**	0.067***
	(0.015)	(0.015)	(0.004)
Num.Obs.	34334762	33842121	33842121
R2	0.010	0.010	0.012
R2 Adj.	0.010	0.010	0.012

Notes: standard errors clustered by AP

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Source: Own elaboration.

Table 8: Dependent variable: Worker earning under 2 minimum wages/month (Y/n) – second-stage estimation, AP level

	Model 1	Model 2	Model 3
fixest = rel_year = -2	-0.005+	-0.005	-0.005+
	(0.002)	(0.002)	(0.002)
fixest = rel_year = 0	0.033***	0.030***	0.013*
	(0.001)	(0.001)	(0.003)
fixest = rel_year = 1	0.053*	0.048*	0.007**
	(0.018)	(0.016)	(0.001)
fixest = rel_year = 2	0.048+	0.045+	0.008*
	(0.019)	(0.017)	(0.002)
fixest = rel_year = 3	0.058+	0.053+	0.024**
	(0.023)	(0.021)	(0.005)
fixest = rel_year = 4	0.067*	0.067*	0.055**
	(0.022)	(0.020)	(0.006)
fixest = rel_year = 5	0.065	0.064+	0.048*
	(0.031)	(0.029)	(0.013)
fixest = rel_year = 6	0.055	0.053	0.032
	(0.034)	(0.031)	(0.017)
fixest = rel_year = 7	0.002	0.013	0.027
	(0.054)	(0.038)	(0.019)
fixest = rel_year = 8	0.052	0.052	0.030+
	(0.030)	(0.026)	(0.013)
fixest = rel_year = 9	0.072+	0.070+	0.042*
	(0.031)	(0.028)	(0.012)
fixest = rel_year = 10	0.082+	0.082*	0.053*
	(0.032)	(0.029)	(0.015)
Num.Obs.	34334762	33842121	33842121
R2	0.004	0.004	0.002
R2 Adj.	0.004	0.004	0.002

Notes: standard errors clustered by AP

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Source: Own elaboration.

Table 9: Dependent variable: Female worker (Y/n) – second-stage estimation, AP level

	Model 1	Model 2	Model 3
fixest = rel_year = -2	-0.002	-0.002	-0.001
	(0.002)	(0.002)	(0.002)
fixest = rel_year = 0	0.004	0.005	0.015*
	(0.004)	(0.004)	(0.005)
fixest = rel_year = 1	-0.039	-0.038	-0.011
	(0.021)	(0.020)	(0.009)
fixest = rel_year = 2	-0.063*	-0.062*	-0.036*
	(0.021)	(0.020)	(0.011)
fixest = rel_year = 3	-0.057*	-0.055*	-0.033*
	(0.019)	(0.019)	(0.010)
fixest = rel_year = 4	-0.058*	-0.057*	-0.045**
	(0.018)	(0.017)	(0.009)
fixest = rel_year = 5	-0.037*	-0.036*	-0.022*
	(0.013)	(0.013)	(0.006)
fixest = rel_year = 6	-0.060**	-0.058**	-0.038**
	(0.012)	(0.011)	(0.005)
fixest = rel_year = 7	-0.028	-0.046*	-0.055***
	(0.035)	(0.016)	(0.005)
fixest = rel_year = 8	-0.046*	-0.045*	-0.032*
	(0.016)	(0.015)	(0.007)
fixest = rel_year = 9	-0.033	-0.032	-0.019+
	(0.017)	(0.016)	(0.008)
fixest = rel_year = 10	-0.026	-0.025	-0.015
	(0.017)	(0.016)	(0.009)
Num.Obs.	34334762	33842121	33842121
R2	0.002	0.002	0.001
R2 Adj.	0.002	0.002	0.001

Notes: standard errors clustered by AP

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Source: Own elaboration.

Table 10: Dependent variable: Non-white worker (Y/n) – second-stage estimation, AP level

	Model 1	Model 2	Model 3
fixest = rel_year = -2	0.0003	0.0002	0.0006
	(0.0002)	(0.0003)	(0.0003)
fixest = rel_year = 0	0.015***	0.015***	0.002*
	(0.0008)	(0.0009)	(0.0005)
fixest = rel_year = 1	0.025*	0.021*	-0.002*
	(0.008)	(0.006)	(0.0005)
fixest = rel_year = 2	0.021*	0.016*	-0.002+
	(0.007)	(0.005)	(0.0008)
fixest = rel_year = 3	0.033*	0.029*	0.019*
	(0.011)	(0.010)	(0.006)
fixest = rel_year = 4	0.016	0.012	0.015
	(0.014)	(0.012)	(0.008)
fixest = rel_year = 5	0.028+	0.026+	0.014
	(0.012)	(0.011)	(0.007)
fixest = rel_year = 6	0.038+	0.035	0.023
	(0.018)	(0.017)	(0.012)
fixest = rel_year = 7	0.012	0.013	0.011
	(0.025)	(0.024)	(0.017)
fixest = rel_year = 8	0.025	0.023	0.011
	(0.023)	(0.021)	(0.017)
fixest = rel_year = 9	0.033	0.031	0.011
	(0.024)	(0.022)	(0.017)
fixest = rel_year = 10	0.029	0.028	0.009
	(0.023)	(0.022)	(0.018)
Num.Obs.	33842121	33842121	33842121
R2	0.0007	0.0006	0.0002
R2 Adj.	0.0007	0.0006	0.0002

Notes: standard errors clustered by AP

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Source: Own elaboration.

Table 11: Dependent variable: Still at work by end of the year (Y/n) – second-stage estimation, AP level

	Model 1	Model 2	Model 3
fixest = rel_year = -2	-0.002	-0.002	-0.0008
	(0.001)	(0.001)	(0.001)
fixest = rel_year = 0	0.006*	0.007*	0.005**
	(0.002)	(0.002)	(0.0008)
fixest = rel_year = 1	-0.025	-0.019	-0.0008
	(0.015)	(0.012)	(0.0008)
fixest = rel_year = 2	-0.027+	-0.020*	0.004
	(0.010)	(0.007)	(0.003)
fixest = rel_year = 3	-0.031+	-0.024+	-0.001
	(0.013)	(0.011)	(0.006)
fixest = rel_year = 4	-0.027+	-0.024+	-0.004
	(0.011)	(0.009)	(0.005)
fixest = rel_year = 5	-0.009	-0.007	-0.0005
	(0.011)	(0.009)	(0.005)
fixest = rel_year = 6	-0.028***	-0.025***	-0.010
	(0.002)	(0.0008)	(0.010)
fixest = rel_year = 7	0.062*	0.101***	0.062**
	(0.021)	(0.003)	(0.011)
fixest = rel_year = 8	-0.036***	-0.035***	-0.019
	(0.002)	(0.0009)	(0.010)
fixest = rel_year = 9	-0.040*	-0.038**	-0.018*
	(0.010)	(0.007)	(0.004)
fixest = rel_year = 10	-0.030*	-0.030**	-0.012+
	(0.009)	(0.006)	(0.004)
Num.Obs.	34334762	33842121	33842121
R2	0.002	0.003	0.001
R2 Adj.	0.002	0.003	0.001

Notes: standard errors clustered by AP

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Source: Own elaboration.

RP level results and first-stage estimations can be found in the annex.

6.4.4 PCA for RAIS "vínculos", 2000 to 2015

For PCA analysis in RAIS "vínculos", observations were first filtered by age where workers between 23 and 65 years of age were kept in the dataset – similar to what had been done in the DiD analysis. Any values carrying NAs were also excluded as a way not to cause issues with running PCA/factor analysis, which still left over 2 million observations for analysis in each year.

Checking for correlation between variables, none of them were over $|0.8|$, which means there were no candidates for exclusion *ex post*. Results from the Kaiser-Meyer-Olkin (KMO) factor adequacy test were > 0.7 for each year, whereas Bartlett's test for sphericity – in which the null hypothesis means the data is essentially non-correlated, and thus impossible for any kind of PCA (Gorsuch 1973; Kaiser 1970) – had a p-value of 0.

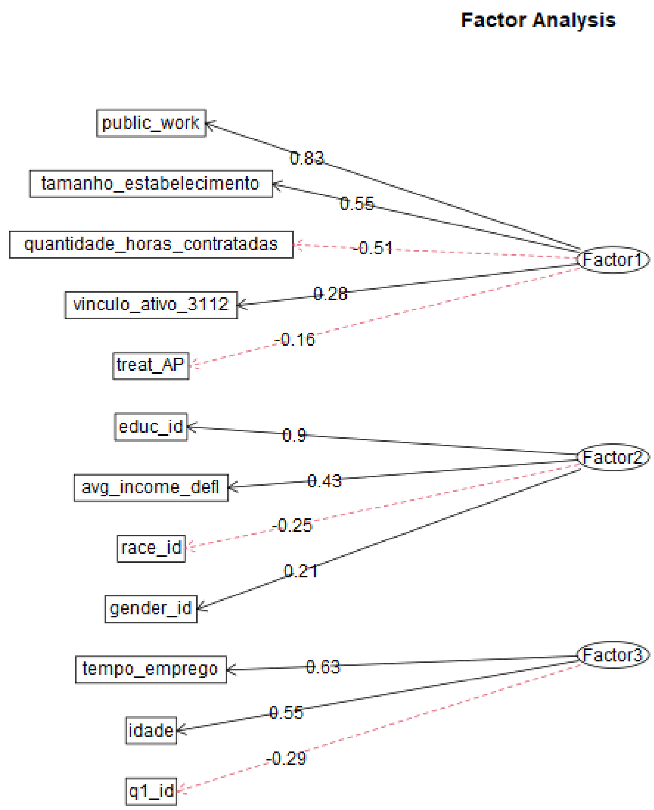
There were three eigenvectors in the dataset with eigenvalues > 1, which under Kaiser's rule mean the research should stick to three factors for analysis (Kaiser 1960). While parallel analysis (Franklin et al. 1995) suggested this could proceed with five factors, it was decided to be best to stick with a conservative number of them.

Although correlation between most variables in the dataset was low, and the assumption of independent factors could be taken from the get-go, promax rotation – generally used for large databases with highly correlated data – was first attempted to confirm the suspicion that factors would be independent. Upon confirmation of that fact the process continued with varimax rotation, which is better suited not just for independent factors but also to make them clearer for interpretation (Watkins 2018).

These are the results for factor analysis in the RAIS "vínculos" database in 2001 and 2015, using varimax rotation on three factors with "**treat-AP**" as a regional indicator [23].

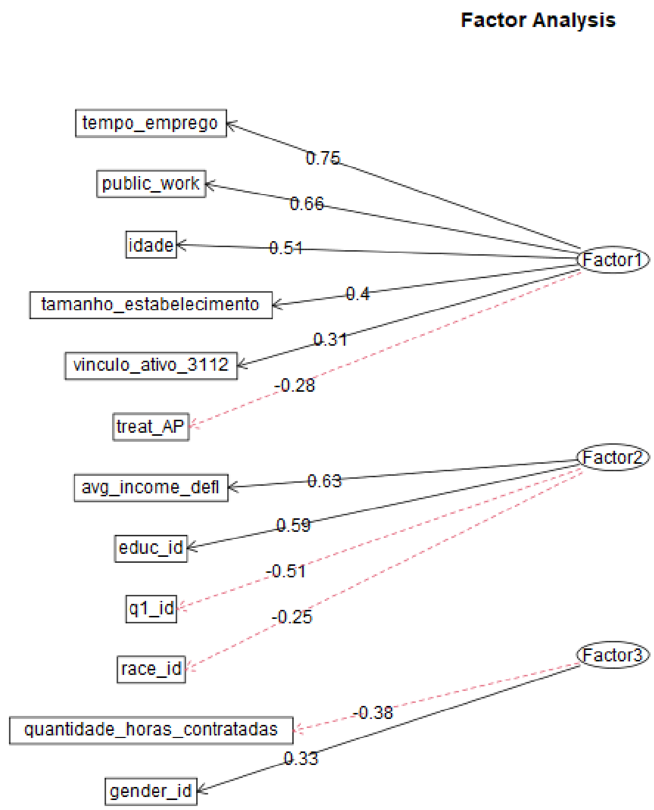
[23] RP level results and factor loadings from years 2000 to 2015 can be found in Annex D, subsection PCA results - RAIS "vínculos" from 2000 to 2015

Figure 8: Factor decomposition for AP treatment in 2000, in RAIS "vínculos" database



Source: Own elaboration.

Figure 9: Factor decomposition for AP treatment in 2015, in RAIS "vínculos" database



Source: Own elaboration.

6.4.5 PCA for RAIS "estabelecimentos", 2003 to 2015

For PCA analysis in RAIS "estabelecimentos", observations were unfiltered. However any values carrying NAs were still excluded. That still left over 200 thousand observations for analysis in each year.

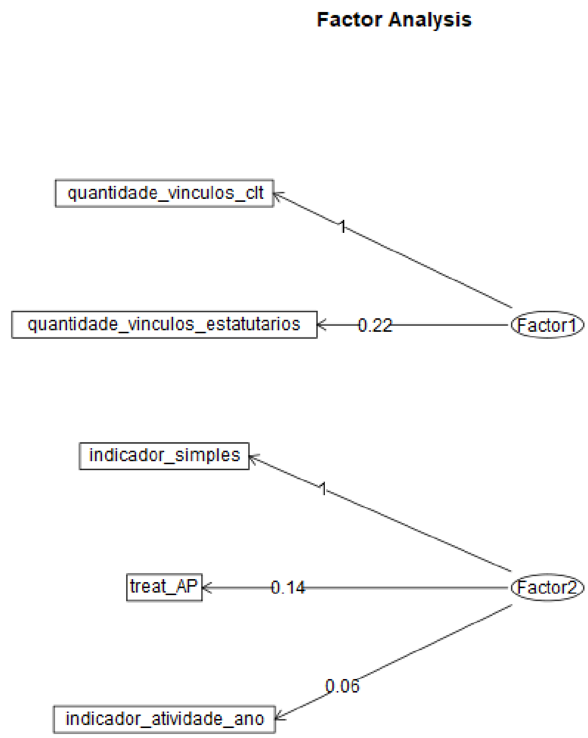
Checking for correlation between variables, none of them were over $|0.8|$. Results from the Kaiser-Meyer-Olkin (KMO) factor adequacy test were > 0.5 for each year, whereas Bartlett's test for sphericity had a p-value of 0.

There were two eigenvectors in the dataset with eigenvalues > 1, which under Kaiser's rule mean the research should stick to two factors for analysis. While parallel analysis suggested this could proceed with three factors, it was decided to be best to stick with a conservative number of them – as done with RAIS "vínculos".

Although correlation between most variables in the dataset was also low in RAIS "estabelecimentos", and the assumption of independent factors could be taken from the get-go, promax rotation was first attempted to confirm the suspicion that factors would be independent. Upon confirmation of that fact the process continued with varimax rotation.

Below are the results for factor analysis in the RAIS "estabelecimentos" database in 2003 and 2015, using varimax rotation on three factors with "**treat-AP**" as a regional indicator [24].

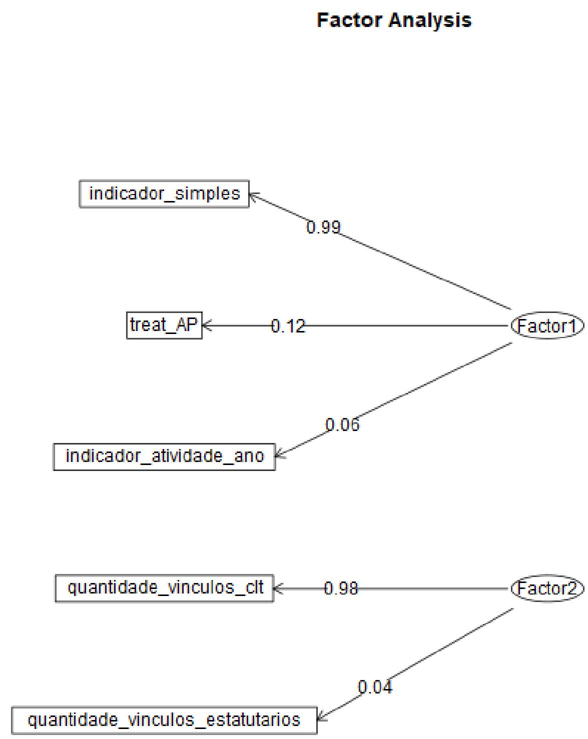
Figure 10: Factor decomposition for AP treatment in 2003, in RAIS "estabelecimentos" database



Source: Own elaboration.

[24] RP level results and factor loadings from years 2003 to 2015 can be found in Annex D, subsection PCA results - RAIS "estabelecimentos" from 2003 to 2015

Figure 11: Factor decomposition for AP treatment in 2015, in RAIS "estabelecimentos" database



Source: Own elaboration.

6.4.6 PCA for PNAD, 2001 to 2015

For PCA analysis in PNAD "individual", observations were filtered for those over the age of 23, whether they are part of the labour market or not. Any values carrying NAs were also excluded, leaving the research with roughly 10 thousand observations in each year.

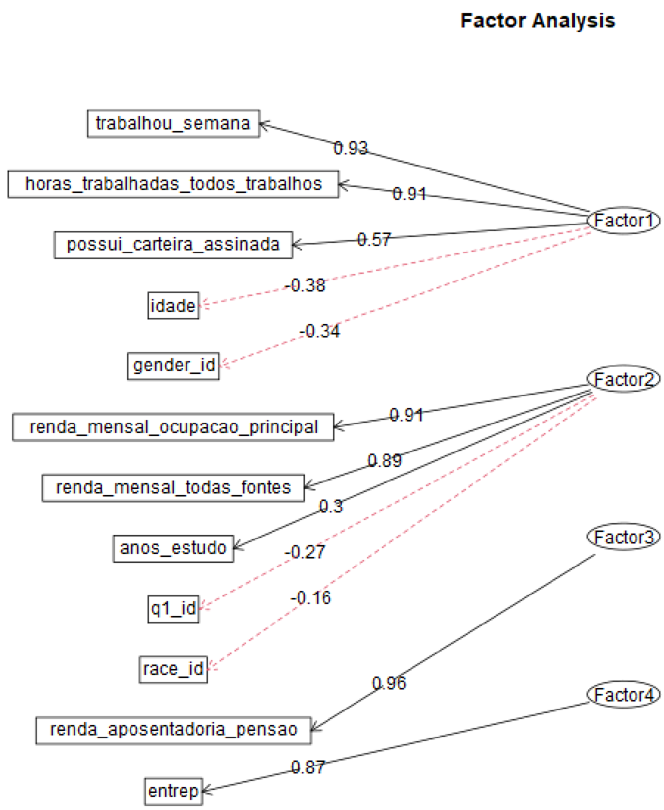
Checking for correlation between variables, none of them were over $|0.8|$. Results from the Kaiser-Meyer-Olkin (KMO) factor adequacy test were > 0.5 for each year, whereas Bartlett's test for sphericity had a p-value of 0.

There were four eigenvectors in the dataset with eigenvalues > 1, which under Kaiser's rule mean the research should stick to four factors for analysis. While parallel analysis suggested this could proceed with six factors, it was decided to be best to stick with a conservative number of them – as done with the two previous databases.

Although correlation between most variables in the dataset was low, and the assumption of independent factors could also be taken from the get-go, promax rotation was first attempted to confirm the suspicion that factors would be independent. Upon confirmation of that fact the process continued with varimax rotation.

Below are the results for factor analysis in the PNAD "individual" database in 2001 and 2015, using varimax rotation on four factors [25].

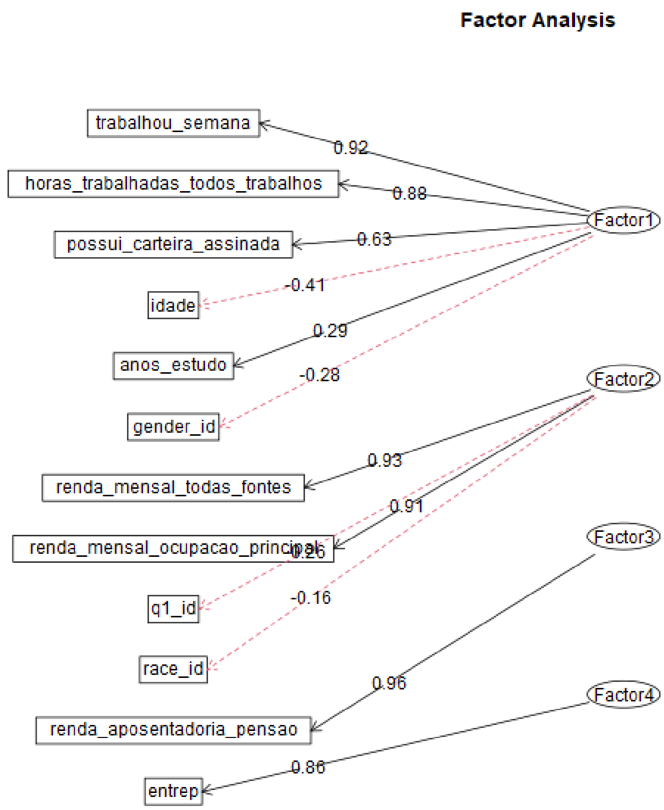
Figure 12: Factor decomposition in 2001, in PNAD "individual" database



Source: Own elaboration.

[25] Factor loadings from years 2001 to 2015 can be found in Annex D, subsection PCA results - PNAD from 2001 to 2015

Figure 13: Factor decomposition in 2015, in PNAD "individual" database



Source: Own elaboration.

6.4.7 Discussion

DiD results for RAIS "vínculos"

Under the complete model (no.3 in the regression tables above), estimations of the treatment effect at AP and RP levels do not differ much either in signal or in estimated impact. In total there were over 33 million observations assessed to gauge whether/how minimum wage growth impacted workers in earnings, and employment status.

The results for earnings are not as strong as in the literature for earnings, as the differential uptick for treated units during treatment was up to 4.9% in 2005 under the full model which accounts for race, gender, age, education, working status (CLT vs. statutory) and company size in the first-stage estimation. Despite a 2% loss of income in 2007, treated units maintained at least 2% gains in earnings two years after treatment, to a total of approximately 27% growth in real income in comparison to non-treated units if they were still working at strongly treated areas over an 11-year span.

First-stage estimation of average income takes an interesting turn between model no.1, assessing only racial, gender and age components; and the complete model, no. 2. Non-white and female workers take 32.8% and 14.8% "hits" to their income in comparison to white and male workers, respectively, when only age is estimated alongside then. But once we control for public work, education and firm size, the racial component negative impact is reduced to 11.5%, and the gender component increases to 30.3%.

Perhaps more noticeable is the increase in employment of workers in "Simples Nacional" companies – i.e, small and medium businesses – in strongly treated regions. From 2003 to 2013 the likelihood of being a worker at such a company at a strongly treated AP region is 5.4% to 9.2% higher than in weakly treated ones. This is in tune with absolute growth in "Simples Nacional" companies specially, as seen in the descriptive analysis section of this thesis and also in PCA/FA.

Unsurprisingly, the number of workers earning under 2 minimum wages also grew in strongly treated regions. It is not a surprising development given the "Simples Nacional" assessment above, as well as first-stage estimate results that reveal how race, gender and age are strong determinants of being at this income level in Rio's labour market.

Another interesting result is how treatment impacted employment of women and non-white workers in strongly treated regions. Whereas the effect on non-white members of the labour force was largely insignificant outside of a 1.9% uptick in hiring in 2006, in comparison to those in weakly treated regions, presence of female workers decreased by up to 5.5% during the years assessed.

Here first-stage estimations are also revealing as female workers are more likely to be found in public work under statutory regimes, with high returns on education and positive impacts across all types of firm size. That is not the case for non-white workers, with firm size playing little to impact in their hiring up until the higher echelon of companies, which hire 1,000+ workers.

Finally, the treatment had no significant impact on workers' ability to maintain themselves at a company by the end of the year. First-stage estimation gives a similar picture, where the bigger the firm the likeliest for a worker to stay at their job.

PCA/FA for RAIS "vínculos"

Three years before treatment began, RAIS "vínculos" has the first and foremost factor – Rio's labour market "health" – largely influenced by the public sector. That is not a surprise, given how influential it has been in Rio de Janeiro's economy since Brazil's years as a Portuguese colony. Nor is it surprising that public workers are hired by the largest "companies" in the city, the city and state governments, all while working fewer hours than private sector counterparts and having more job security to boot. And since AP 1 - Centro is where these jobs are officially located, it is no wonder "**treat_AP**" and "**treat_RP**" enter the first RAIS "vínculos" factor as a negative contributor.

The second most influential factor in the database back in 2000 can be interpreted as the importance of returns to education to a worker, especially when it comes to employment. Interestingly being a non-white workers leads not just to being less educated than white counterparts but also to earning fewer income in Rio's labour market. Whereas women in Rio's labour market get higher returns to education in comparison to male counterparts in hiring likelihood as also seen in DiD estimation, although this is likely helped by their below-40% presence in it.

The third and last factor in 2000 is in itself an indicator of job longevity, another important determinant of wages in Rio's labour market given public worker's influence in the city's labour market state. The older you are, the longer you are employed, and the more likely it is that you are not one of the people earning fewer than two minumum wages per month on average in the city – these are younger folks usually entering firms outside of weakly treated areas in Rio de Janeiro.

Fifteen years later, in 2015, time at employment – which has diminished on average in the city, as shown by the descriptive analysis – "jumps" to become the largest contributor to Rio de Janeiro's labour market "health" alongside public workers and age. This could be an effect of the aforementioned economic crisis that hit Brazil in 2014, with private workers suffering the most with unemployment and lost income. "**treat_AP**" remains a negative contributor since strongly treated regions largely employ younger workers in small enterprises, whose job security is also frailer in comparison to other regions in the city.

The second factor in 2015 remains as a measure of returns to education in the city, although with income overtaking employment likelihood. There is little change when it comes to race affecting both employment and income. In the meantime, those earning fewer that two minimum wages on average per month become a more influential slice of Rio's labour market during an economic crisis throughout the country.

Once a component of job longevity, now the third factor in Rio de Janeiro's labour market in 2015 is more related to hours worked – something that can also be seen as linked to public workers. Same is the case with women being more likely to enter contracts with fewer than 44 hours per week, which is the norm for statutory workers, given how they are more likely to be employed under such a regime as per DiD analysis.

There is an interesting difference in the analysis at the RP-level for Rio de Janeiro in the year 2015, which is "**treat_RP**" contributing negatively to the income/educational factor in the analysis. Everywhere else it stays the same, although with some differences in "weights" of each variable's contribution to the estimated factors.

PCA/FA for RAIS "estabelecimentos"

The low number of variables for PCA lead to a diminished number of factors. But that does not mean they are not as important as in other databases, especially when looking at "**treat_AP**" and "**treat_RP**" influences.

First year in the analysis, 2003, has the first factor as employment numbers in the city. As expected the largest contributor are CLT workers, with statutory ones being roughly 1/5th of them in terms of weight in the aforementioned factor.

The interesting picture is within the second factor, which could be interpreted as an entrepreneurial measure with "**indicador_simples**" – as in, SME under the "Simples Nacional" tax regime – as the largest contributor. What is quite telling is that both geographical indicators enter this factor positively in 2003, as well as "**indicador_atividade_ano**".

Then in 2015, the picture flips with the aforementioned entrepreneurial measure overtaking its employment counterpart. In the latter statutory workers become a much weaker contributor, which is on par with the diminishing number of employees under the statutory regime as years go by in the city's labour market.

PCA/FA for PNAD

Since PNAD also captures non-workers – either those still studying, or those removed from the labour force due to age/retirement, accidents and other factors that lead them to becoming pensionists – it is a very useful survey to capture the entire Rio de Janeiro's metropolitan area economic outlook. While it has the disadvantage of not having AP/RP geographical definitions, nor a city-wide one, it can still be a way to show how the region evolved over the years in this analysis.

For that matter, PCA in 2001 for PNAD shows how labour market participants are the largest contributors to Rio de Janeiro's economy. That is especially true to those under CLT, who also work more often than their "informal" counterparts who are more often than not under "piecemeal" working regimes.

Two interesting things of note in the first PNAD factor in 2001 are age as a negative contributor, and gender as well. The latter can be understood as an effect of the gender pay gap (Morello and Anjolim 2021), whereas the latter is likely due to the aforementioned pensionists who have been out of the labour market due to retirement.

As is the case with RAIS "vínculos", income returns to education for year 2001 are somewhat strong in Rio's metropolitan area. Here both "**q1_id**" and race are negative contributors.

The last two factors in 2001 analysis are also quite interesting. The third factor is income from pensions, which is an important component of demand at the regional level. There are many families in Brazil, and also in Rio, whose main income source are pensions from their older members (Pamplona and Cucolo 2019). PCA shows that they are a stronger contributor than entrepreneurs, which are those either working by themselves as freelancers/liberal professionals, or employers.

In 2015 the scenario undergoes a few changes. While people in the labour market remain the largest contributor to Rio's economy, years of studying and gender enter the first factor as positive and negative contributors to labour market participation, respectively. The income factor is now only negatively affected by "**q1_id**" and race, whereas the picture for pensionists' and entrepreneurs' contributions to the city's economy remain the same even in "loading" terms.

7 Conclusion

This master thesis assesses how Brazil's minimum wage policy from 2003 to 2010, which saw major growth at the base level of earnings for workers in the country, affected economic outcomes of Rio de Janeiro's state capital workers, which live in a very diverse and highly unequal region.

By using a difference-in-differences (DiD) design inspired by Derenoncourt et al (2021)'s specification utilising strongly and weakly impacted areas given income levels pre-treatment, this dissertation finds that minimum wage growth had a positive impact on earnings of workers in strongly treated regions of up to 27% for units treated over a 11-year span, while not having major negative impacts in hiring of non-white workers. It also boosted employment at small and medium enterprises (SMEs) at regions more susceptible to the impact of minimum wage growth policy.

And by using principal component analysis (PCA) and factor analysis (FA) to capture economic activity development at the microlevel, this research finds that over time, strongly treated regions become entrepreneurial hubs thanks to small and medium enterprises that are created within them. This could be explained by the demand pull of workers who live in strongly treated regions, which are oftentimes "bedroom neighbourhoods", who seek to be provided with better commerce options closer to home.

At the same time pensionists are very influential in Rio's city and metropolitan area on the demand side. This could be explained by the fact Rio still has a large influence of public workers in the city – as also shown by PCA/FA – with large pensions in comparison to the rest of the country. These pensions are often used as main income for several families, helping younger folks stay in school for better returns in education in the future.

Further research on this subject would include more methodologies to assess robustness of these findings, as well as expanding the assessed regions (from the state's capital to the entire metropolitan area). The research design for both employment effects and economic activity could also be applied to different regions of Brazil such as São Paulo, Brasília and Fortaleza, which also allow for economic analysis at the neighbourhood level via the databases used in this dissertation.

Appendix

Appendix A – Econometric methodologies

Appendix A1 – Difference-in-differences (DiD)

A DiD estimator can be defined as first the difference within groups after and before treatment; and then the difference between the two groups, such as the equation below:

$$\beta_{\text{DiD}} = [\bar{Y}^{\text{treat,after}} - \bar{Y}^{\text{treat,before}}] - [\bar{Y}^{\text{control,after}} - \bar{Y}^{\text{control,before}}]$$

The four groups in the DiD estimator are:

- $\bar{Y}^{\text{control,before}}$ – average outcome of control group before the experiment;
- $\bar{Y}^{\text{control,after}}$ – average outcome of control group after the experiment;
- $\bar{Y}^{\text{treat,before}}$ – average outcome of treatment group before the experiment;
- $\bar{Y}^{\text{treat,after}}$ – average outcome of treatment group after the experiment.

The general form of a DiD equation is:

$$Y_{it} = \beta_0 + \beta_1 \text{after}_t + \beta_2 \text{treat}_i + \beta_3 (\text{treat} \times \text{after})_{it} + u_{it}$$

- Where: treat_i is a dummy variable equal to 1, if the observation belongs to the treatment group;
- And after_t is also a dummy variable equal to 1, if the observation takes place after the treatment.

To work properly, DiD needs three basic assumptions:

- **Common time trend** (CTT), which is the basis as to why treatment and control groups would have evolved in the same way in the absence of treatment – and also allows the DiD estimator to draw differences within and then between treatment and control groups;
- **Stable unit treatment value**, where observations in the control group are unaffected by assignment of treatment to the other observations – meaning control group observations are independent from treatment, and generate no spill-over nor general equilibrium effects;
- And **no anticipatory effects**, whereby the treatment group cannot change their behaviour in anticipation of the treatment

But since these can be too rigid – especially the common time trend assumption – one way to overcome limitations is to condition on covariates that make said assumption hold. Thus the estimation can take the form:

$$\Delta Y_i = \beta_0 + \beta_1 X_i + \beta_2 W_{1i} + ... + \beta_{r+1} W_{ri} + u_i$$

Where:

- X_i is the set of covariates that can help conditioning CTT;
- And $W_{1i}, ..., W_{ri}$ is a set of regressors that allow checking and adjusting for conditional randomization, as well as improving overall efficiency of the estimation. (Cunningham 2021).

To reign in on CTT, a two-way fixed-effects (TWFE) is often utilised as a way to draw the average effect of the treatment on the treated (ATT), assuming said effect is constant due to non-heterogeneity across groups and over time. That however is also a very rigid assumption.

An alternative that allows heterogeneity across groups and over time is to draw group-time average treatment effects instead, in which a treatment coefficient τ_{gt} would be allowed to vary. From this coefficient an average can be drawn to find the overall average treatment effect, τ , from the sum of the averages of group-time pairs N_{gt} .

Still, even the TWFE in its "static" form does not deliver consistent estimate for overall average treatment effect due to negative weighting from $\hat{\tau} \equiv \sum_{g,t} w_{gt} \hat{\tau}_{gt}$. That is due to the weight of estimated treatment effects for each group-time pair, w_{gt} , being often inconstant due to timing of treatment and/or heterogeneity across group and time (Borusyak, Jaravel, and Spiess 2022; de Chaisemartin and D'Haultfoeuille 2020; Goodman-Bacon 2021; Sun and Abraham 2021).

One way to further address heterogeneity issues, is to deploy dynamic/event-study estimations. These try to assess average treatment effect of being exposed for k periods, but they also fail for the same aforementioned reasons (Sun and Abraham 2021).

A solution to this issue provided by Gardner (2021) and implemented by Butts and Gardner (2022) is to employ a robust two-stage difference-in-differences estimation, for both static and dynamic models. This estimation differs from the traditional TWFE by avoiding joint estimation of group and time effects, and ATT.

The first step is to estimate a $y_{igt} = \delta_t + \delta_g + \varepsilon_{igt}$ model using only untreated/not-yet-treated observations to form $\tilde{y}_{igt} = y_{igt} - \delta_t - \delta_g$, which is in practice the residuals of the former equation. The second step is to regress $\tilde{y}_{igt}$ on treatment status across all observations, which enables the researcher to find treatment effects either on static of dynamic form.

Since standard errors from treatment will be incorrect due to $\tilde{y}_{igt}$ being an estimate itself, the estimator takes the form of a two-stage GMM estimator (Butts and Gardner 2022), for asymptotically correct standard errors. And covariates can be added as well, for same reasoning as above – i.e, conditioning randomization of treatment and improving overall efficiency of the estimation (Butts and Gardner 2022; Borusyak, Jaravel, and Spiess 2022; Gardner 2021).

Appendix A2 – Predictive mean matching (PMM)

In PMM, estimation of missing entries takes values only within the range of observed data elsewhere, thus limiting the room for unrealistic or meaningless imputations. It also does not have an explicit model, reducing the chance of misspecification that arises from parametric estimation techniques (Buuren 2018)

One reason why PMM is so attractive, beyond its efficiency in estimating values (Young 2017), is the possibility of combining the technique with MICE (Multivariate Imputation by Chained Equations) by using the "mice" package in R. In practice, MICE works similarly to a Markov chain Monte Carlo method/algorithm by processing

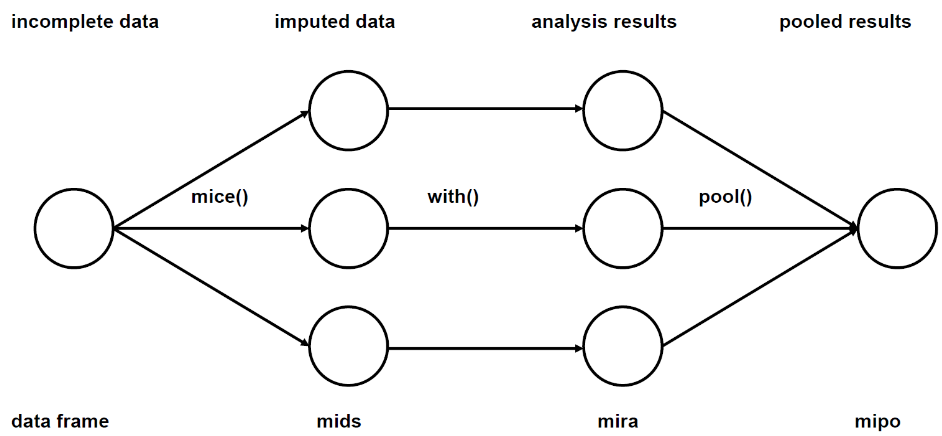
$$
\begin{gathered}
P\left(Y_1 \mid Y_{-1}, \theta_1\right) \\
\vdots \\
P\left(Y_p \mid Y_{-p}, \theta_p\right)
\end{gathered}
$$

Where Y is the complete data of a partially observed random sample, and $P(Y \mid \theta)$ is the p-variate multivariate distribution. The key is finding θ, which is a vector of unknown parameters that completely specify the distribution of Y. By doing the iterative sampling of conditional distributions given by the expression above, MICE gets the posterior distribution of θ to impute missing values that will complete Y. This takes the form

$$
\begin{aligned}
\theta_1^{*(t)} & \sim P\left(\theta_1 \mid Y_1^{\text{obs}}, Y_2^{(t-1)}, \ldots, Y_p^{(t-1)}\right) \\
Y_1^{*(t)} & \sim P\left(Y_1 \mid Y_1^{\text{obs}}, Y_2^{(t-1)}, \ldots, Y_p^{(t-1)}, \theta_1^{*(t)}\right) \\
& \vdots \\
\theta_p^{*(t)} & \sim P\left(\theta_p \mid Y_p^{\text{obs}}, Y_1^{(t)}, \ldots, Y_{p-1}^{(t)}\right) \\
Y_p^{*(t)} & \sim P\left(Y_p \mid Y_p^{\text{obs}}, Y_1^{(t)}, \ldots, Y_p^{(t)}, \theta_p^{*(t)}\right)
\end{aligned}
$$

Where $Y_j^{(t)} = \left(Y_j^{obs}, Y_j^{*(t)}\right)$ is the j-th imputed variable at iteration t. That way MICE converges quite fast to the real Y, given enough imputations and iterations with the data (Buuren and Groothuis-Oudshoorn 2011).

Figure 14: "mice" scheme when running three imputations at the same time



Source: Buuren and Groothuis-Oudshoorn (2011)

Appendix A3 – Principal component analysis (PCA)

The key concept of PCA is the principal component (PC), which is a set of unit vectors (i.e., with length $= 1$) where each of these are in the direction of a best-of-fit line given the data that produced said component, all while being orthogonal to the vectors before them. This generates an orthonormal basis, where different individual dimensions of the data are linearly uncorrelated.

Thus, for each p variable in the analysis, a unit vector given the aforementioned principles is calculated and then added to a set which will be the PC. The process will generate as many PCs as there are variables to explain the entirety of variance that a dataset contains.

Each PC, which share a common coordinating system thanks to a change of basis process, will be the weighted linear combination of variables used to produce them. They will all be uncorrelated and orthogonal to each other (Dutt 2021).

Usually, PCs are produced in an order of most variance explained. The first PC will explain the most variance in the dataset, the second one will be the most explanatory once variance assessed by the first is excluded, and so on until the p-th PC.

Doing PCA starts with mean-centering, where variable averages are computed and then subtracted from the data so the coordinating system for each n-observation and p-variable shares a common origin point at zero. In practice, this is the standardisation of each analysed variable so issues with differences in range between them does not lead to biased results in the analysis. With standardisation, these variables are scaled to a form where they can be compared between one another.

The formula for standardisation is quite simple, where:

$$z = \frac{\text{value - mean}}{\text{standard deviation}}$$

A covariance matrix is then drawn, just so we can check the variance of each variable in PCA and calculate eigenvectors and eigenvalues to identify principal components. Eigenvectors will indicate the direction in which most variance is found after a linear transformation in the variable(s) vector(s), becoming principal components per se. Whereas eigenvalues are scalar factors, or coefficients, that "stretches" eigenvectors the most – in PCA, they indicate how much variance each eigenvector, or PC, carry.

Therefore, eigenvectors in PCA point to where most information can be found in the p-dimensional system of variables. Meanwhile eigenvalues will help us rank PCs by order of variance, just by putting them on a decreasing scale.

Finally, unit vectors will be found by normalizing the orthogonal eigenvectors. Via diagonal transformation of the covariance matrix, we can find "scores" which will give out the importance of each PC drawn in the PCA process. These "scores" can be seen as "feature vectors", or "summary indexes", that will be used to finalise the process either with dimensionality reduction, singular value decomposition, score plots, indexes, and/or many other forms of interpretation (Jaadi 2021).

Mathematically, there is a A matrix with $n \times p$ dimensions, where n are observations and p, variables of a data set. This matrix has been standardised so each column's sample mean is shifted to zero.

For PCA, the A matrix undergoes an orthogonal linear transformation where a set m of p-dimensional vector weights, $w_{(k)} = (w_1, \ldots, w_p)_{(k)}$, will map each row vector $a_{(i)}$ from A to create principal component scores, $s_{(i)} = (s_1, \ldots, s_l)_{(i)}$. This is done via

$$s_{k(i)} = a_{(i)} \cdot w_{(k)}$$

Where $i = 1, \ldots, n$ and $k = 1, \ldots, m$. This way, each score in $s_{k(i)}$ will take maximum variance from the A matrix, with each $w_{(k)}$ weight being a unit vector – or eigenvector.

Drawing PCs is conditional on weight vectors. For the first PC, which carries the most variance from the A matrix, the condition that needs to be fulfilled is

$$w_{(1)} = \arg\max_{\|w\|=1} \left\{ \sum_i (t_1)^2_{(i)} \right\} = \arg\max_{\|w\|=1} \left\{ \sum_i \left(a_{(i)} \cdot w \right)^2 \right\}$$

In matrix form, this is equivalent to

$$w_{(1)} = \arg\max_{\|w\|=1} \left\{ \|Aw\|^2 \right\} = \arg\max_{\|w\|=1} \left\{ w^\top A^\top A w \right\}$$

It can also be drawn as

$$w_{(1)} = \arg\max \left\{ \frac{w^\top A^\top A w}{w^\top w} \right\}$$

Since $w_{(k)}$ are unit vectors. Therefore,

$$s_{1(i)} = a_{(i)} \cdot w_{(1)}$$

Can be seen as a score for the first principal component. This can also be drawn as a vector in the original variables, in the form $a_{(i)} \cdot w_{(1)} w_{(1)}$.

As for the other PCs, they are found when you subtract the first $k-1$ PCs from A. This means

$$\hat{A}_k = A - \sum_{v=1}^{k-1} A w_{(v)} w_{(v)}^\top$$

And

$$w_{(k)} = \arg\max_{\|w\|=1} \left\{ \left\| \hat{A}_k w \right\|^2 \right\} = \arg\max \left\{ \frac{w^\top \hat{A}_k^\top \hat{A}_k w}{w^T w} \right\}$$

To find the respective weight/unit vectors for each PC from A. Therefore,

$$s_{k(i)} = a_{(i)} \cdot w_{(k)}$$

Are scores for each PC found via linear transformation of A. And its full principal components' decomposition can be summed up as

$$S = AW$$

With W as a matrix of weights whose columns are not just eigenvectors for the PCs, but that can also be utilised for "loadings" in PCA and factor analysis (Jolliffe 2002).

This entire process helps with addressing dimensionality issues. While it is quite useful to have millions of observations and dozens of variables for the task at hand, computation and even interpretation of such data can become problems if it is assessed under normal circumstances and with a limited amount of time. Methods such as PCA help with addressing the "curse of dimensionality" that comes from handling high-dimensional spaces in big data (Sohil, Sohali, and Shabbir 2022).

References

Abdala, V. 2021. "Taxa de informalidade no mercado de trabalho sobe para 40%, diz IBGE." *Agência Brasil*. https://agenciabrasil.ebc.com.br/economia/noticia/2021-07/taxa-de-informalidade-no-mercado-de-trabalho-sobe-para-40-diz-ibge.

ADVFN. 2018. "A Evolução Do Salário Mínimo No Brasil." Document. *ADVFN*. https://br.advfn.com/indicadores/salario-minimo/evolucao.

Agência O Globo. 2017. "A história do Rio: do esplendor à crise." *Época Negócios*. https://epocanegocios.globo.com/Economia/noticia/2017/09/historia-do-rio-do-esplendor-crise.html.

Ahlfeldt, G. M., D. Roth, and T. Seidel. 2018. "The Regional Effects of Germany's National Minimum Wage." *Economics Letters* 172 (November): 127–30. https://doi.org/10.1016/j.econlet.2018.08.032.

Amaral, P. V., M. Lemos, R. Simões, and F. Chein. 2010. "Regional Imbalances and Market Potential in Brazil." *Spatial Economic Analysis* 5 (4): 463–82. https://doi.org/10.1080/17421772.2010.516441.

Amri, K. 2018. "The Macroeconomic Impact of Regional Minimum Wages: A Cross-Province Data Evidence from Indonesia." *Regional Science Inquiry* X (3): 163–76.

Arestis, P., C. T. Baltar, and D. M. Prates. 2016. "Brazilian Economic Performance Since the Emergence of the Great Recession: The Effects of Income Distribution on Consumption." *Panoeconomicus* 63 (2): 157–74. https://doi.org/10.2298/PAN1602157A.

Askenazy, P. 2003. "Minimum Wage, Exports and Growth." *European Economic Review* 47 (1): 147–64. https://doi.org/10.1016/S0014-2921(02)00187-3.

Blackman, A. W., E. J. Seligman, and G. C. Sogliero. 1973. "An Innovation Index Based on Factor Analysis." *Technological Forecasting and Social Change* 4 (3): 301–16. https://doi.org/10.1016/0040-1625(73)90060-7.

Boadway, R., and K. Cuff. 2001. "A Minimum Wage Can Be Welfare-Improving and Employment-Enhancing." *European Economic Review* 45 (3): 553–76. https://doi.org/10.1016/S0014-2921(00)00066-0.

Boeri, T., P. Garibaldi, and M. Ribeiro. 2011. "The Lighthouse Effect and Beyond." *Review of Income and Wealth* 57 (s1): S54–78. https://doi.org/10.1111/j.1475-4991.2011.00455.x.

Borusyak, K., X. Jaravel, and J. Spiess. 2022. "Revisiting Event Study Designs: Robust and Efficient Estimation." arXiv. https://doi.org/10.48550/arXiv.2108.12419.

Braunstein, E., and S. Seguino. 2018. "The Impact of Economic Policy and Structural Change on Gender Employment Inequality in Latin America, 1990–2010." *Review of Keynesian Economics* 6 (3): 307–32. https://doi.org/10.4337/roke.2018.03.02.

Brinca, P., and J. Costa-Filho. 2021. "Economic Depression in Brazil: The 2014-2016 Fall." {{MPRA Paper}}. https://mpra.ub.uni-muenchen.de/107298/.

Brito, A., M. Foguel, and C. Kerstenetzky. 2017. "The Contribution of Minimum Wage Valorization Policy to the Decline in Household Income Inequality in Brazil: A Decomposition Approach." *Journal of Post Keynesian Economics* 40 (4): 540–75. https://doi.org/10.1080/01603477.2017.1333436.

Butts, K. 2021. "Did2s: Two-Stage Difference-in-Differences Following Gardner (2021)." https://github.com/kylebutts/did2s/.

Butts, K., and J. Gardner. 2022. "{Did2s}: Two-Stage Difference-in-Differences." arXiv. https://arxiv.org/abs/2109.05913.

Buuren, S. van. 2018. *Flexible Imputation of Missing Data*.

Buuren, S. van, and K. Groothuis-Oudshoorn. 2011. "Mice: Multivariate Imputation by Chained Equations in R." *Journal of Statistical Software* 45 (December): 1–67. https://doi.org/10.18637/jss.v045.i03.

Cahuc, P., and P. Michel. 1996. "Minimum Wage Unemployment and Growth." *European Economic Review* 40 (7): 1463–82. https://doi.org/10.1016/0014-2921(95)00035-6.

Callaway, B., A. Goodman-Bacon, and P. H. C. Sant'Anna. 2021. "Difference-in-Differences with a Continuous Treatment." *arXiv:2107.02637 [Econ]*, July. https://arxiv.org/abs/2107.02637.

Callaway, B., T. Li, and T. Oka. 2018. "Quantile Treatment Effects in Difference in Differences Models Under Dependence Restrictions and with Only Two Time Periods." *Journal of Econometrics*, Special issue on Advances in Econometric Theory: Essays in honor of Takeshi Amemiya, 206 (2): 395–413. https://doi.org/10.1016/j.jeconom.2018.06.008.

Callaway, B., and P. H. C. Sant'Anna. 2022. "Introduction to DiD with Multiple Time Periods." https://bcallaway11.github.io/did/articles/multi-period-did.html.

Card, D., and A. B. Krueger. 1994. "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania." *The American Economic Review* 84 (4): 772–93.

Carino, J. 2019. "Qual a origem da palavra carioca?" https://immub.org/noticias/qual-a-origem-da-palavra-carioca.

Casa Fluminense. 2020. "Mapa da Desigualdade." *Casa Fluminense*.

Chateau, D., C. Metge, H. Prior, and R.-A. Soodeen. 2012. "Learning From the Census: The Socio-economic Factor Index (SEFI) and Health Outcomes in Manitoba." *Canadian Journal of Public Health / Revue Canadienne de Sante'e Publique* 103: S23–27.

Chimed-Ochir, O., T. Ikaga, S. Ando, T. Ishimaru, T. Kubo, S. Murakami, and Y. Fujino. 2021. "Effect of Housing Condition on Quality of Life." *Indoor Air* 31 (4): 1029–37. https://doi.org/10.1111/ina.12819.

Christl, M., M. Köppl-Turyna, and D. Kucsera. 2019. "Employment Effects of Minimum Wages." *Ifo DICE Report* 16 (4): 1–8.

Columbia Mailman School of Public Health. 2013. "Difference-in-Difference Estimation." https://www.publichealth.columbia.edu/research/population-health-methods/difference-difference-estimation.

Crone, T. M., and A. Clayton-Matthews. 2005. "Consistent Economic Indexes for the 50 States." *The Review of Economics and Statistics* 87 (4): 593–603.

Cunningham, S. 2021. *Causal Inference: The Mixtape*.

Dahis, R. 2022. "Base Dos Dados." *Base Dos Dados*. https://basedosdados.org/.

Dai, P.-F., X. Xiong, and W.-X. Zhou. 2021. "A Global Economic Policy Uncertainty Index from Principal Component Analysis." *Finance Research Letters* 40 (May): 101686. https://doi.org/10.1016/j.frl.2020.101686.

de Chaisemartin, C., and X. D'Haultfoeuille. 2020. "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects," 51.

Derenoncourt, E., F. Gérard, L. Lagos, and C. Montialoux. 2021. "Racial Inequality, Minimum Wage Spillovers, and the Informal Sector." *Working Paper*, 75.

Derenoncourt, E., and C. Montialoux. 2021. "Minimum Wages and Racial Inequality." *The Quarterly Journal of Economics* 136 (1): 169–228. https://doi.org/10.1093/qje/qjaa031.

Dong, Y., and C.-Y. J. Peng. 2013. "Principled Missing Data Methods for Researchers." *SpringerPlus* 2 (May): 222. https://doi.org/10.1186/2193-1801-2-222.

Du, P., and S. Wang. 2020. "The Effect of Minimum Wage on Firm Markup: Evidence from China." *Economic Modelling* 86 (March): 241–50. https://doi.org/10.1016/j.econmod.2019.10.012.

Dube, A. 2019a. *Impacts of Minimum Wages: Review of the International Evidence.*

———. 2019b. "Minimum Wages and the Distribution of Family Incomes." *American Economic Journal: Applied Economics* 11 (4): 268–304. https://doi.org/10.1257/app.20170085.

Dube, A., T. W. Lester, and M. Reich. 2010. "Minimum Wage Effects Across State Borders: Estimates Using Contiguous Counties." *The Review of Economics and Statistics* 92 (4): 945–64. https://doi.org/10.1162/REST_a_00039.

Dutt, A. 2021. "A Step By Step Implementation of Principal Component Analysis." *Medium*. https://towardsdatascience.com/a-step-by-step-implementation-of-principal-component-analysis-5520cc6cd598.

Engbom, N., and C. Moser. 2021. "Earnings Inequality and the Minimum Wage: Evidence from Brazil." Working {{Paper}} 28831. Working Paper Series. National Bureau of Economic Research. https://doi.org/10.3386/w28831.

European Central Bank. 2016. "What Is Driving Brazil's Economic Downturn?" no. 1: 3.

Fanti, L., and L. Gori. 2011. "On Economic Growth and Minimum Wages." *Journal of Economics* 103 (1): 59–82. https://doi.org/10.1007/s00712-011-0190-3.

Franklin, S. B., D. J. Gibson, P. A. Robertson, J. T. Pohlmann, and J. S. Fralish. 1995. "Parallel Analysis: A Method for Determining Significant Principal Components." *Journal of Vegetation Science* 6 (1): 99–106. https://doi.org/10.2307/3236261.

Gaglioni, C. 2022. "O que você precisa saber sobre o Censo 2022." *Nexo Jornal*. https://www.nexojornal.com.br/expresso/2022/08/01/O-que-voc%C3%AA-precisa-saber-sobre-o-Censo-2022.

Gardner, J. 2021. "Two-Stage Differences in Differences," 34.

Gerard, F., L. Lagos, E. Severnini, and D. Card. 2021. "Assortative Matching or Exclusionary Hiring? The Impact of Employment and Pay Policies on Racial Wage Differences in Brazil." *American Economic Review* 111 (10): 3418–57. https://doi.org/10.1257/aer.20181596.

Godoey, A., and M. Reich. 2021. "Are Minimum Wage Effects Greater in Low-Wage Areas?" *Industrial Relations: A Journal of Economy and Society* 60 (1): 36–83. https://doi.org/10.1111/irel.12267.

Godshalk, E. B., and D. H. Timothy. 1988. "Factor and Principal Component Analyses as Alternatives to Index Selection." *Theoretical and Applied Genetics* 76 (3): 352–60. https://doi.org/10.1007/BF00265334.

Goodman-Bacon, A. 2021. "Difference-in-Differences with Variation in Treatment Timing." *Journal of Econometrics*, Themed Issue: Treatment Effect 1, 225 (2): 254–77. https://doi.org/10.1016/j.jeconom.2021.03.014.

Gorsuch, R. L. 1973. "Using Bartlett's Significance Test to Determine the Number of Factors to Extract." *Educational and Psychological Measurement* 33 (2): 361–64. https://doi.org/10.1177/001316447303300216.

Governo Federal do Brasil. 2022. "Optar Pelo Simples Nacional." https://www.gov.br/pt-br/servicos/optar-pelo-simples-nacional.

Guimarães, C. A. 2022. "Em 150 anos, conheça a história que o Censo conta." *Agência de Notícias - IBGE*. https://agenciadenoticias.ibge.gov.br/agencia-noticias/2012-agencia-de-noticias/noticias/33495-em-150-anos-conheca-a-historia-que-o-censo-conta.

Hacker, S. R., J. Klaesson, L. Pettersson, and P. Sjölander. 2013. "Regional Economic Concentration and Growth." In *Metropolitan Regions: Knowledge Infrastructures of the Global Economy*, edited by Johan Klaesson, Börje Johansson, and Charlie Karlsson, 117–39. Advances in Spatial Science. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-32141-2_6.

Hapfelmeier, A., and K. Ulm. 2014. "Variable Selection by Random Forests Using Data with Missing Values." *Computational Statistics & Data Analysis* 80 (December): 129–39. https://doi.org/10.1016/j.csda.2014.06.017.

Hermann, B. M. 2014. "Formação, declínio e inflexão: reflexões sobre a centralidade da cidade do Rio de Janeiro," February.

Hinojosa, P. G. 2019. "The Role of the Minimum Wage on the Declining Wage Inequality in Latin America: Evidence from Brazil." *Review of Economic Analysis* 11 (1): 39–82. https://doi.org/10.15353/rea.v11i1.1517.

Ho, T. K. 1995. "Random Decision Forests." In *Proceedings of 3rd International Conference on Document Analysis and Recognition*, 1:278–282 vol.1. https://doi.org/10.1109/ICDAR.1995.598994.

"Indice de Progresso Social Do Rio de Janeiro." 2020. http://ipsrio.com.br/metodologia.
Instituto Brasileiro de Geografia e Estatística. 2013. "Censos Demográficos - Memória IBGE." https://memoria.ibge.gov.br/historia-do-ibge/historico-dos-censos/censos-demograficos.html.
———. 2019. "Produto Interno Bruto Dos Municípios - IBGE [Data File]."
———. 2022a. "Estimativas Da População Residente Para Os Municípios e Para as Unidades Da Federação." https://www.ibge.gov.br/estatisticas/sociais/populacao/9103-estimativas-de-populacao.html?=&t=resultados.
———. 2022b. "Relação Anual de Informações Sociais – RAIS." https://ces.ibge.gov.br/base-de-dados/metadados/mte/relacao-anual-de-informacoes-sociais-rais.
———. 2022c. "PNAD Contínua Trimestral: desocupação recua em 15 das 27 UFs no 4° trimestre de 2021." *Agência de Notícias - IBGE*. https://agenciadenoticias.ibge.gov.br/agencia-sala-de-imprensa/2013-agencia-de-noticias/releases/33033-pnad-continua-trimestral-desocupacao-recua-em-15-das-27-ufs-no-4-trimestre-de-2021.
Instituto de Pesquisa Econômica Aplicada. 2022. "Salário Mínimo Real." Document. *Ipeadata*. http://www.ipeadata.gov.br/ExibeSerie.aspx?serid=37667.
Instituto Mauro Borges. 2018. "Produto Interno Bruto do Brasil a preços de mercado, Brasil, segundo Grandes Regiões e Unidades da Federação - IMB." *Instituto Mauro Borges de Estatísticas e Estudos Socioeconômicos*. https://www.imb.go.gov.br/index.php?option=com_content&view=article&id=1170:produto-interno-bruto-do-brasil-a-pre%C3%A7os-de-mercado,-brasil,-segundo-grandes-regi%C3%B5es-e-unidades-da-federa%C3%A7%C3%A3o&catid=218&Itemid=101.
Instituto Municipal de Urbanismo Pereira Passos. 2019. "Regiões de Planejamento - Indicadores 2018." https://www.data.rio/datasets/PCRJ::regi%C3%B5es-de-planejamento-indicadores-2018/explore.
———. 2022. "Instituto Municipal de Urbanismo Pereira Passos - História." https://www.rio.rj.gov.br/web/ipp/historia.
International Labour Organization. 2015. "A Short History of Minimum Wage Laws." Document. *ILO - International Labour Organization*. http://www.ilo.org/global/topics/wages/minimum-wages/definition/WCMS_439071/lang–en/index.htm.
Jaadi, Z. 2021. "A Step-by-Step Explanation of Principal Component Analysis (PCA) - Built In." https://builtin.com/data-science/step-step-explanation-principal-component-analysis.
Jales, H. 2018. "Estimating the Effects of the Minimum Wage in a Developing Country: A Density Discontinuity Design Approach." *Journal of Applied Econometrics* 33 (1): 29–51. https://doi.org/10.1002/jae.2586.
Jolliffe, I. T. 2002. *Principal Component Analysis*. Springer Series in Statistics. New York: Springer-Verlag. https://doi.org/10.1007/b98835.
Jung, Y. C., A. McFarlane, and A. Das. 2021. "The Effect of Minimum Wages on Consumption in Canada." *The Economic and Labour Relations Review* 32 (1): 65–89. https://doi.org/10.1177/1035304620949950.
Kaiser, H. F. 1960. "The Application of Electronic Computers to Factor Analysis." *Educational and Psychological Measurement* 20 (1): 141–51. https://doi.org/10.1177/001316446002000116.
———. 1970. "A Second Generation Little Jiffy." *Psychometrika* 35 (4): 401–15. https://doi.org/10.1007/BF02291817.
Katovich, E. S., and A. G. Maia. 2018-Jan-Apr. "The Relation Between Labor Productivity and Wages in Brazil:" *Nova Economia* 28 (2018-Jan-Apr): 7–38. https://doi.org/10.1590/0103-6351/3943.
Komatsu, B., and N. M. Filho. 2016. "Does the Rise of the Minimum Wage Explain the Fall of Wage Inequality in Brazil?" *INSPER Working Paper*, 30.
Lai, D. 2003. "Principal Component Analysis on Human Development Indicators of China." *Social Indicators Research* 61 (3): 319–30. https://doi.org/10.1023/A:1021951302937.
Lang, K., and S. Kahn. 1998. "The Effect of Minimum-Wage Laws on the Distribution of Employment:

Theory and Evidence." *Journal of Public Economics* 69 (1): 67–82. https://doi.org/10.1016/S0047-2727(97)00085-6.

Lemos, S. 2004. "The Effect of the Minimum Wage on Prices in Brazil." *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.515482.

Madley-Dowd, P., R. Hughes, K. Tilling, and J. Heron. 2019. "The Proportion of Missing Data Should Not Be Used to Guide Decisions on Multiple Imputation." *Journal of Clinical Epidemiology* 110 (June): 63–73. https://doi.org/10.1016/j.jclinepi.2019.02.016.

Manning, A. 2021. "The Elusive Employment Effect of the Minimum Wage." *Journal of Economic Perspectives* 35 (1): 3–26. https://doi.org/10.1257/jep.35.1.3.

Marques, C. R., and P. D. Neves. 2001. "USING THE FIRST PRINCIPAL COMPONENT AS A CORE INFLATION INDICATOR." *Economic Bulletin*, 7.

McKenzie, D. 2020. "Revisiting the Difference-in-Differences Parallel Trends Assumption: Part I Pre-Trend Testing." https://blogs.worldbank.org/impactevaluations/revisiting-difference-differences-parallel-trends-assumption-part-i-pre-trend.

Meghir, C., R. Narita, and J.-M. Robin. 2015. "Wages and Informality in Developing Countries." *The American Economic Review* 105 (4): 1509–46.

Meireles, F. 2018. "deflateBR: Deflate Nominal Brazilian Reais." https://CRAN.R-project.org/package=deflateBR.

Montelpare, W. J., E. Read, T. McComber, A. Mahar, and K. Ritchie. 2020. "Working with Missing Data." *Applied Statistics in Healthcare Research*.

Morello, T., and J. Anjolim. 2021. "Gender Wage Discrimination in Brazil from 1996 to 2015: A Matching Analysis." *EconomiA* 22 (2): 114–28. https://doi.org/10.1016/j.econ.2021.03.002.

Morris, T. P., I. R. White, and P. Royston. 2014. "Tuning Multiple Imputation by Predictive Mean Matching and Local Residual Draws." *BMC Medical Research Methodology* 14 (1): 75. https://doi.org/10.1186/1471-2288-14-75.

Navarro, T. M., F. F. M. Durán, and J. L. S. Bartolomé. 2017. "Un índice de competitividad regional para España." *Revista de estudios regionales*, no. 109: 67–94.

New Zealand Government. 2022. "Benefits of Using Open Data." https://www.data.govt.nz/toolkit/open-data/benefits-of-using-open-data/.

Nooraee, N., G. Molenberghs, J. Ormel, and E. R. Van den Heuvel. 2018. "Strategies for Handling Missing Data in Longitudinal Studies with Questionnaires." *Journal of Statistical Computation and Simulation* 88 (17): 3415–36. https://doi.org/10.1080/00949655.2018.1520854.

Oreiro, J. L. 2017-Jan-Apr. "A grande recessão brasileira: diagnóstico e uma agenda de política econômica." *Estudos Avançados* 31 (2017-Jan-Apr): 75–88. https://doi.org/10.1590/s0103-40142017.31890009.

Organisation for Economic Co-operation and Development. 2021. "Real Minimum Wages [Data File]."

Pamplona, N., and E. Cucolo. 2019. "No Nordeste, quase 25% da renda vem de aposentadorias e programas sociais." *Folha de S.Paulo*. https://www1.folha.uol.com.br/mercado/2019/10/no-nordeste-quase-25-da-renda-vem-de-aposentadorias-e-programas-sociais.shtml.

Paun, C. V., R. Nechita, A. Patruti, and M. V. Topan. 2021. "The Impact of the Minimum Wage on Employment: An Eu Panel Data Analysis." *Sustainability (Switzerland)* 13 (16). https://doi.org/10.3390/su13169359.

Prefeitura do Rio de Janeiro. 1981. "LEI Nº 282 DE 9 DE NOVEMBRO DE 1981." http://mail.camara.rj.gov.br/APL/Legislativos/contlei.nsf/2ed241833abd7a5b8325787100687ecc/497bd8f0d1c231a9032576ac

———. 2011. "Lei Complementar nº 111 de 01/02/2011 - Municipal - Rio de Janeiro." https://www.legisweb.com.br/legislacao/?id=178079.

Rebitzer, J. B., and L. J. Taylor. 1995. "The Consequences of Minimum Wage Laws Some New Theoretical

Ideas." *Journal of Public Economics* 56 (2): 245–55. https://doi.org/10.1016/0047-2727(93)01411-3.

Receita Federal. 2007. "Simples Nacional." https://www8.receita.fazenda.gov.br/SimplesNacional/Documentos/Pagina.aspx?i

Silva, M. O. da, and M. H. Versiani. 2015. "História de Capitalidade do Rio De Janeiro." *Cadernos do Desenvolvimento Fluminense*, no. 7: 75–90. https://doi.org/10.12957/cdf.2015.19711.

Sistema Único de Saude. 2011. "Índice de Gini Da Renda Domiciliar Per Capita - Rio de Janeiro." http://tabnet.datasus.gov.br/cgi/ibge/censo/cnv/ginirj.def.

Sohil, F., M. U. Sohali, and J. Shabbir. 2022. "An Introduction to Statistical Learning with Applications in R: By Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani, New York, Springer Science and Business Media, 2013, $41.98, eISBN: 978-1-4614-7137-7." *Statistical Theory and Related Fields* 6 (1): 87–87. https://doi.org/10.1080/24754269.2021.1980261.

Spanache, I. 2020. "Data Science for Social Good: Best Sources for Free Open Data." *Medium*. https://towardsdatascience.com/data-science-for-social-good-best-sources-for-free-open-data-5120070caf02.

Stock, J. H., and M. W. Watson. 1999. "Forecasting Inflation." *Journal of Monetary Economics* 44 (2): 293–335. https://doi.org/10.1016/S0304-3932(99)00027-6.

Sun, L., and S. Abraham. 2021. "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects." *Journal of Econometrics*, Themed Issue: Treatment Effect 1, 225 (2): 175–99. https://doi.org/10.1016/j.jeconom.2020.09.006.

The World Bank. 2021. "GDP (Constant 2015 US$) - Brazil [Data File]."

The World Bank, Development Research Group. 2021. "Gini Index (World Bank Estimate) - Brazil [Data File]."

Vartanian, P. R., and H. de S. Garbe. 2018. "The Brazilian Economic Crisis During the Period 2014-2016: Is There Precedence of Internal or External Factors?" 21.

Vidal, P., L. P. Sierra, J. Sanabria, and J. A. Collazos. 2017. "A Monthly Regional Indicator of Economic Activity: An Application for Latin America." *Latin American Research Review* 52 (4): 589–605. https://doi.org/10.25222/larr.38.

Villela, D. 2013. "Nordeste cresce acima da média do País." *O Estado de São Paulo*. https://economia.estadao.com.br/noticias/geral,nordeste-cresce-acima-da-media-do-pais,165312e.

Vyas, S., and L. Kumaranayake. 2006. "Constructing Socio-Economic Status Indices: How to Use Principal Components Analysis." *Health Policy and Planning* 21 (6): 459–68. https://doi.org/10.1093/heapol/czl029.

Watkins, M. W. 2018. "Exploratory Factor Analysis: A Guide to Best Practice." *Journal of Black Psychology* 44 (3): 219–46. https://doi.org/10.1177/0095798418771807.

Wickham, H. 2014. "Tidy Data." *Journal of Statistical Software* 59 (September): 1–23. https://doi.org/10.18637/jss.v059.i10.

Young, J. 2017. "Imputation for Random Forests." *All Graduate Plan B and Other Reports*, August. https://doi.org/10.26076/d5e1-3083.

Zainal, N. R., G. Kaur, N. 'Aisah. Ahmad, and J. Mhd. Khalili. 2012. "Housing Conditions and Quality of Life of the Urban Poor in Malaysia." *Procedia - Social and Behavioral Sciences*, AcE-Bs 2012 Bangkok, Sukosol Hotel, Bangkok, Thailand, 16-18 July, 2012, 50 (January): 827–38. https://doi.org/10.1016/j.sbspro.2012.08.085.

Annex

Annex A – Literature review support tables

Figure 15: Minimum wage and employment

	Main objective(s)	Methodology	Dependent variable(s) (y)	Explanatory variable(s) (x)		Estimated effect(s)	Conclusion(s)
Ahlfeldt et al. (2018)	To assess the regional effects of the introduction of a minimum wage in Germany	Difference-in-differences (DiD)	Hourly wages	Minimum wage "bite"	Share of workers earning less than minimum wage, before minimum wage introduction, in a region	(+)(insignificant)	Introducing a minimum wage had no employment effects, all while raising low-wage workers' earnings, thanks to spatial wage convergence
			Unemployment rate			(-)(insignificant)	
Card & Krueger (1994)	To verify the effect of minimum wage increase on low-wage jobs, comparing employment effects of said increase in New Jersey and Pennsylvania stores (with the former increasing minimum wages, and the former not doing so)	Difference-in-differences (DiD)	Change in employment in a store from time-period 1 to time-period 2	GAP	Measure of the impact of minimum wage (increase) at a store, based on the initial wage on said store - or (proportional) increase in wages necessary to bring wages at a store to the new minimum wage (in New Jersey)	(+)	Contrary to general theory of employment (based on classical labour market model), there is an increase - and not a decrease - in employment from growth in minimum wage
Christl et al. (2019)	To test whether minimum wage effects on employment are non-linear	Ordinary least squares (OLS)	Youth employment rate (for people aged 15-24)	Real minimum wage	In annual terms	non-linear - inverted U-shape	Low-level minimum wages estimulate the "acceptance rate" of job offers (and employment), whereas high-level minimum wage reduce labour demand and thus employment opportunities
				Kaitz index	Relationship between levels of minimum and average wage	non-linear - inverted U-shape	
		Instrumental variables (IV)		Minimum wage, interated with oil price and share of left-wing parties in countries' parliament		non-linear - inverted U-shape	
Dube (2019)	To review international evidence on impacts of minimum wage	Literature review, using estimated elasticities of various studies	Own-wage employment elasticity	Minimum wage employment elasticity	Percentage change in employment, divided by percentage change in minimum wage	(-)	There is a "muted effect" of minimum wage on employment, with little to no decrease on it, all while significantly increasing earnings of low-paid workers
				Average wage elasticity	Percentage change in average wage, divided by percentage change in minimum wage	(+)	
Dube et al. (2010)	To identify effects of minimum wages on both earnings and employment of workers in low-wage sectors, along state borders in the United States	Panel data, with fixed effects estimation	Log of earnings	Minimum wage	At county level, in log values	(+)(-)	By using local identification instead of a generalised, regional/state-wide analysis, minimum wage effects are positive on earnings and insignificant on employment
			Log of employment	Minimum wage differences between county-pairs		(+)(insignificant)	
Meghir et al (2015)	To model and test effects of higher formal employment (and minimum wage) enforcement on labour markets with several informal workers	Stationary equilibrium wage-posting search model, with homogenous workers and heterogenous firms	Formal and informal employment	Cost of informality	Proxy of higher regulation on larger firms who take in informal workers	(+)(+)	In a market with search frictions, increasing regulation of informal workers does not decrease employment while improving workers' welfare (with higher wages, labour rights etc)
			Formal and informal wages				
Paun et al. (2021)	To check the effect of the minimum wage on employment, with focus on vulnerable groups	Panel data, with random and fixed effects estimation	Yearly change for total, youth, elder and female employments	Minimum wage dynamic	In log values	(-)	Increasing minimum wage negatively affects employment, especially vulnerable groups, in the assessed groups (mostly developed nations)
				Yearly change in labour productivity		(-)	
				Yearly change in total labour		(+)	

Source: Own elaboration.

Figure 16: Minimum wage heterogeneities

	Main objective(s)	Methodology	Dependent variable(s) (y)	Explanatory variable(s) (x)		Estimated effect(s)	Conclusion(s)
Braunstein & Seguino (2018)	To assess what motivated sharp declines in gender employment and inequality in Latin American countries during the late 20th/early 21st centuries	Panel data, with country-level fixed effects	Female and male employment/unemployment, rates, although with focus on the female perspective	Social and employment policies	. Social public expenditures, as share of GDP . Minimum wage, relative to average wage (Kaitz index)	(+)(insignificant) (+)(-)	Social spending, higher minimum wages and public investment provided positive effects on employment in Latin America, especially for women
		Two-stage least squares (2SLS)		Macroeconomic policies	. Real effective exchange rate . Real interest rate . Public investment as share of GDP	(+)(insignificant) (insignificant)(-) (+)(-)	
				Measures of economic structure	. Ratio of manufacturing exports to imports . Fuel and ores, as share of merchandised exports	(insignificant)(+) (+)(insignificant)	
Callaway et al. (2018)	To estimate the effect of minimum wage increases on earnings distribution for several subgroups	Difference-in-differences (DiD), with conditional quantile treatment effect on the treated (CQTT)	Earnings	Minimum wage	Analysed by quantile (of income), subdivided by gender, race, education	(-)	Minimum wage effect is negative, and concentrated in the lower part of income distribution for white female college graduates, and non-white male and female non-college graduates
Derenoncourt & Montialoux (2021)	To check the effects of expanding minimum wage coverage (in 1966) in the United States across industry and racial working groups	Difference-in-differences (DiD)	Log annual earnings of a worker	Covered 1967	Dummy term interacted with another dummy term for years before and after reform (1966)	(+)(insignificant)(-)	Expansion of minimum wage laws in the United States increased earnings for low-wage workers, and reduced racial gap earnings, while not having a significant impact on employment for any of the groups affected
					Term also analysed for different race groups (blacks and whites)	(+), more for blacks than for whites (insignificant) (N/A)	
			Log annual hours worked, employment probability Racial gap in earnings	Strongly treated states	States which had no minimum wage law prior to 1966	(+)(insignificant)(N/A)	
Dube (2019)	To estimate minimum wage increases and its effect on family income distribution	Two-way fixed effects	Equivalised real family income (i.e ration of family income to federal povery threshold, depending on family size, number of children, year)	Minimum wage	By state of residence	(+)(-)	Higher minimum wages increases incomes at the bottom of the family income distribution, reducing poverty rate, especially for vulnerable groups (i.e. youth (under 30), non-high school graduates, blacks and Latinos) in the United States
		Quantile partial effects	Dummy indicating whether individual is in a family at or below the poverty rate				
Dube et al. (2007)	To study how a citywide minimum wage affected outcomes for low-wage employees	Difference-in-differences (DiD)	Average hourly pay Distribution of pay (to capture wage inequality) Total employment Part-time/full-time employment	Treatment intensity	Proportion of workers earning under $8.50, if the restaurant is covered by (SF minimum wage) law	(+)(-)(insignificant)(insignificant)	The implementation of a citywide minimum wage increased pay and reduced wage inequality, without negative effects on employment
Godoey & Reich (2021)	To assess the effects of minimum wage increases, while comparing high- and low-income areas at the county level	Difference-in-differences (DiD)	Employment Earnings Poverty rate	Minimum wage	i.e. the change in log minimum wage over an event window, using event studies	(insignificant)(+)(-)	Positive minimum wage increase effects are higher in low-income areas, with reductions in household and child poverty while not having any adverse effects on employment

Source: Own elaboration.

Figure 17: Minimum wage effects in Brazil

	Main objective(s)	Methodology	Dependent variable (y)	Explanatory variable(s) (x)		Estimated effect(s)	Conclusion(s)
Boeri et al. (2011)	To show how official minimum wage increases also affect earnings in informal labour relationships, via sorting and composition effects	Fixed effects estimation, using panel data	Hourly wages	(Employment) tenure	i.e. whether worker is in formal or informal sector, plus length of employment	(+)(-)(+)	Minimum wage increases lead to higher earnings for informal workers, while also shifting low-wage workers from informal to formal labour relationships, and increasing educational attainment, thanks to sorting effects. However, there is at least a temporary decrease in employment
		Difference-in-differences (DiD)	Employment	Education		(+)(-)(+)	
			Educational attainment				
Brito et al. (2017)	To gauge the effect of minimum wage increases in overall household inequality	Decomposition method	Gini index	Minimum wage	Increases from 1995 to 2014	(-)	Over 64% of Brazil's 14.6 reduction in inequaltiy obseved between 1995 and 2014 was due to minimum wage increases, confirmed by analysis at regional and formal/informal labour relationships, especially via pensions
Derenoncourt et al. (2021)	To study how the minimum wage affected Brazilian racial inequalities in earnings	Quantile regression	Earnings, at different quantiles	White	Whether a worker is white or non-white	(-)(insignificant)(insignificant)	Minimum wage increases greatly diminished racial wage gaps, with effects concentrated at the lower deciles of earnings, without reallocation of workers from formal to informal sectors nor disemployment effects for non-white workers
		Difference-in-differences (DiD)	Employment level	Minimum wage	Increases from 2000 to 2009	(+)(insignificant)(insignificant)	
			Employment relationship	Treated states	Whether a state is "strongly" treated (i.e. if there are a lot of low-wage workers in the region)	(+)(insignificant)(insignificant)	
Engbom & Moser (2021)	To show the contribution of minimum wage hikes in reducing inequality in Brazil	Two-way fixed effects	Income (as mean wages in multiples of the minimum wage)	Time-varying worker characteristics	Education, work hours, type of occupation	(+)(-)(insignificant)	Minimum wage increases from 1996 to 2012 had spillover effects up to the 90th percentile of distribution, reducing inequality and having little negative employment effects thanks to workers moving from low to high productivity firms
		Quantile regression	Inequality measures	Firm pay		(+)(-)(insignificant)	
		Spatial OLS	Employment level	Kaitz index	Generalised as log difference between minimum wage and wage distribution of a	(+)(-)(insignificant)	
		Instrumental variables					
Hinojosa (2019)	To see the dynamics of minimum wage effects on income distribution post-hyper-inflation Brazil	Panel data, with regional and time fixed effects	Difference between wage percentile and central wage (at the 60th percentile)	Minimum wage	In "normal" and quadratic forms	(-)(-)(-)	Minimum wage increases represent 35-50% of wage inequality reduction from 1995 to 2015, with larger effects on formal workers, and showing very small adverse employment effects that do not outweigh inequality reduction
		Instrumental variables (2SLS)	Employment rate	Difference between actual minimum wage and prior central wage, plus square of said difference	First instrument	(-)(-)(-)	
			Working hours	Fraction of hourly wages at or below minimum wage "bind" wage distributions at a certain region	Second instrument	(-)(-)(-)	
Jales (2018)	To estimate minimum wage effects in Brazil, given its large informal sector	Regression discontinuity design (RDD)	Unemployment	Minimum wage	Growth analysed for each year of the sample, at the monthly level	(+)(+)(+)(-)(+)(+)(-)	An increase in minimum wage increases the likelihood of a formal worker transitioning to the informal sector, with unemployment effects being found as well from the minimum wage policy. While average wages grow and wage inequality decreases, labour tax revenue (to pay for e.g. pensions) decreases as well
			Non-compliance to the minimum wage				
		Nonparametric kernel estimation (for density and probability analyses)	Average wages				
			Wage inequality				
			Sector mobility (from formal to informal, and vice-versa)				
			Size of informal sector				
			Labour tax revenues				

Source: Own elaboration.

Figure 18: Regional economic activity indexes

	Main objective(s)	Methodology	Dependent variable (y)	Explanatory variable(s) (x)		Estimated effect(s)	Conclusion(s)
Amaral et al. (2010)	To estimate effects of market potential on regional wage imbalances in Brazilian munipalities	Panel data using KKP spatial specfication (from Kapoor et al. (2007), using spatial first-order AR error term for each period of time), and random effects	Wage	Market potential	Sum of income of a municipality, with sum of income of other municipalities divided by distance between their centres	(+)(insignificant)(+)(+)	Market potential, measured as sum of internal and external incomes, and productive concentration of a location have strong links to nomimal wages of that locality
				Literacy ratio	For adults over 25		
				Levels of instruction	Different levels		
		Instrumental variables (2SLS)		Absolute latitute of municipalties' centroid, plus its squared term	Instrument		
Crone & Clayton-Matthews (2005)	To show indexes to measure regional economic activity in the United States	Stock-Watson basic method (time series OLS), using the Kalman filter/smoother to get maximum likelihood and smoothed state estimators	Coincident index, for economic activity of a state	Nonagricultural employment		57.4% median contribution to the coincident index	Despite constraints such as availability of data and different time-periods to integrate into indexes, the indexes that can be produced using the paper's methodology are still consistent in providing a good, and single, measure of various regional economic aspects from poverty to general economic activity, even in absence of GSP (gross state product)
				Unemployment rate	Using common factor, as calculate in Stock-Watson's methodology	17.7% median contribution to the coincident index	
		Combining linear transformation recalibrations of the Kalman filter/smoother, to give an index where its trend rate of growth is equal to each state's real gross state product over period of estimation, and variance around the trend given by weighted average of the		Average hours worked in manufacturing	Seasonally adjusted for each state	4.9% median contribution to the coincident index	
				Real wage/salary disembursements	Deflated by national CPI-U, and excluding proprietor's incomes	14.3% median contribution to the coincident index	
Hacker et al. (2013)	To analyse whether concentration of population leads to economic growth, or vice-versa; and whether the relationship varies by type of region	Time-series OLS	Average wages (wage-sums to employment ratio)	Population concentration	Using Shannon index, going from 0 for extreme concentration to 1 for the opposite	(+ for cities, metro regions; insignificant for rural regions)	The correct path is population agglomeration, leading towards economic growth, in metro and city regions; whereas rural regions have no defined direction, which shows urban setting have greater dependence on economics of agglomeration
		First-differenced vector autoregressive (DVAR) Granger causality approach, to check whether (the prediction of) future events can cause present effects to take place (when past events do so)					
		Wald F-test					
Navarro et al. (2017)	To create a competitiveness index for Spanish regions, using 60+ variables	Weighted index, for every region	Productive capital index	Business culture, sectorial concentration/specialization, internationalisation,		(varied)	Madrid, Basque Country and Catalunya are the most competitive regions in Spain, while Castilla la Mancha, Extremadura and Canarias are on the opposite side of the specturm
			Human capital index	Workforce availability, highly qualified workforce, managerial qualifications, post-			
			Public capital index	Knowledge infrastructures, transportation and communication infrastructures,			
Stock & Watson (1999)	To check the best method/variables to forecast inflation in the United States, given the reliance on the Phillips curve over several decades	Time-series OLS	Inflation (CPI and PCE deflated)	Industrial production			The best-performing forecast of inflation is done via a combination of the Phillips curve with a composite index of 168 individual activity measures, which consistently overperforms traditional and other alternative methods of forecasting inflation in the United States
				Real personal income			
				Total real manufacturing and trade sales		Best performer (individually)	
		Kalman filter/smoother		Number of employees on nonagricultural payrolls			
				Utilization rate in manufacturing		Best performer (individually)	
				Housing starts			
				Unemployment rate (of males, age 25-54)		Base comparison	
		Ridge regression		Other 168 economics indicators	Interest rates, nominal money, exchange rates, prices and wages, output, employment, sales, consumption, housing, inventories/orders, stock prices, etc.	Best performer, when combined with the Phillips curve	
Vidal et al. (2017)	To develop (monthly) indicators that will assess regional economic activity, using Stock-Watson (1991)'s model for national economies, and Latin American data	Dynamic factor model (DFM) using maximum likelihood, along with a Kalman fiter	Economic activity of Valle del Cauca region (common factor estimation)	Ground sugarcane	Agricultural activity	0.06 contribution to the estimated common factor	Regional industrial production index, vehicle sales and exports are the most relevant indicators of economic activity in the Valle del Cauca region, which is a peculiar region in Colombia since it is somewhat "detached" from national cycles due to a certain degree of independence
				Cement shipments	Civil construction activity	0.06 contribution to the estimated common factor	
		Litterman (1983) method, for temporal disaggregation of annual GDP		Nonresidential energy consumption	Energy industry/demand	0.12 contribution to the estimated common factor	
				New vehicle sales	Consumption	0.24 contribution to the estimated common factor	
		Univariate structural time-series methodology by Harvey (1989)		Exports at constant prices	External impact, industrial dynamics	0.16 contribution to the estimated common factor	
				Regional Industrial Production Index	Industrial dynamics	0.22 contribution to the estimated common factor	
				Imports at constant prices	External impact, consumption	0.15 contribution to the estimated common factor	

Source: Own elaboration.

Figure 19: Minimum wage and growth

	Main objective(s)	Methodology	Dependent variable (y)	Explanatory variable(s) (x)		Estimated effect(s)	Conclusion(s)
Amri (2018)	To determine the existence, and direction, of the causal relationship between economic growth, regional minimum wages, unemployment rate and labor force participation	Panel co-integration test	Economic growth	Economic growth		(+ to MW) (insignificant to unemployment) (- to LFP)	Via co-integration, vector and VAR Granger causality tests, there is evidence of positive "flow" from minimum wages to regional economic growth; and bidirectional links between growth and minimum wages, and minimum wages and LFP
		Panel vector error correction model	Regional minimum wages	Regional minimum wages		(+ to growth) (+ to unemployment) (+ to LFP)	
		Panel data VAR Granger causality test	Unemployment rate	Unemployment rate		(+ to growth) (+ to MW) (- to LFP)	
			Labor force participation	Labor force participation		(+ to growth) (+ to MW) (- to unemployment)	
Arestis et al. (2016)	To assess the origin of Brazil's economic growth issues since years post Great Recession in 2008	Generalized method of moments (GMM), instrumenting explanatory variables with lagged versions of themselves	Consumption	Wage income		(+)(+)(+)	Pre- and post-financial crisis, consumption is a big driver of general economic growth in Brazil
				Non-wage income		(+)(+)(+)	
				Interest rate	Proxies for liquid assets	(-)(N/A)(N/A)	
			Imports	Credit		(+)(N/A)(N/A)	
		Vector error correction (VEC)		Real exchange rate		(N/A)(-)(N/A)	
			(Private) investment	Market interest rate		(N/A)(N/A)(-)	
				Profit share		(N/A)(N/A)(-)	
Askenazy (2003)	To study the impact of minimum wages on growth for an "innovator country"	OLS	Economic growth	Kaitz index	The ratio of minimum wage to average wage	(insignificant)	Minimum wage, combined with an increase in exports, can lead to economic growth and to a Pareto-optimal state even in case of employment losses taking place, as long as there is not an "oversubstitution" of unskilled workers for skilled counterparts
				Exports	Interacted with Kaitz index difference, with US index as a benchmark	(+)	
Du & Wang (2020)	To examine the impact of minimum wages on firm markups	Fixed effects	Firm markup	Minimum wages	At the county level	(+)	Across various specifications, methods and controls, the minimum wage has a positive impact on firm markup, most likely thanks to incentivising value-enhancing activities and more efficient resource allocation
		Instrumental variables (IV)		Prediction of local average wage (proxy of minimum wages)	Instrumented by average wage of a certain industry in China, ratio of employment of	(+)	
		Difference-in-differences (DiD)					
Jung et al. (2021)	To gauge the effect of minimum wage hikes on consumption	Autoregreesive distributed lag (ARDL)	Quarterly real retail trade sales	Minimum wage	Weighted average of provincial minimum wages, adjusted by consumer price index over the years	(+)	In Canada, 1% increases in the minimum wage are associated with close to 0.5% increases in real retail sales, showing how minimum wages can serve to boost aggregate demand of an economy when needed
				Stock price index		(+)	
				House price index		(insignificant)	
				Unemployment rate		(-)	
				Treasure bill rate	Three-month time-period	(-)	
Katovich & Maia (2018)	To analyse the relation between labor productivity and wages in Brazil from a sectoral point of view	Fixed effects, with heteroscedasticity-consistent standard errors	Individual labour income per hour, and per sector	Minimum wage	At constant 2014 USD prices	(+)	While productivity is an important determinant of wages in several sectors, institutional factors such as the minimum wage may reign above it in determining earnings to workers in Brazil
				Schooling	Proxied by years of education	(+)	
				Female	Proportion	(-)	
				Non-white	Proportion	(-)	
				Age		(+ if linear) (- if squared/non-linear)	
				Productivity	By state	(+)	
				Union rate	Proportion, by state	(+)	
				Formal rate	Proportion, by state	(+)	

Source: Own elaboration.

Annex B – Rio de Janeiro tables and maps

Table 12: Real minimum wage in Brazil from 1970 to 2022 (in January 2022 R$)

1970.01	R$ 852.87	1997.01	R$ 516.80
1971.01	R$ 845.61	1998.01	R$ 530.47
1972.01	R$ 846.69	1999.01	R$ 561.85
1973.01	R$ 867.46	2000.01	R$ 542.30
1974.01	R$ 885.07	2001.01	R$ 571.05
1975.01	R$ 880.76	2002.01	R$ 620.15
1976.01	R$ 862.55	2003.01	R$ 592.33
1977.01	R$ 859.00	2004.01	R$ 654.40
1978.01	R$ 877.97	2005.01	R$ 669.70
1979.01	R$ 848.46	2006.01	R$ 737.00
1980.01	R$ 937.20	2007.01	R$ 835.39
1981.01	R$ 929.38	2008.01	R$ 860.82
1982.01	R$ 984.95	2009.01	R$ 883.32
1983.01	R$ 949.90	2010.01	R$ 1,040.15
1984.01	R$ 826.35	2011.01	R$ 1,033.84
1985.01	R$ 762.47	2012.01	R$ 1,127.38
1986.01	R$ 787.74	2013.01	R$ 1,152.47
1987.01	R$ 783.36	2014.01	R$ 1,169.17
1988.01	R$ 725.35	2015.01	R$ 1,187.88
1989.01	R$ 703.97	2016.01	R$ 1,191.78
1990.01	R$ 681.93	2017.01	R$ 1,203.56
1991.01	R$ 540.19	2018.01	R$ 1,202.85
1992.01	R$ 702.99	2019.01	R$ 1,214.98
1993.01	R$ 716.73	2020.01	R$ 1,212.70
1994.01	R$ 663.17	2021.01	R$ 1,216.59
1995.01	R$ 427.25	2022.01	R$ 1,212.00
1996.01	R$ 500.27		

Source: IPEA. Own elaboration.

Table 13: Deflators for years 2000-2015, based on IPCA-IBGE, for R$ in January 2022 values

2000	3.655004
2001	3.394527
2002	3.016546
2003	2.759864
2004	2.564942
2005	2.426871
2006	2.352953
2007	2.252542
2008	2.126992
2009	2.039074
2010	1.925313
2011	1.807749
2012	1.708025
2013	1.612703
2014	1.515592
2015	1.369432

Source: IPEA. Own elaboration.

Table 14: Neighbourhoods in AP1

	AP	RP	Planning area	Planning region
Benfica	1	1.1	Centro	Centro
Caju	1	1.1	Centro	Centro
Catumbi	1	1.1	Centro	Centro
Centro	1	1.1	Centro	Centro
Cidade Nova	1	1.1	Centro	Centro
Estacio	1	1.1	Centro	Centro
Gamboa	1	1.1	Centro	Centro
Mangueira	1	1.1	Centro	Centro
Paqueta	1	1.1	Centro	Centro
Rio Comprido	1	1.1	Centro	Centro
Santa Teresa	1	1.1	Centro	Centro
Santo Cristo	1	1.1	Centro	Centro
Sao Cristovao	1	1.1	Centro	Centro
Saude	1	1.1	Centro	Centro

Source: Prefeitura do Rio de Janeiro. Own elaboration.

Table 15: Neighbourhoods in AP2

	AP	RP	Planning area	Planning region
Botafogo	2	2.1	Zona Sul e Grande Tijuca	Zona Sul
Catete	2	2.1	Zona Sul e Grande Tijuca	Zona Sul
Copacabana	2	2.1	Zona Sul e Grande Tijuca	Zona Sul
Cosme Velho	2	2.1	Zona Sul e Grande Tijuca	Zona Sul
Flamengo	2	2.1	Zona Sul e Grande Tijuca	Zona Sul
Gavea	2	2.1	Zona Sul e Grande Tijuca	Zona Sul
Gloria	2	2.1	Zona Sul e Grande Tijuca	Zona Sul
Humaita	2	2.1	Zona Sul e Grande Tijuca	Zona Sul
Ipanema	2	2.1	Zona Sul e Grande Tijuca	Zona Sul
Jardim Botanico	2	2.1	Zona Sul e Grande Tijuca	Zona Sul
Lagoa	2	2.1	Zona Sul e Grande Tijuca	Zona Sul
Laranjeiras	2	2.1	Zona Sul e Grande Tijuca	Zona Sul
Leblon	2	2.1	Zona Sul e Grande Tijuca	Zona Sul
Leme	2	2.1	Zona Sul e Grande Tijuca	Zona Sul
Rocinha	2	2.1	Zona Sul e Grande Tijuca	Zona Sul
Sao Conrado	2	2.1	Zona Sul e Grande Tijuca	Zona Sul
Urca	2	2.1	Zona Sul e Grande Tijuca	Zona Sul
Vidigal	2	2.1	Zona Sul e Grande Tijuca	Zona Sul
Alto da Boa Vista	2	2.2	Zona Sul e Grande Tijuca	Tijuca
Andarai	2	2.2	Zona Sul e Grande Tijuca	Tijuca
Grajau	2	2.2	Zona Sul e Grande Tijuca	Tijuca
Maracana	2	2.2	Zona Sul e Grande Tijuca	Tijuca
Praca da Bandeira	2	2.2	Zona Sul e Grande Tijuca	Tijuca
Tijuca	2	2.2	Zona Sul e Grande Tijuca	Tijuca
Vila Isabel	2	2.2	Zona Sul e Grande Tijuca	Tijuca

Source: Prefeitura do Rio de Janeiro. Own elaboration.

Table 16: Neighbourhoods in AP3 - RP 3.1 & 3.2

	AP	RP	Planning area	Planning region
Bonsucesso	3	3.1	Zona Norte	Ramos
Manguinhos	3	3.1	Zona Norte	Ramos
Mare	3	3.1	Zona Norte	Ramos
Olaria	3	3.1	Zona Norte	Ramos
Ramos	3	3.1	Zona Norte	Ramos
Abolicao	3	3.2	Zona Norte	Meier
Agua Santa	3	3.2	Zona Norte	Meier
Cachambi	3	3.2	Zona Norte	Meier
Encantado	3	3.2	Zona Norte	Meier
Engenho de Dentro	3	3.2	Zona Norte	Meier
Engenho Novo	3	3.2	Zona Norte	Meier
Jacare	3	3.2	Zona Norte	Meier
Lins de Vasconcelos	3	3.2	Zona Norte	Meier
Meier	3	3.2	Zona Norte	Meier
Piedade	3	3.2	Zona Norte	Meier
Pilares	3	3.2	Zona Norte	Meier
Riachuelo	3	3.2	Zona Norte	Meier
Rocha	3	3.2	Zona Norte	Meier
Sampaio	3	3.2	Zona Norte	Meier
Sao Francisco Xavier	3	3.2	Zona Norte	Meier
Todos os Santos	3	3.2	Zona Norte	Meier

Source: Prefeitura do Rio de Janeiro. Own elaboration.

Table 17: Neighbourhoods in AP3 - RP 3.3 & 3.4

	AP	RP	Planning area	Planning region
Bento Ribeiro	3	3.3	Zona Norte	Madureira
Campinho	3	3.3	Zona Norte	Madureira
Cascadura	3	3.3	Zona Norte	Madureira
Cavalcanti	3	3.3	Zona Norte	Madureira
Colegio	3	3.3	Zona Norte	Madureira
Engenheiro Leal	3	3.3	Zona Norte	Madureira
Honorio Gurgel	3	3.3	Zona Norte	Madureira
Iraja	3	3.3	Zona Norte	Madureira
Madureira	3	3.3	Zona Norte	Madureira
Marechal Hermes	3	3.3	Zona Norte	Madureira
Oswaldo Cruz	3	3.3	Zona Norte	Madureira
Quintino Bocaiuva	3	3.3	Zona Norte	Madureira
Rocha Miranda	3	3.3	Zona Norte	Madureira
Turiacu	3	3.3	Zona Norte	Madureira
Vaz Lobo	3	3.3	Zona Norte	Madureira
Vicente de Carvalho	3	3.3	Zona Norte	Madureira
Vila Cosmos	3	3.3	Zona Norte	Madureira
Vila da Penha	3	3.3	Zona Norte	Madureira
Vista Alegre	3	3.3	Zona Norte	Madureira
Del Castilho	3	3.4	Zona Norte	Inhauma
Engenho da Rainha	3	3.4	Zona Norte	Inhauma
Higienopolis	3	3.4	Zona Norte	Inhauma
Inhauma	3	3.4	Zona Norte	Inhauma
Maria da Graca	3	3.4	Zona Norte	Inhauma
Tomas Coelho	3	3.4	Zona Norte	Inhauma

Source: Prefeitura do Rio de Janeiro. Own elaboration.

Table 18: Neighbourhoods in AP3 - RP 3.5, 3.6 & 3.7

	AP	RP	Planning area	Planning region
Bras de Pina	3	3.5	Zona Norte	Penha
Cordovil	3	3.5	Zona Norte	Penha
Jardim America	3	3.5	Zona Norte	Penha
Parada de Lucas	3	3.5	Zona Norte	Penha
Penha	3	3.5	Zona Norte	Penha
Penha Circular	3	3.5	Zona Norte	Penha
Vigario Geral	3	3.5	Zona Norte	Penha
Acari	3	3.6	Zona Norte	Pavuna
Anchieta	3	3.6	Zona Norte	Pavuna
Barros Filho	3	3.6	Zona Norte	Pavuna
Coelho Neto	3	3.6	Zona Norte	Pavuna
Costa Barros	3	3.6	Zona Norte	Pavuna
Guadalupe	3	3.6	Zona Norte	Pavuna
Parque Anchieta	3	3.6	Zona Norte	Pavuna
Pavuna	3	3.6	Zona Norte	Pavuna
Ricardo de Albuquerque	3	3.6	Zona Norte	Pavuna
Bancarios	3	3.7	Zona Norte	Ilha do Governador
Cacuia	3	3.7	Zona Norte	Ilha do Governador
Cidade Universitaria	3	3.7	Zona Norte	Ilha do Governador
Cocota	3	3.7	Zona Norte	Ilha do Governador
Freguesia-Ilha	3	3.7	Zona Norte	Ilha do Governador
Galeao	3	3.7	Zona Norte	Ilha do Governador
Jardim Carioca	3	3.7	Zona Norte	Ilha do Governador
Jardim Guanabara	3	3.7	Zona Norte	Ilha do Governador
Monero	3	3.7	Zona Norte	Ilha do Governador
Pitangueiras	3	3.7	Zona Norte	Ilha do Governador
Portuguesa	3	3.7	Zona Norte	Ilha do Governador
Praia da Bandeira	3	3.7	Zona Norte	Ilha do Governador
Ribeira	3	3.7	Zona Norte	Ilha do Governador
Taua	3	3.7	Zona Norte	Ilha do Governador
Zumbi	3	3.7	Zona Norte	Ilha do Governador

Source: Prefeitura do Rio de Janeiro. Own elaboration.

Table 19: Neighbourhoods in AP4

	AP	RP	Planning area	Planning region
Anil	4	4.1	Baixada de Jacarepagua	Jacarepagua
Cidade de Deus	4	4.1	Baixada de Jacarepagua	Jacarepagua
Curicica	4	4.1	Baixada de Jacarepagua	Jacarepagua
Freguesia-Jpa	4	4.1	Baixada de Jacarepagua	Jacarepagua
Gardenia Azul	4	4.1	Baixada de Jacarepagua	Jacarepagua
Jacarepagua	4	4.1	Baixada de Jacarepagua	Jacarepagua
Pechincha	4	4.1	Baixada de Jacarepagua	Jacarepagua
Praca Seca	4	4.1	Baixada de Jacarepagua	Jacarepagua
Tanque	4	4.1	Baixada de Jacarepagua	Jacarepagua
Taquara	4	4.1	Baixada de Jacarepagua	Jacarepagua
Vila Valqueire	4	4.1	Baixada de Jacarepagua	Jacarepagua
Barra da Tijuca	4	4.2	Baixada de Jacarepagua	Barra da Tijuca
Camorim	4	4.2	Baixada de Jacarepagua	Barra da Tijuca
Grumari	4	4.2	Baixada de Jacarepagua	Barra da Tijuca
Itanhanga	4	4.2	Baixada de Jacarepagua	Barra da Tijuca
Joa	4	4.2	Baixada de Jacarepagua	Barra da Tijuca
Recreio dos Bandeirantes	4	4.2	Baixada de Jacarepagua	Barra da Tijuca
Vargem Grande	4	4.2	Baixada de Jacarepagua	Barra da Tijuca
Vargem Pequena	4	4.2	Baixada de Jacarepagua	Barra da Tijuca

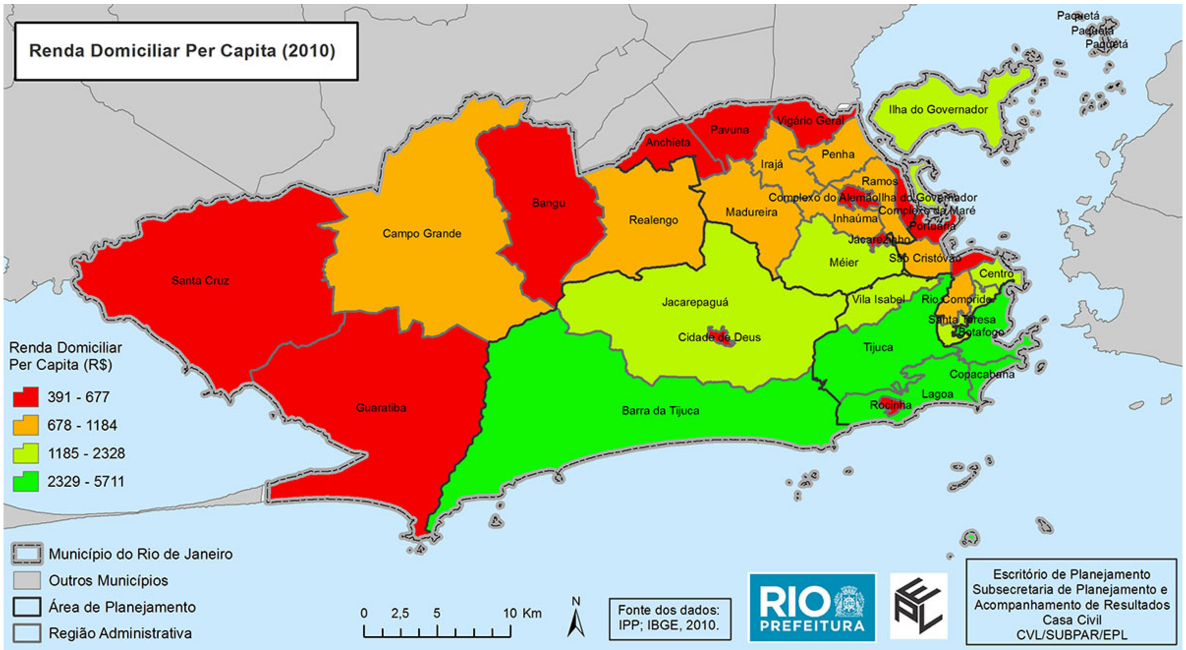
Source: Prefeitura do Rio de Janeiro. Own elaboration.

Table 20: Neighbourhoods in AP5

	AP	RP	Planning area	Planning region
Bangu	5	5.1	Zona Oeste	Bangu
Campo dos Afonsos	5	5.1	Zona Oeste	Bangu
Deodoro	5	5.1	Zona Oeste	Bangu
Jardim Sulacap	5	5.1	Zona Oeste	Bangu
Magalhaes Bastos	5	5.1	Zona Oeste	Bangu
Padre Miguel	5	5.1	Zona Oeste	Bangu
Realengo	5	5.1	Zona Oeste	Bangu
Senador Camara	5	5.1	Zona Oeste	Bangu
Vila Kennedy	5	5.1	Zona Oeste	Bangu
Vila Militar	5	5.1	Zona Oeste	Bangu
Campo Grande	5	5.2	Zona Oeste	Campo Grande
Cosmos	5	5.2	Zona Oeste	Campo Grande
Inhoaiba	5	5.2	Zona Oeste	Campo Grande
Santissimo	5	5.2	Zona Oeste	Campo Grande
Senador Vasconcelos	5	5.2	Zona Oeste	Campo Grande
Paciencia	5	5.3	Zona Oeste	Santa Cruz
Santa Cruz	5	5.3	Zona Oeste	Santa Cruz
Sepetiba	5	5.3	Zona Oeste	Santa Cruz
Barra de Guaratiba	5	5.4	Zona Oeste	Guaratiba
Guaratiba	5	5.4	Zona Oeste	Guaratiba
Pedra de Guaratiba	5	5.4	Zona Oeste	Guaratiba

Source: Prefeitura do Rio de Janeiro. Own elaboration.

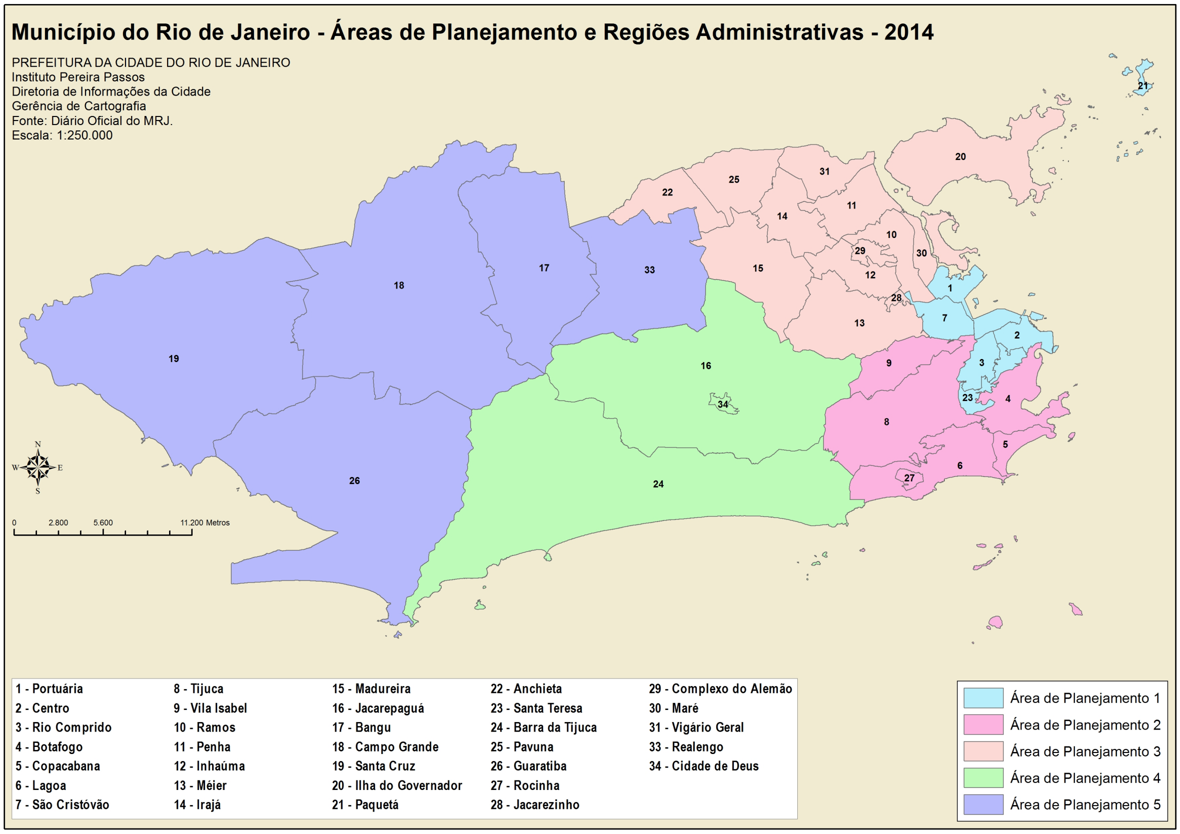
Figure 20: Average household income of Rio de Janeiro's capital in 2010 (in R$)



Available in < https://pds-pcrj.hub.arcgis.com/pages/renda>.

Source: Painel Rio.

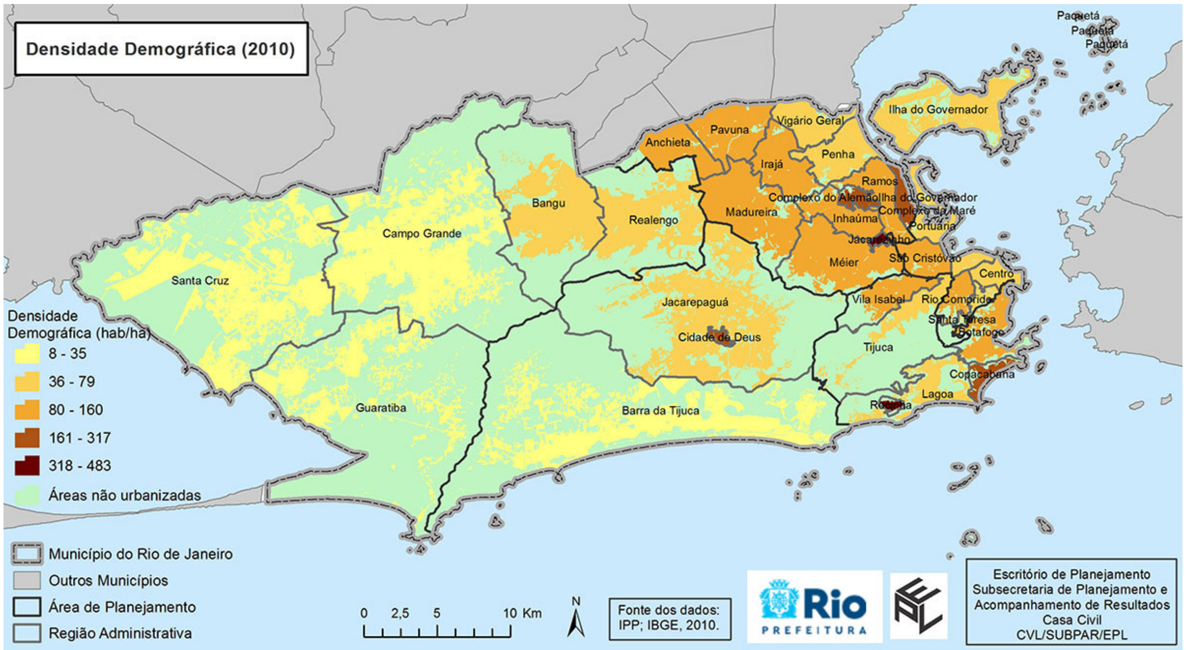
Figure 21: APs (planning areas) and RAs (administrative regions) in Rio



Available in <https://rioonwatch.org.br/?p=21068>.

Source: Rio On Watch.

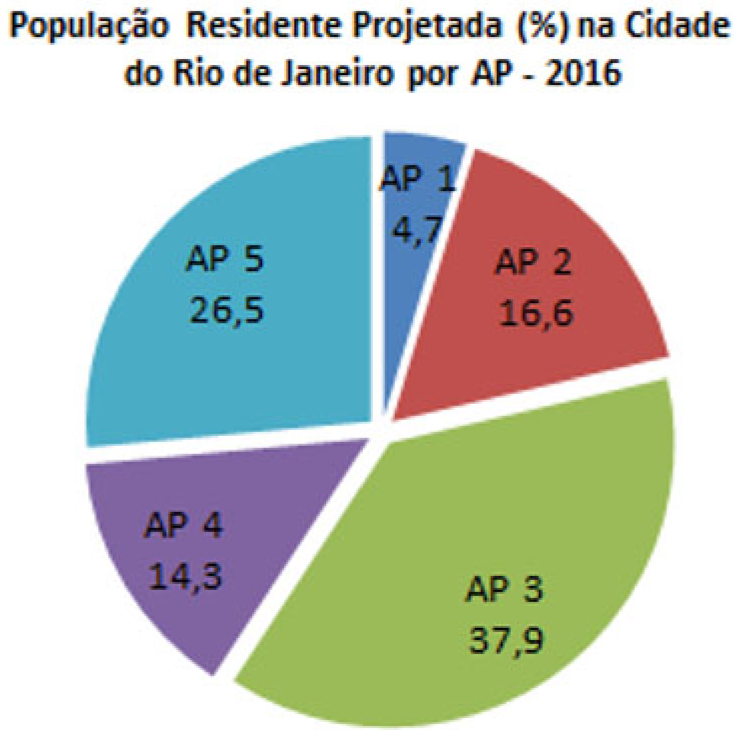
Figure 22: Demographic density of Rio de Janeiro in 2010



Available in <https://pds-pcrj.hub.arcgis.com/pages/painel>.

Source: Portal Rio.

Figure 23: 2016 projected population in Rio by AP



Available in <https://pds-pcrj.hub.arcgis.com/pages/painel>.

Source: Portal Rio.

Annex C – Descriptive statistics of Rio de Janeiro between 2000 and 2015

Table 21: Descriptive statistics - Censo 2000 vs. Censo 2010

			2000	2010
Earnings (in R$)		Mean	975.3	2073.4
		Median	453.0	950.0
		Std. dev.	2770.1	5372.3
Age		Mean	32.4	36.5
		Median	30.0	35.0
		Std. dev.	20.9	20.1
Working population			46.1%	59.8%
Working hours		Mean	43.0	37.3
		Median	40.0	40.0
		Std. dev.	14.7	16.8
Race				
	White		57.7%	51.1%
	Non-white		42.3%	48.9%
Gender				
	Male		47.6%	49.3%
	Female		52.4%	50.7%
Education				
	Less than primary education		53.7%	38.1%
	Less than secondary education		16.6%	16.3%
	Secondary education		20.3%	29.6%
	Tertiary education and above		9.1%	15.4%
PC at home				
	Yes		22.3%	59.2%
	No		77.7%	40.8%

Source: IBGE. Own elaboration.

Table 22: Descriptive statistics - PNAD 2001 vs. PNAD 2015

			2001	**2015**
Earnings (in R$)		Mean	817.7	2279.9
		Median	420.0	1270.0
		Std. dev.	1507.8	4463.1
Age		Mean	32.3	37.0
		Median	30.0	36.0
		Std. dev.	20.9	21.7
Working population			48.0%	50.4%
Workers under CLT			67.2%	75.9%
Working hours		Mean	42.7	38.3
		Median	40.0	40.0
		Std. dev.	13.6	13.4
Race				
	White		59.5%	44.3%
	Non-white		40.5%	55.7%
Gender				
	Male		46.8%	46.9%
	Female		53.2%	53.1%
Education				
	Less than primary education		57.2%	39.1%
	Less than secondary education		16.8%	17.3%
	Secondary education		19.2%	30.6%
	Tertiary education and above		6.9%	13.1%
People per domicile				
	1 to 3		60.0%	73.0%
	4 or more		40.0%	27.0%

Source: IBGE. Own elaboration.

Table 23: Descriptive statistics - RAIS "estabelecimentos" 2003 vs. 2014

		2003						2014					
		City	Baixada de Jacarepagua	Centro	Zona Norte	Zona Oeste	Zona Sul e Grande Tijuca	City	Baixada de Jacarepagua	Centro	Zona Norte	Zona Oeste	Zona Sul e Grande Tijuca
No. of workers	Mean	8.2	6.0	14.2	8.0	5.8	6.2	9.6	7.2	18.1	9.3	7.2	6.6
	Median	0.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
	Std. dev.	288.8	39.5	486.8	352.1	45.1	81.9	287.4	56.7	626.7	131.0	59.7	100.2
No. of CLT workers	Mean	6.4	6.0	8.4	6.4	5.5	5.9	7.9	7.2	10.3	8.7	7.1	6.4
	Median	0.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
	Std. dev.	63.7	39.5	83.1	54.4	41.6	74.8	91.3	56.4	95.6	115.1	59.7	95.4
No. of statutory workers	Mean	1.9	0.0	5.8	1.7	0.3	0.4	1.7	0.0	7.7	0.6	0.0	0.3
	Median	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Std. dev.	268.3	0.7	462.4	324.8	17.5	33.2	268.1	5.8	610.1	53.9	1.9	30.4
% of companies		100.0%	12.2%	22.8%	27.7%	9.2%	28.1%	100.0%	18.4%	20.3%	24.6%	10.2%	26.5%
Simples Nacional adopters		42.3%	40.6%	37.0%	50.5%	55.7%	35.7%	41.9%	40.5%	34.2%	52.3%	54.0%	35.4%
RAIS Negativa		47.5%	42.7%	49.9%	49.2%	54.1%	40.0%	49.7%	48.4%	52.6%	48.2%	50.1%	47.9%
Active company during the year (from 2005)		82.5%	84.4%	80.7%	81.3%	79.7%	85.3%	79.7%	82.0%	77.6%	78.7%	77.2%	83.2%
Company size (by no. of workers)													
	0 to 19	95.0%	94.4%	94.2%	95.0%	95.3%	95.8%	93.9%	93.9%	92.7%	93.4%	94.7%	95.5%
	20 to 49	3.2%	3.7%	3.5%	3.1%	2.9%	2.9%	3.9%	3.9%	4.4%	4.2%	3.5%	3.0%
	50 to 99	0.9%	1.1%	1.0%	1.0%	1.0%	0.7%	1.2%	1.2%	1.4%	1.3%	1.1%	0.9%
	100 to 499	0.7%	0.7%	1.0%	0.7%	0.7%	0.5%	0.8%	0.8%	1.2%	0.9%	0.7%	0.5%
	500 or more	0.2%	0.1%	0.3%	0.1%	0.1%	0.1%	0.2%	0.1%	0.4%	0.2%	0.1%	0.1%

Source: Ministério do Trabalho e Previdência. Own elaboration.

Table 24: Descriptive statistics - RAIS "vínculos" 2003 vs. 2014

			2003						2014					
			City	Baixada de Jacarepagua	Centro	Zona Norte	Zona Oeste	Zona Sul e Grande Tijuca	City	Baixada de Jacarepagua	Centro	Zona Norte	Zona Oeste	Zona Sul e Grande Tijuca
Earnings (in R$)		Mean	1188.5	924.8	1524.8	874.9	782.3	1188.3	2688.4	2209.2	3764.2	1866.9	1707.8	2575.8
		Median	593.3	479.0	792.4	526.4	451.1	563.1	1378.0	1244.2	1969.0	1158.9	1104.9	1366.8
		Std. dev.	1890.2	1698.7	2286.5	1205.4	1099.2	1947.3	4163.8	3797.2	5184.3	2571.8	2175.6	4173.8
Age		Mean	36.1	33.3	37.7	36.0	33.8	35.6	36.7	34.7	38.8	36.4	34.8	36.1
		Median	35.0	31.0	37.0	35.0	32.0	34.0	35.0	33.0	37.0	34.0	33.0	34.0
		Std. dev.	11.4	10.5	11.7	11.3	10.6	11.4	12.1	11.2	12.3	12.1	11.6	12.0
Working hours		Mean	40.4	42.2	38.3	41.7	41.7	41.1	40.9	41.8	39.5	41.8	41.7	40.8
		Median	44.0	44.0	40.0	44.0	44.0	44.0	44.0	44.0	40.0	44.0	44.0	44.0
		Std. dev.	6.7	5.6	7.1	5.8	7.1	6.7	6.6	6.2	6.6	6.1	6.7	7.1
Statutory workers			17.4%	0.1%	32.1%	15.8%	3.5%	4.5%	12.6%	0.2%	32.8%	4.6%	0.3%	2.5%
% of workers			100%	9.6%	37.3%	26.1%	6.5%	20.5%	100%	15.3%	34.0%	23.8%	8.0%	18.8%
Race (from 2006)														
	White		58.7%	58.5%	60.4%	54.3%	54.6%	62.2%	49.8%	52.3%	53.5%	43.9%	47.5%	52.7%
	Non-white		41.3%	41.5%	39.6%	45.7%	45.4%	37.8%	50.2%	47.7%	46.5%	56.1%	52.5%	47.3%
Gender														
	Male		60.3%	62.4%	59.2%	59.5%	63.1%	60.8%	58.1%	59.6%	56.4%	60.9%	58.3%	55.5%
	Female		39.7%	37.6%	40.8%	40.5%	36.9%	39.2%	41.9%	40.4%	43.6%	39.1%	41.7%	44.5%
Education														
	Less than primary education		21.4%	26.7%	15.5%	24.9%	20.6%	25.5%	11.2%	14.3%	7.6%	12.7%	10.7%	13.2%
	Less than secondary education		25.9%	28.9%	20.1%	31.0%	35.8%	25.0%	19.5%	22.6%	13.0%	25.1%	24.5%	19.9%
	Secondary education		33.3%	32.4%	35.7%	31.5%	34.4%	31.3%	47.7%	48.2%	44.2%	51.1%	54.5%	47.2%
	Tertiary education and above		19.3%	11.9%	28.6%	12.5%	9.2%	18.3%	21.7%	14.9%	35.2%	11.1%	10.3%	19.7%
Time at employment														
	Less than a year		33.7%	44.0%	30.3%	33.8%	39.3%	32.2%	41.4%	51.7%	32.0%	44.7%	48.2%	42.2%
	One to five years		34.7%	40.1%	28.6%	37.8%	40.5%	38.6%	35.0%	35.5%	32.3%	36.8%	37.9%	36.1%
	More than five years		31.6%	15.8%	41.1%	28.4%	20.2%	29.2%	23.4%	12.8%	35.8%	18.5%	13.8%	21.7%
% of workers who kept job at end of year			73.8%	66.7%	75.9%	74.1%	71.5%	74.4%	66.5%	59.1%	73.3%	65.5%	62.2%	63.4%

Source: Ministério do Trabalho e Previdência. Own elaboration.

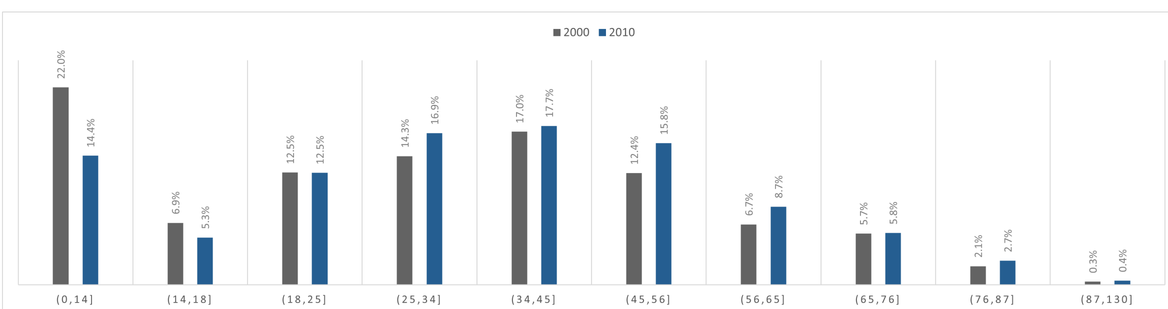
Figure 24: Working population in Rio between 2000 and 2010

Source: IBGE. Own elaboration.

Figure 25: Race in Rio between 2000 and 2010

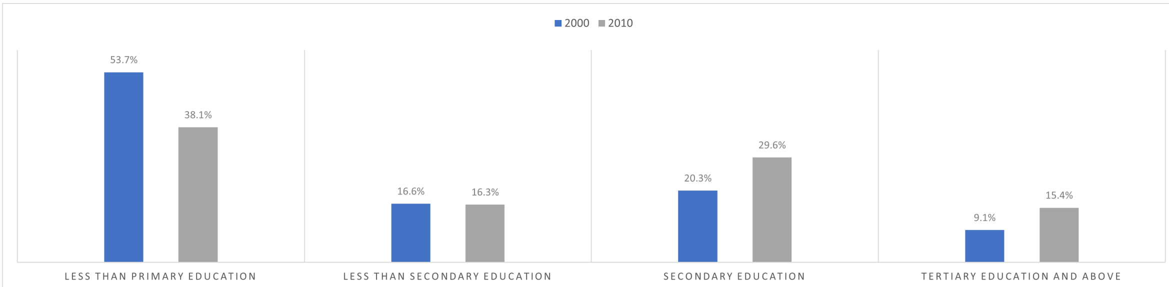
Source: IBGE. Own elaboration.

Figure 26: Age distribution in Rio between 2000 and 2010



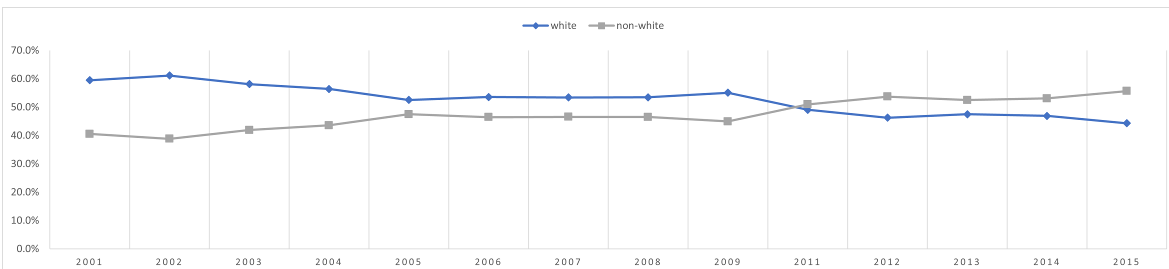
Source: IBGE. Own elaboration.

Figure 27: Education in Rio between 2000 and 2010



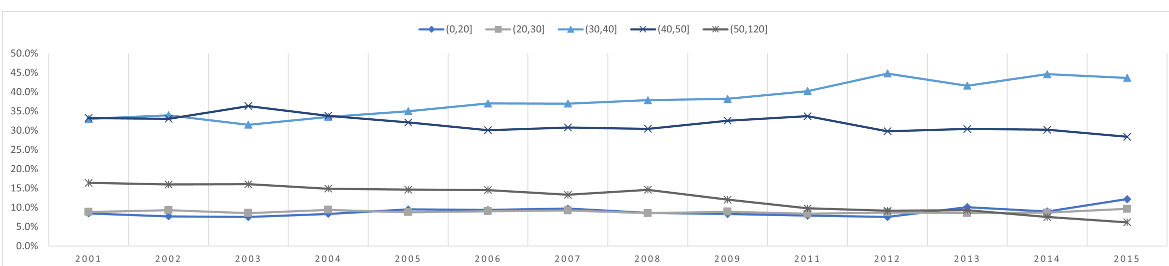
Source: IBGE. Own elaboration.

Figure 28: Racial composition in Rio between 2001 and 2015



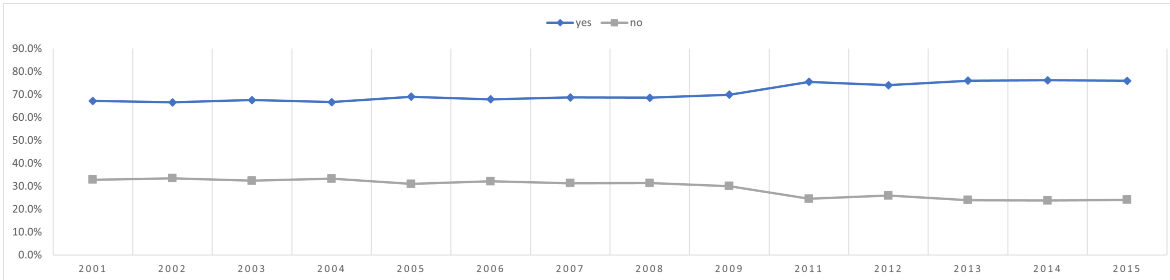
Source: IBGE. Own elaboration.

Figure 29: Worked hours per week between 2001 and 2015



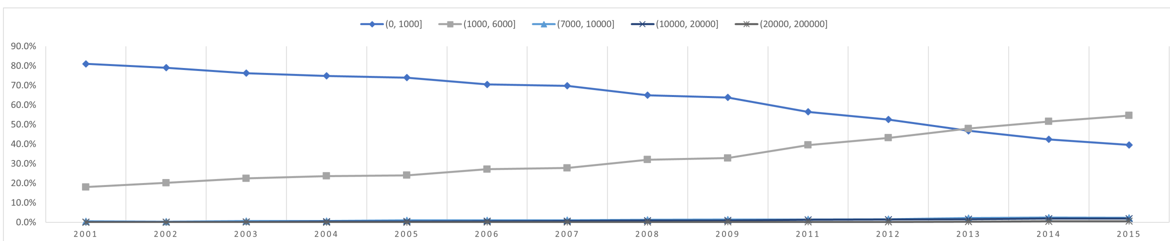
Source: IBGE. Own elaboration.

Figure 30: Workers under CLT in Rio between 2001 and 2015



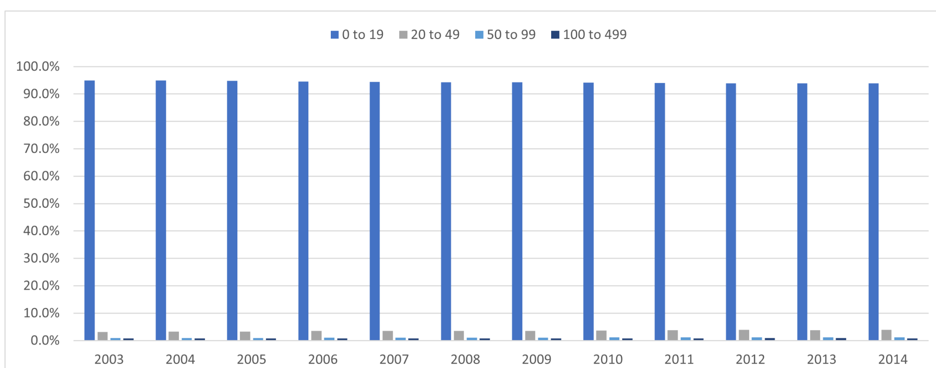
Source: IBGE. Own elaboration.

Figure 31: Income per individual in Rio (in R$) between 2001 and 2015



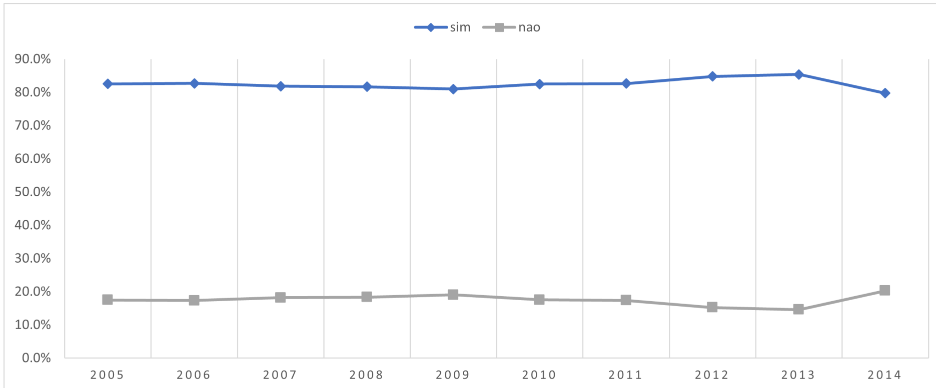
Source: IBGE. Own elaboration.

Figure 32: Company size in Rio between 2003 and 2014



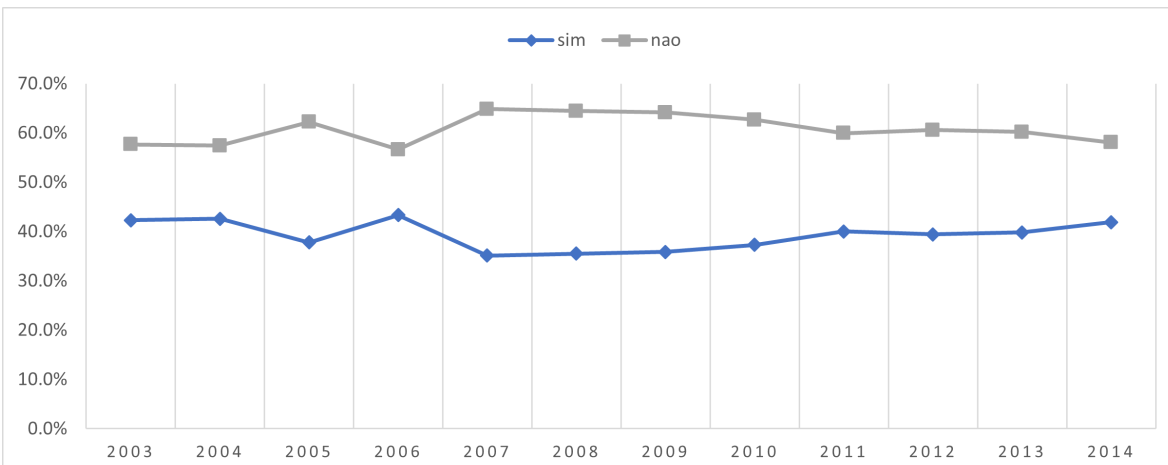
Source: Ministério do Trabalho e Previdência. Own elaboration.

Figure 33: Active companies in Rio, between 2005 and 2014



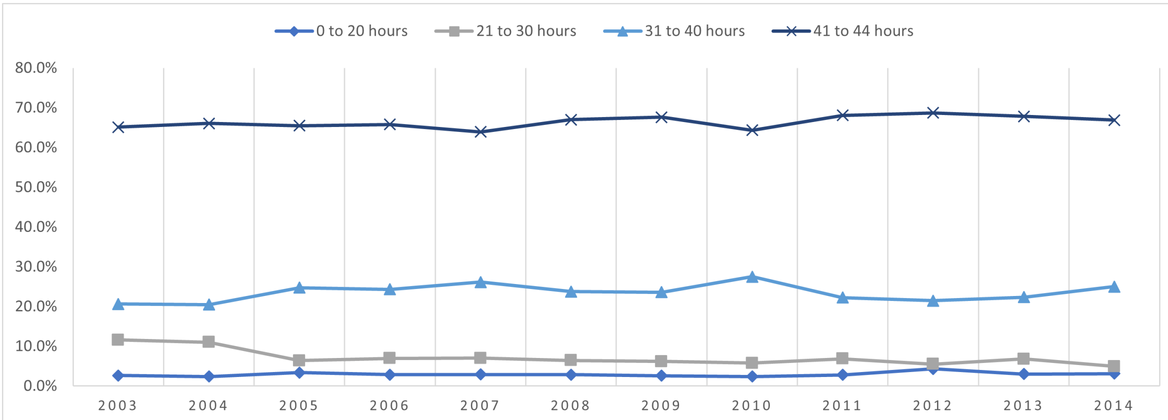
Source: Ministério do Trabalho e Previdência. Own elaboration.

Figure 34: Companies under simplified tax regime in Rio between 2003 and 2014



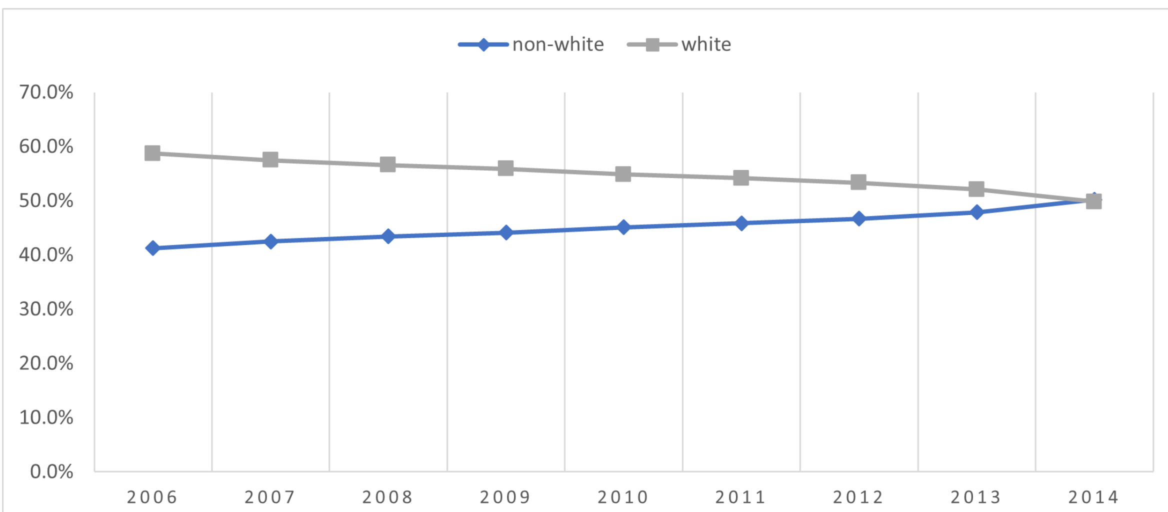
Source: Ministério do Trabalho e Previdência. Own elaboration.

Figure 35: Working hours per week in Rio between 2003 and 2014



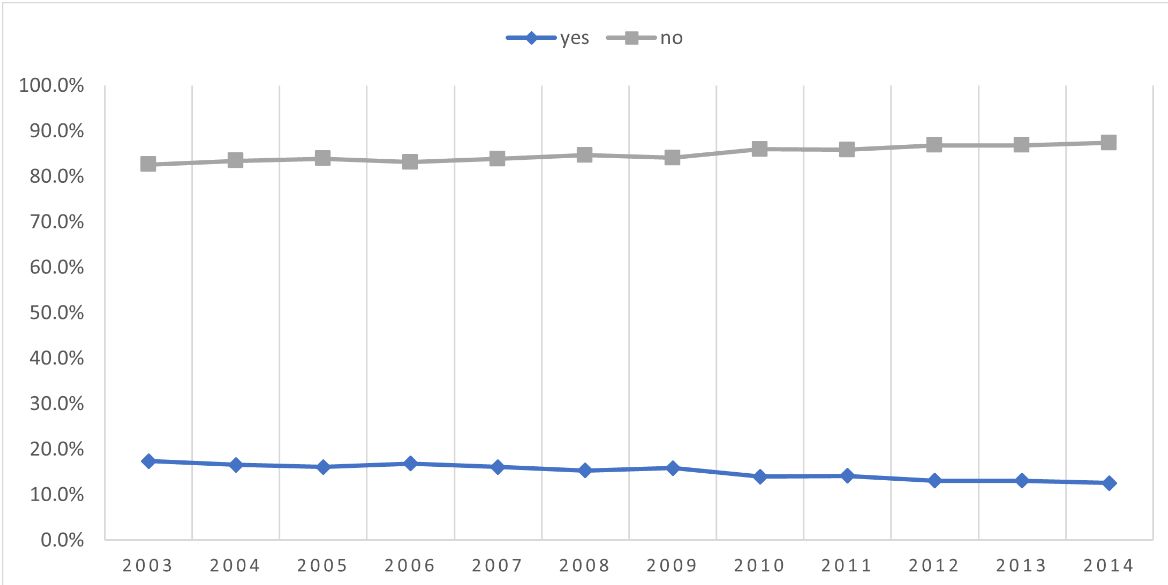
Source: Ministério do Trabalho e Previdência. Own elaboration.

Figure 36: White/non-white workers in Rio between 2003 and 2014



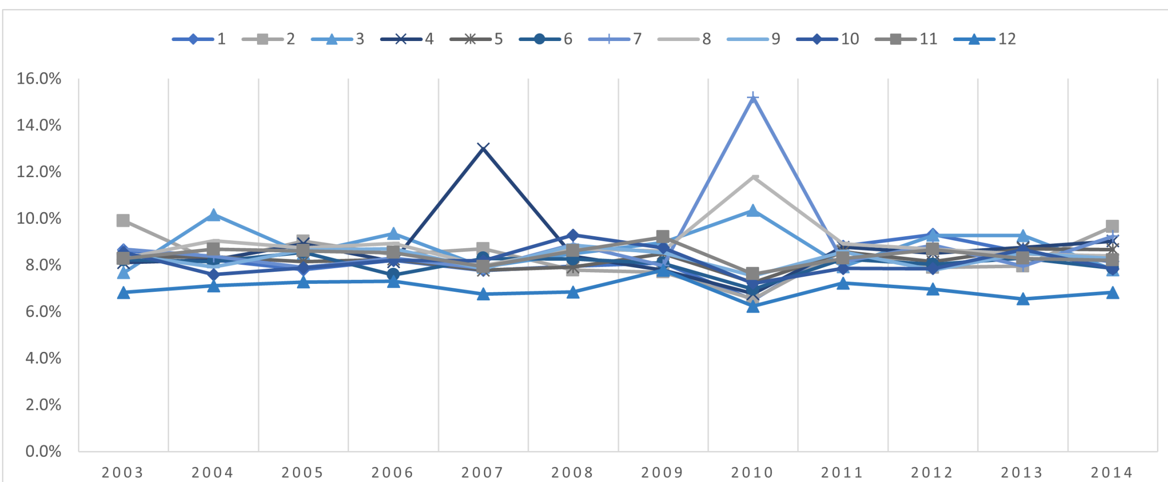
Source: Ministério do Trabalho e Previdência. Own elaboration.

Figure 37: Statutory/non-statutory workers in Rio between 2003 and 2014



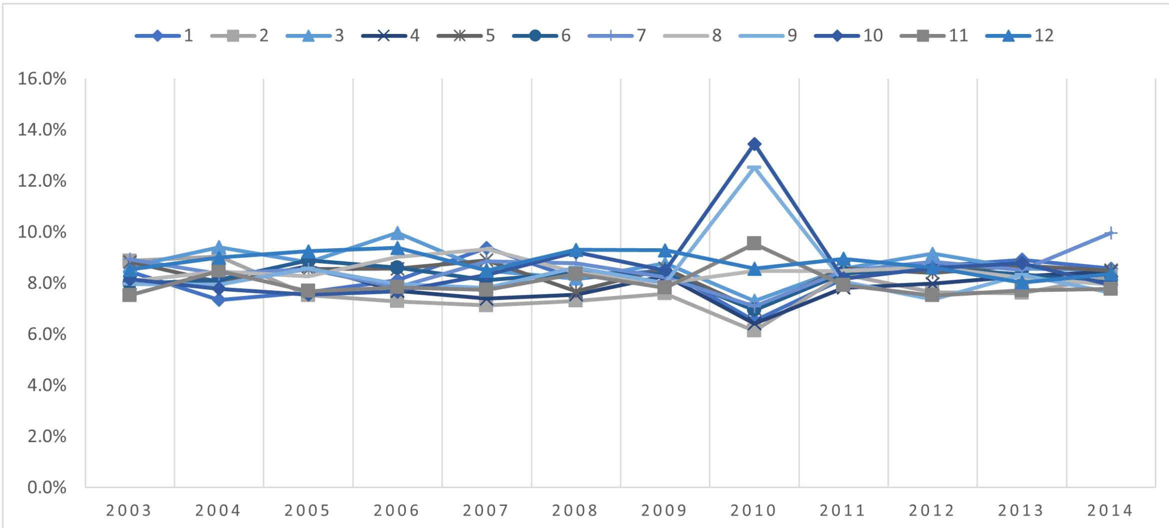
Source: Ministério do Trabalho e Previdência. Own elaboration.

Figure 38: Hirings per month in Rio between 2003 and 2014



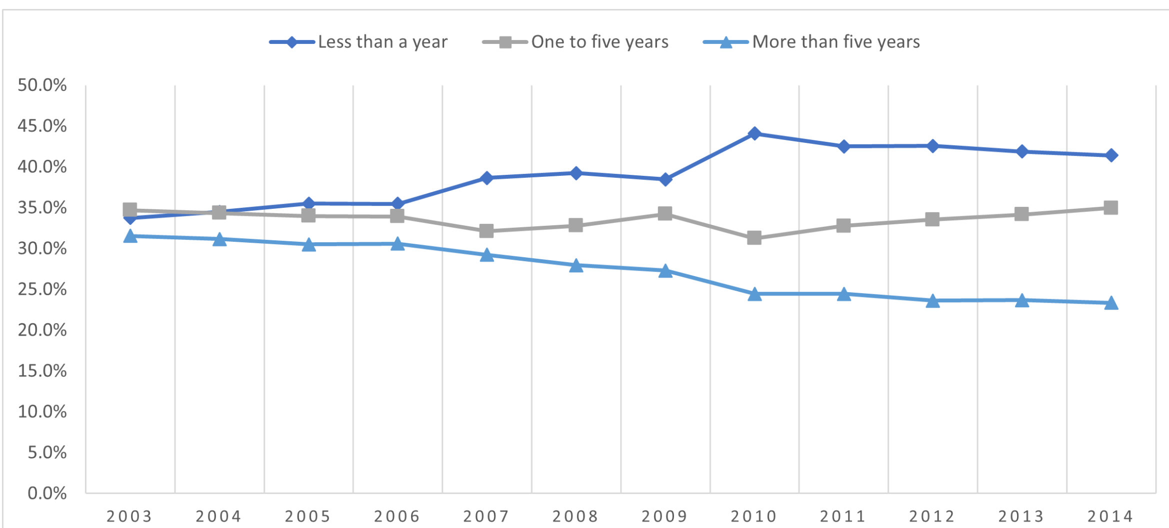
Source: Ministério do Trabalho e Previdência. Own elaboration.

Figure 39: Firings per month in Rio between 2003 and 2014



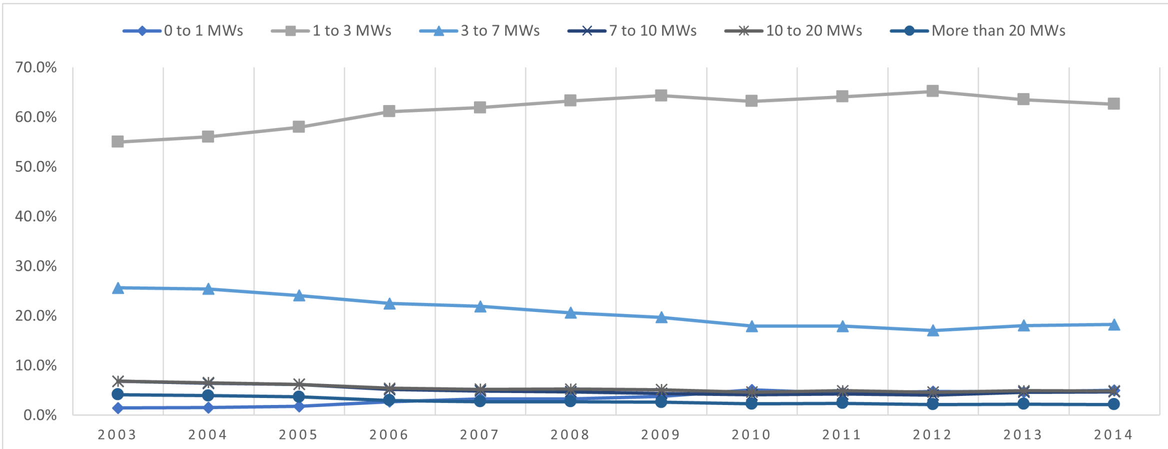
Source: Ministério do Trabalho e Previdência. Own elaboration.

Figure 40: Job longevity in Rio between 2003 and 2014



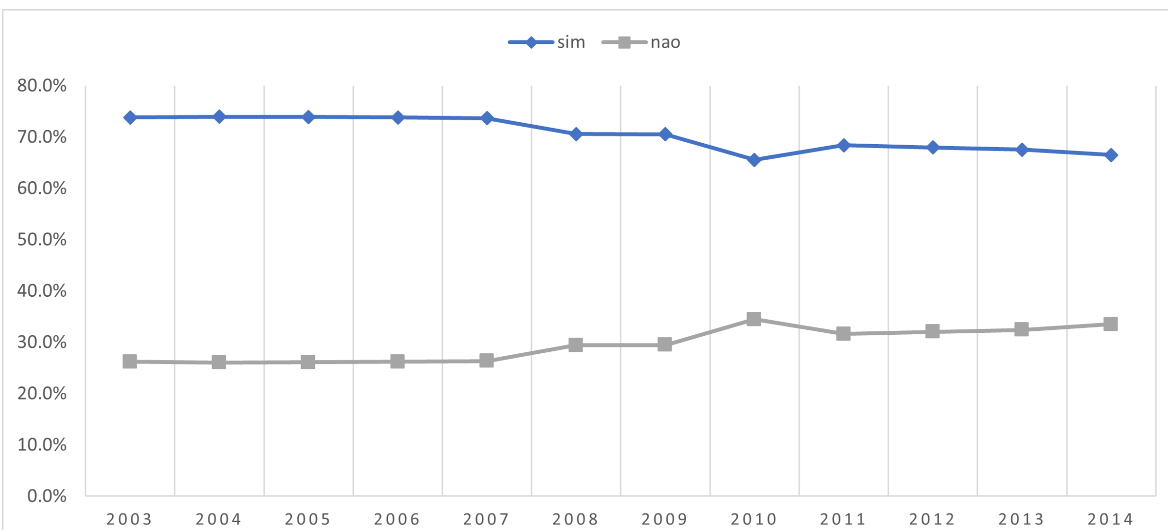
Source: Ministério do Trabalho e Previdência. Own elaboration.

Figure 41: Average income per worker in Rio (by minimum wage units) between 2003 and 2014



Source: Ministério do Trabalho e Previdência. Own elaboration.

Figure 42: Workers still at company by the end of the year in Rio, between 2003 and 2014



Source: Ministério do Trabalho e Previdência. Own elaboration.

Annex D – Regression results

PMM-MICE results in the RAIS "estabelecimentos" database

Table 25: Frequency of companies per AP from 2003 to 2015, in the original RAIS "estabelecimentos" database

	2003	2004	2005	2006	2007	2008	2009
OG1	20.9%	20.6%	20.0%	20.5%	20.8%	21.0%	20.8%
OG2	25.7%	25.4%	24.5%	25.5%	24.4%	25.0%	25.2%
OG3	25.4%	24.6%	23.6%	24.1%	23.7%	23.7%	23.7%
OG4	11.2%	11.2%	12.2%	13.0%	13.7%	14.3%	14.9%
OG5	8.5%	8.5%	8.3%	8.8%	9.1%	9.3%	9.5%
OGNA	8.3%	9.7%	11.4%	8.0%	8.2%	6.7%	5.8%

	2010	2011	2012	2013	2014	2015
OG1	20.4%	20.4%	19.8%	19.6%	19.0%	19.1%
OG2	25.3%	25.4%	25.0%	25.0%	24.9%	25.0%
OG3	23.8%	24.1%	23.7%	23.5%	23.1%	23.2%
OG4	15.7%	16.3%	16.5%	17.0%	17.3%	17.8%
OG5	9.8%	9.7%	9.6%	9.7%	9.6%	9.7%
OGNA	5.0%	4.1%	5.4%	5.2%	6.1%	5.2%

Source: Ministério do Trabalho e Previdência. Own elaboration.

Table 26: Frequency of companies per AP from 2003 to 2015, in the PMM-MICE-treated RAIS "estabelecimentos" database

	2003	2004	2005	2006	2007	2008	2009
mod1	22.0%	25.2%	24.2%	23.4%	23.2%	23.3%	22.5%
mod2	27.1%	28.1%	26.8%	27.0%	26.5%	25.9%	25.9%
mod3	29.8%	26.0%	26.8%	26.4%	26.0%	24.8%	25.5%
mod4	11.8%	12.0%	13.6%	13.9%	14.7%	16.3%	15.4%
mod5	9.3%	8.8%	8.6%	9.2%	9.7%	9.8%	10.7%
modNA	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

	2010	2011	2012	2013	2014	2015
mod1	21.4%	21.9%	20.9%	21.5%	21.1%	20.9%
mod2	26.6%	25.7%	25.5%	26.0%	25.2%	25.9%
mod3	25.3%	25.7%	25.6%	24.8%	25.7%	24.6%
mod4	16.4%	16.7%	18.2%	17.6%	17.8%	18.6%
mod5	10.2%	10.0%	9.8%	10.0%	10.1%	10.1%
modNA	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Source: Ministério do Trabalho e Previdência. Own elaboration.

Table 27: Comparison of companies per AP in the RAIS "estabelecimentos" database (original vs. treated)

	2003	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
1	1.1%	4.5%	4.1%	3.0%	2.3%	2.3%	1.7%
2	1.3%	2.7%	2.3%	1.5%	2.0%	0.9%	0.7%
3	4.4%	1.3%	3.2%	2.3%	2.2%	1.1%	1.8%
4	0.6%	0.8%	1.5%	0.9%	1.0%	2.0%	0.4%
5	0.9%	0.3%	0.3%	0.4%	0.7%	0.4%	1.1%
NA	-8.3%	-9.7%	-11.4%	-8.0%	-8.2%	-6.7%	-5.8%

	2010	**2011**	**2012**	**2013**	**2014**	**2015**
1	1.0%	1.5%	1.1%	1.9%	2.1%	1.8%
2	1.3%	0.3%	0.5%	1.0%	0.4%	0.9%
3	1.5%	1.5%	1.9%	1.3%	2.6%	1.4%
4	0.8%	0.5%	1.7%	0.6%	0.6%	0.8%
5	0.4%	0.3%	0.2%	0.4%	0.5%	0.4%
NA	-5.0%	-4.1%	-5.4%	-5.2%	-6.1%	-5.2%

Source: Ministério do Trabalho e Previdência. Own elaboration.

Table 28: Frequency of companies per RP from 2003 to 2015, in the original RAIS "estabelecimentos" database

	2003	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
OG1.1	20.9%	20.6%	20.0%	20.5%	20.8%	21.0%	20.8%
OG2.1	18.3%	18.0%	17.3%	18.1%	17.1%	17.7%	18.0%
OG2.2	7.4%	7.4%	7.2%	7.4%	7.3%	7.3%	7.3%
OG3.1	3.6%	3.4%	3.2%	3.3%	3.2%	3.3%	3.3%
OG3.2	6.0%	5.9%	5.7%	5.7%	5.7%	5.6%	5.6%
OG3.3	6.4%	6.2%	5.7%	6.0%	6.0%	5.9%	6.0%
OG3.4	1.3%	1.3%	1.3%	1.3%	1.3%	1.3%	1.3%
OG3.5	3.8%	3.6%	3.5%	3.6%	3.4%	3.4%	3.4%
OG3.6	1.8%	1.8%	1.7%	1.8%	1.8%	1.8%	1.8%
OG3.7	2.6%	2.5%	2.5%	2.4%	2.3%	2.4%	2.5%
OG4.1	4.3%	4.4%	5.0%	5.3%	5.6%	5.7%	5.9%
OG4.2	6.8%	6.8%	7.2%	7.8%	8.1%	8.6%	9.1%
OG5.1	3.5%	3.7%	3.5%	3.8%	4.0%	4.1%	4.1%
OG5.2	3.3%	3.2%	3.1%	3.3%	3.4%	3.5%	3.6%
OG5.3	1.3%	1.2%	1.2%	1.3%	1.3%	1.4%	1.4%
OG5.4	0.4%	0.4%	0.4%	0.4%	0.4%	0.4%	0.4%
OGNA	8.3%	9.7%	11.4%	8.0%	8.2%	6.7%	5.8%

	2010	**2011**	**2012**	**2013**	**2014**	**2015**
OG1.1	20.4%	20.4%	19.8%	19.6%	19.0%	19.1%
OG2.1	18.1%	18.2%	18.1%	18.2%	18.2%	18.3%
OG2.2	7.2%	7.2%	6.9%	6.8%	6.7%	6.7%
OG3.1	3.3%	3.3%	3.2%	3.2%	3.1%	3.1%
OG3.2	5.5%	5.6%	5.5%	5.5%	5.4%	5.4%
OG3.3	6.0%	6.1%	5.9%	5.9%	5.8%	5.8%
OG3.4	1.3%	1.3%	1.3%	1.3%	1.3%	1.3%
OG3.5	3.3%	3.4%	3.3%	3.2%	3.2%	3.2%
OG3.6	1.9%	2.0%	2.0%	1.9%	1.9%	1.9%
OG3.7	2.5%	2.6%	2.5%	2.5%	2.5%	2.5%
OG4.1	6.2%	6.4%	6.4%	6.5%	6.5%	6.7%
OG4.2	9.5%	9.9%	10.2%	10.5%	10.7%	11.1%
OG5.1	4.1%	3.9%	3.7%	3.7%	3.6%	3.6%
OG5.2	3.8%	3.8%	3.9%	4.0%	4.0%	4.1%
OG5.3	1.5%	1.5%	1.5%	1.5%	1.5%	1.5%
OG5.4	0.5%	0.5%	0.5%	0.5%	0.5%	0.5%
OGNA	5.0%	4.1%	5.4%	5.2%	6.1%	5.2%

Source: Ministério do Trabalho e Previdência. Own elaboration.

Table 29: Frequency of companies per RP from 2003 to 2015, in the PMM-MICE-treated RAIS "estabelecimentos" database

	2003	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
mod1.1	23.2%	23.3%	23.4%	23.5%	22.3%	23.1%	22.6%
mod2.1	19.3%	19.4%	19.6%	19.4%	19.3%	18.1%	18.8%
mod2.2	9.1%	7.6%	7.3%	7.6%	7.8%	8.5%	7.6%
mod3.1	4.9%	3.4%	3.8%	3.6%	3.4%	3.6%	3.5%
mod3.2	6.2%	7.2%	6.5%	6.2%	6.5%	5.8%	5.7%
mod3.3	7.6%	6.3%	6.0%	6.1%	6.0%	6.2%	6.1%
mod3.4	1.3%	1.3%	1.3%	1.3%	1.8%	1.3%	1.3%
mod3.5	3.8%	3.7%	4.2%	3.7%	4.1%	3.4%	3.8%
mod3.6	1.8%	3.4%	2.3%	1.9%	1.8%	1.9%	1.8%
mod3.7	2.6%	2.8%	2.6%	2.5%	2.8%	2.8%	2.5%
mod4.1	4.3%	5.1%	5.4%	5.9%	5.7%	6.6%	6.7%
mod4.2	7.1%	7.0%	8.2%	8.9%	8.4%	9.0%	9.9%
mod5.1	3.5%	3.9%	4.4%	4.0%	4.7%	4.1%	4.2%
mod5.2	3.6%	3.8%	3.3%	3.3%	3.7%	3.9%	3.6%
mod5.3	1.3%	1.3%	1.3%	1.8%	1.3%	1.4%	1.4%
mod5.4	0.4%	0.4%	0.4%	0.4%	0.4%	0.5%	0.4%
modNA	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

	2010	**2011**	**2012**	**2013**	**2014**	**2015**
mod1.1	22.9%	21.5%	21.4%	21.0%	21.1%	20.2%
mod2.1	18.8%	18.6%	18.4%	19.5%	18.7%	19.6%
mod2.2	7.5%	7.3%	7.0%	7.2%	7.1%	6.9%
mod3.1	3.3%	3.7%	3.6%	3.4%	3.9%	3.1%
mod3.2	5.5%	6.2%	5.8%	5.7%	5.4%	5.7%
mod3.3	6.1%	6.2%	6.1%	6.4%	5.8%	6.3%
mod3.4	1.3%	1.3%	1.3%	1.3%	1.3%	1.5%
mod3.5	3.5%	3.5%	3.6%	3.3%	3.6%	3.4%
mod3.6	2.2%	2.0%	2.0%	2.1%	1.9%	2.0%
mod3.7	3.1%	2.7%	3.0%	2.8%	3.5%	2.5%
mod4.1	6.4%	6.8%	6.8%	6.5%	6.8%	6.8%
mod4.2	9.5%	10.2%	10.8%	10.9%	10.8%	11.8%
mod5.1	4.1%	4.0%	4.1%	3.8%	3.6%	3.7%
mod5.2	3.8%	4.0%	4.0%	4.1%	4.5%	4.3%
mod5.3	1.5%	1.5%	1.6%	1.6%	1.5%	1.5%
mod5.4	0.5%	0.5%	0.5%	0.5%	0.5%	0.6%
modNA	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Source: Ministério do Trabalho e Previdência. Own elaboration.

Table 30: Comparison of work contracts per RP in the RAIS "estabelecimentos" database (original vs. treated)

	2003	2004	2005	2006	2007	2008	2009
1.1	2.2%	2.6%	3.3%	3.0%	1.5%	2.2%	1.9%
2.1	1.0%	1.5%	2.4%	1.2%	2.2%	0.4%	0.8%
2.2	1.7%	0.2%	0.1%	0.2%	0.5%	1.2%	0.3%
3.1	1.3%	0.0%	0.6%	0.3%	0.2%	0.3%	0.2%
3.2	0.3%	1.3%	0.9%	0.5%	0.8%	0.2%	0.1%
3.3	1.2%	0.1%	0.3%	0.1%	0.0%	0.2%	0.1%
3.4	0.0%	0.0%	0.1%	0.0%	0.5%	0.0%	0.0%
3.5	0.0%	0.1%	0.7%	0.0%	0.7%	0.0%	0.5%
3.6	0.0%	1.7%	0.6%	0.2%	0.0%	0.1%	0.0%
3.7	0.0%	0.3%	0.0%	0.0%	0.5%	0.4%	0.0%
4.1	0.0%	0.7%	0.4%	0.6%	0.1%	0.9%	0.8%
4.2	0.3%	0.2%	0.9%	1.1%	0.3%	0.4%	0.8%
5.1	0.0%	0.2%	0.9%	0.2%	0.8%	0.0%	0.1%
5.2	0.3%	0.6%	0.2%	0.0%	0.3%	0.4%	0.0%
5.3	0.0%	0.1%	0.0%	0.5%	0.0%	0.0%	0.0%
5.4	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%
NA	-8.3%	-9.7%	-11.4%	-8.0%	-8.2%	-6.7%	-5.8%

	2010	2011	2012	2013	2014	2015
1.1	2.4%	1.1%	1.6%	1.4%	2.0%	1.1%
2.1	0.7%	0.3%	0.3%	1.3%	0.6%	1.2%
2.2	0.3%	0.2%	0.1%	0.3%	0.5%	0.2%
3.1	0.0%	0.4%	0.4%	0.2%	0.8%	0.1%
3.2	0.0%	0.5%	0.3%	0.2%	0.0%	0.3%
3.3	0.1%	0.1%	0.2%	0.5%	0.0%	0.4%
3.4	0.1%	0.0%	0.0%	0.0%	0.0%	0.2%
3.5	0.2%	0.2%	0.3%	0.0%	0.4%	0.3%
3.6	0.3%	0.1%	0.0%	0.1%	0.0%	0.1%
3.7	0.6%	0.2%	0.5%	0.3%	1.0%	0.0%
4.1	0.2%	0.4%	0.4%	0.0%	0.3%	0.1%
4.2	0.0%	0.3%	0.6%	0.4%	0.0%	0.7%
5.1	0.1%	0.1%	0.3%	0.1%	0.0%	0.2%
5.2	0.0%	0.2%	0.2%	0.1%	0.5%	0.3%
5.3	0.0%	0.0%	0.1%	0.1%	0.0%	0.0%
5.4	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
NA	-5.0%	-4.1%	-5.4%	-5.2%	-6.1%	-5.2%

Source: Ministério do Trabalho e Previdência. Own elaboration.

PMM-MICE results in the RAIS “vínculos” database

Table 31: Frequency of work contracts per AP from 2000 to 2015, in the original RAIS "vínculos" database

	2000	2001	2002	2003	2004	2005	2006	2007
1	33.7%	33.4%	33.2%	35.0%	32.7%	35.6%	37.3%	39.9%
2	18.9%	18.9%	18.5%	19.3%	18.4%	17.2%	18.5%	17.2%
3	23.6%	24.6%	24.9%	24.6%	20.1%	19.5%	20.5%	20.9%
4	7.5%	7.9%	8.3%	9.0%	8.8%	9.6%	11.2%	11.3%
5	5.6%	5.8%	5.9%	6.1%	5.8%	5.4%	6.2%	6.1%
NA	10.7%	9.4%	9.2%	6.0%	14.2%	12.7%	6.3%	4.6%
	2008	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**
1	32.7%	36.0%	36.4%	35.1%	33.8%	33.9%	32.4%	31.4%
2	18.4%	18.1%	16.8%	17.9%	17.6%	17.6%	17.9%	17.7%
3	21.9%	21.8%	21.0%	22.1%	22.7%	22.0%	22.7%	24.0%
4	12.3%	12.7%	12.7%	13.9%	14.3%	14.5%	14.6%	15.2%
5	7.0%	7.2%	9.7%	7.7%	7.5%	7.7%	7.6%	8.0%
NA	7.8%	4.2%	3.4%	3.4%	4.1%	4.2%	4.8%	3.8%

Source: Ministério do Trabalho e Previdência. Own elaboration.

Table 32: Frequency of work contracts per AP from 2000 to 2015, in the PMM-MICE-treated RAIS "vínculos" database

	2000	2001	2002	2003	2004	2005	2006	2007
1	37.7%	36.9%	36.5%	37.1%	38.6%	40.3%	39.9%	41.7%
2	21.2%	20.9%	20.5%	20.6%	21.5%	19.9%	19.7%	18.1%
3	26.4%	27.1%	27.4%	26.1%	23.1%	22.5%	21.9%	22.0%
4	8.4%	8.7%	9.2%	9.6%	10.1%	11.1%	11.9%	11.8%
5	6.3%	6.4%	6.5%	6.5%	6.6%	6.2%	6.6%	6.4%
NA	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	2008	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**
1	36.2%	37.6%	36.7%	36.4%	35.2%	35.3%	33.9%	32.5%
2	19.9%	18.9%	16.9%	18.5%	18.4%	18.4%	18.9%	18.4%
3	23.3%	22.8%	21.3%	22.8%	23.7%	23.0%	23.9%	25.0%
4	13.1%	13.3%	12.9%	14.3%	14.9%	15.2%	15.4%	15.8%
5	7.5%	7.5%	9.8%	8.0%	7.9%	8.0%	8.0%	8.3%
NA	0.0%	0.0%	2.4%	0.0%	0.0%	0.0%	0.0%	0.0%

Source: Ministério do Trabalho e Previdência. Own elaboration.

Table 33: Comparison of work contracts per AP in the RAIS "vínculos" database (original vs. treated)

	2000	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**
1	4.0%	3.5%	3.3%	2.1%	5.8%	4.7%	2.6%	1.8%
2	2.4%	2.0%	1.9%	1.3%	3.2%	2.7%	1.3%	0.9%
3	2.8%	2.5%	2.5%	1.6%	3.0%	3.0%	1.4%	1.0%
4	0.9%	0.8%	0.8%	0.6%	1.3%	1.5%	0.7%	0.6%
5	0.6%	0.6%	0.6%	0.4%	0.9%	0.8%	0.4%	0.3%
NA	-10.7%	-9.4%	-9.2%	-6.0%	-14.2%	-12.7%	-6.3%	-4.6%

	2008	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**
1	3.6%	1.5%	0.3%	1.3%	1.4%	1.4%	1.5%	1.1%
2	1.5%	0.9%	0.2%	0.6%	0.8%	0.8%	1.0%	0.7%
3	1.5%	1.0%	0.3%	0.8%	1.0%	1.0%	1.2%	1.0%
4	0.7%	0.6%	0.2%	0.5%	0.6%	0.7%	0.8%	0.6%
5	0.4%	0.3%	0.1%	0.3%	0.3%	0.3%	0.4%	0.3%
NA	-7.8%	-4.2%	-1.0%	-3.4%	-4.1%	-4.2%	-4.8%	-3.8%

Source: Ministério do Trabalho e Previdência. Own elaboration.

Table 34: Frequency of work contracts per RP from 2000 to 2015, in the original RAIS "vínculos" database

	2000	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**
1.1	33.7%	33.4%	33.2%	35.0%	32.7%	35.6%	37.3%	39.9%
2.1	12.5%	12.7%	12.6%	12.9%	12.4%	12.0%	12.4%	11.4%
2.2	6.3%	6.2%	5.9%	6.4%	6.0%	5.2%	6.1%	5.8%
3.1	3.9%	3.8%	3.7%	3.9%	3.7%	4.1%	4.3%	4.3%
3.2	4.6%	4.8%	4.8%	4.7%	4.7%	4.0%	4.2%	4.7%
3.3	6.6%	7.6%	7.6%	7.4%	4.0%	3.6%	4.0%	3.8%
3.4	0.9%	0.9%	0.9%	0.9%	1.1%	1.1%	1.1%	1.1%
3.5	3.8%	3.7%	3.8%	3.7%	3.3%	3.1%	3.2%	3.2%
3.6	1.4%	1.5%	1.5%	1.7%	1.5%	1.4%	1.7%	1.7%
3.7	2.3%	2.3%	2.6%	2.3%	1.9%	2.0%	2.0%	2.2%
4.1	3.1%	3.3%	3.3%	3.3%	3.5%	3.8%	4.5%	4.6%
4.2	4.4%	4.6%	5.1%	5.7%	5.3%	5.8%	6.7%	6.6%
5.1	2.1%	2.1%	2.1%	2.2%	2.0%	1.9%	2.4%	2.4%
5.2	2.5%	2.6%	2.7%	2.8%	2.6%	2.4%	2.5%	2.6%
5.3	0.9%	1.0%	1.0%	1.0%	1.0%	1.0%	1.1%	1.0%
5.4	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.2%	0.2%
NA	10.7%	9.4%	9.2%	6.0%	14.2%	12.7%	6.3%	4.6%
	2008	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**
1.1	32.7%	36.0%	36.4%	35.1%	33.8%	33.9%	32.4%	31.4%
2.1	12.5%	12.3%	11.3%	12.1%	11.8%	12.0%	12.7%	12.3%
2.2	5.8%	5.8%	5.5%	5.8%	5.8%	5.6%	5.2%	5.4%
3.1	4.5%	4.5%	4.6%	4.7%	5.0%	4.6%	4.3%	4.3%
3.2	4.8%	5.0%	4.7%	4.8%	5.0%	4.5%	5.0%	5.0%
3.3	3.9%	3.9%	3.6%	3.8%	4.0%	4.2%	4.5%	5.0%
3.4	1.1%	1.0%	1.0%	1.1%	1.0%	1.0%	1.1%	1.3%
3.5	3.4%	3.3%	3.2%	3.5%	3.5%	3.6%	3.5%	3.5%
3.6	1.8%	1.8%	1.7%	1.8%	1.9%	1.9%	2.0%	2.1%
3.7	2.3%	2.3%	2.2%	2.3%	2.3%	2.3%	2.3%	2.7%
4.1	4.9%	5.1%	4.9%	5.3%	5.7%	5.6%	5.7%	5.8%
4.2	7.4%	7.6%	7.8%	8.6%	8.6%	8.9%	8.9%	9.4%
5.1	2.5%	2.5%	2.6%	2.7%	2.7%	2.7%	2.6%	2.8%
5.2	2.9%	2.8%	5.3%	3.3%	3.1%	3.3%	3.1%	3.2%
5.3	1.4%	1.6%	1.6%	1.4%	1.4%	1.4%	1.5%	1.5%
5.4	0.2%	0.2%	0.2%	0.3%	0.3%	0.3%	0.4%	0.4%
NA	7.8%	4.2%	3.4%	3.4%	4.1%	4.2%	4.8%	3.8%

Source: Ministério do Trabalho e Previdência. Own elaboration.

Table 35: Frequency of work contracts per RP from 2000 to 2015, in the PMM-MICE-treated RAIS "vínculos" database

	2000	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**
1.1	37.7%	36.9%	36.5%	37.1%	38.6%	40.3%	39.9%	41.7%
2.1	14.1%	14.0%	13.9%	13.8%	14.5%	13.8%	13.2%	12.0%
2.2	7.1%	6.9%	6.6%	6.8%	7.0%	6.1%	6.5%	6.1%
3.1	4.4%	4.2%	4.1%	4.1%	4.2%	4.8%	4.6%	4.5%
3.2	5.2%	5.3%	5.3%	5.0%	5.3%	4.6%	4.5%	5.0%
3.3	7.4%	8.3%	8.3%	7.8%	4.6%	4.2%	4.3%	4.0%
3.4	1.0%	1.0%	1.0%	1.0%	1.2%	1.3%	1.1%	1.1%
3.5	4.2%	4.1%	4.1%	4.0%	3.8%	3.6%	3.4%	3.4%
3.6	1.6%	1.7%	1.7%	1.8%	1.7%	1.7%	1.8%	1.7%
3.7	2.6%	2.6%	2.9%	2.4%	2.2%	2.3%	2.2%	2.3%
4.1	3.5%	3.6%	3.6%	3.5%	4.1%	4.4%	4.8%	4.9%
4.2	4.9%	5.1%	5.6%	6.1%	6.1%	6.7%	7.1%	7.0%
5.1	2.3%	2.3%	2.3%	2.3%	2.3%	2.2%	2.6%	2.5%
5.2	2.8%	2.9%	3.0%	3.0%	3.0%	2.7%	2.7%	2.7%
5.3	1.0%	1.1%	1.1%	1.1%	1.2%	1.1%	1.1%	1.0%
5.4	0.1%	0.1%	0.1%	0.1%	0.2%	0.2%	0.2%	0.2%
NA	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	2008	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**
1.1	36.1%	37.6%	36.7%	36.3%	35.2%	35.3%	33.9%	32.5%
2.1	13.6%	12.9%	11.4%	12.5%	12.4%	12.6%	13.4%	12.8%
2.2	6.3%	6.1%	5.5%	6.0%	6.0%	5.9%	5.5%	5.6%
3.1	4.8%	4.7%	4.6%	4.9%	5.2%	4.8%	4.5%	4.5%
3.2	5.1%	5.2%	4.8%	5.0%	5.2%	4.6%	5.2%	5.2%
3.3	4.2%	4.0%	3.6%	3.9%	4.1%	4.4%	4.8%	5.2%
3.4	1.2%	1.1%	1.0%	1.1%	1.0%	1.1%	1.1%	1.4%
3.5	3.6%	3.5%	3.2%	3.6%	3.6%	3.8%	3.7%	3.7%
3.6	1.9%	1.9%	1.8%	1.9%	2.0%	1.9%	2.2%	2.2%
3.7	2.6%	2.4%	2.2%	2.4%	2.4%	2.4%	2.4%	2.8%
4.1	5.2%	5.3%	4.9%	5.5%	5.9%	5.9%	6.0%	6.0%
4.2	7.9%	7.9%	8.0%	8.9%	9.0%	9.3%	9.4%	9.8%
5.1	2.6%	2.6%	2.6%	2.8%	2.8%	2.8%	2.8%	2.9%
5.2	3.1%	3.0%	5.4%	3.4%	3.3%	3.4%	3.3%	3.4%
5.3	1.5%	1.7%	1.6%	1.5%	1.5%	1.5%	1.6%	1.6%
5.4	0.2%	0.2%	0.2%	0.3%	0.3%	0.3%	0.4%	0.5%
NA	0.0%	0.0%	2.4%	0.0%	0.0%	0.0%	0.0%	0.0%

Source: Ministério do Trabalho e Previdência. Own elaboration.

Table 36: Comparison of work contracts per RP in the RAIS "vínculos" database (original vs. treated)

	2000	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**
1.1	4.0%	3.4%	3.3%	2.1%	5.8%	4.7%	2.6%	1.8%
2.1	1.6%	1.4%	1.3%	0.9%	2.2%	1.9%	0.9%	0.6%
2.2	0.8%	0.7%	0.6%	0.4%	1.0%	0.8%	0.4%	0.3%
3.1	0.5%	0.4%	0.4%	0.3%	0.5%	0.6%	0.3%	0.2%
3.2	0.5%	0.5%	0.5%	0.3%	0.7%	0.6%	0.3%	0.2%
3.3	0.8%	0.8%	0.7%	0.5%	0.6%	0.6%	0.3%	0.2%
3.4	0.1%	0.1%	0.1%	0.1%	0.2%	0.2%	0.1%	0.1%
3.5	0.4%	0.4%	0.4%	0.2%	0.5%	0.5%	0.2%	0.2%
3.6	0.2%	0.2%	0.2%	0.1%	0.2%	0.2%	0.1%	0.1%
3.7	0.3%	0.3%	0.3%	0.1%	0.3%	0.3%	0.1%	0.1%
4.1	0.4%	0.3%	0.3%	0.2%	0.5%	0.6%	0.3%	0.2%
4.2	0.5%	0.5%	0.5%	0.4%	0.8%	0.9%	0.4%	0.3%
5.1	0.2%	0.2%	0.2%	0.1%	0.3%	0.3%	0.2%	0.1%
5.2	0.3%	0.3%	0.3%	0.2%	0.4%	0.4%	0.2%	0.1%
5.3	0.1%	0.1%	0.1%	0.1%	0.2%	0.1%	0.1%	0.1%
5.4	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
NA	-10.7%	-9.4%	-9.2%	-6.0%	-14.2%	-12.7%	-6.3%	-4.6%

	2008	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**
1.1	3.5%	1.5%	0.3%	1.2%	1.4%	1.4%	1.5%	1.1%
2.1	1.0%	0.6%	0.1%	0.4%	0.5%	0.5%	0.7%	0.5%
2.2	0.5%	0.3%	0.1%	0.2%	0.3%	0.3%	0.3%	0.2%
3.1	0.3%	0.2%	0.1%	0.2%	0.2%	0.2%	0.2%	0.2%
3.2	0.3%	0.2%	0.1%	0.2%	0.2%	0.2%	0.3%	0.2%
3.3	0.3%	0.1%	0.0%	0.1%	0.2%	0.2%	0.2%	0.2%
3.4	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%
3.5	0.2%	0.1%	0.1%	0.1%	0.2%	0.2%	0.2%	0.1%
3.6	0.1%	0.1%	0.0%	0.1%	0.1%	0.1%	0.1%	0.1%
3.7	0.2%	0.1%	0.0%	0.1%	0.1%	0.1%	0.1%	0.1%
4.1	0.3%	0.2%	0.1%	0.2%	0.2%	0.3%	0.3%	0.2%
4.2	0.5%	0.3%	0.1%	0.3%	0.4%	0.4%	0.5%	0.4%
5.1	0.2%	0.1%	0.0%	0.1%	0.1%	0.1%	0.1%	0.1%
5.2	0.2%	0.1%	0.0%	0.1%	0.1%	0.1%	0.2%	0.1%
5.3	0.1%	0.1%	0.0%	0.1%	0.1%	0.1%	0.1%	0.1%
5.4	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
NA	-7.8%	-4.2%	-1.0%	-3.4%	-4.1%	-4.2%	-4.8%	-3.8%

Source: Ministério do Trabalho e Previdência. Own elaboration.

Table 37: Frequency of white/non-white workers (0/1) from 2000 to 2015, in the original RAIS "vínculos" database

	2000	2001	2002	2003	2004	2005	2006	2007
0	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	47.4%	44.9%
1	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	33.3%	33.2%
NA	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	19.4%	21.8%

	2008	2009	2010	2011	2012	2013	2014	2015
0	46.3%	45.4%	42.4%	44.7%	44.2%	43.1%	41.0%	39.8%
1	35.5%	35.8%	34.9%	37.9%	38.7%	39.6%	41.4%	41.9%
NA	18.2%	18.8%	22.7%	17.4%	17.1%	17.3%	17.6%	18.3%

Source: Ministério do Trabalho e Previdência. Own elaboration.

Table 38: Frequency of white/non-white workers (0/1) from 2000 to 2015, in the PMM-MICE-treated RAIS "vínculos" database

	2000	2001	2002	2003	2004	2005	2006	2007
0	60.7%	60.5%	60.3%	60.8%	60.4%	60.6%	60.7%	59.1%
1	39.3%	39.5%	39.7%	39.2%	39.6%	39.4%	39.3%	40.9%
NA	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

	2008	2009	2010	2011	2012	2013	2014	2015
0	58.3%	57.9%	46.6%	56.0%	55.3%	54.1%	52.1%	51.1%
1	41.7%	42.1%	38.1%	44.0%	44.7%	45.9%	47.9%	48.9%
NA	0.0%	0.0%	15.3%	0.0%	0.0%	0.0%	0.0%	0.0%

Source: Ministério do Trabalho e Previdência. Own elaboration.

DiD second-stage results - RP level

Table 39: Dependent variable: Log Average Income (deflated) – second-stage estimation, RP level

	Model 1	Model 2	Model 3
fixest = rel_year = -2	0.020***	0.020***	0.017***
	(0.005)	(0.005)	(0.003)
fixest = rel_year = 0	-0.029**	-0.026*	0.003
	(0.010)	(0.009)	(0.008)
fixest = rel_year = 1	0.005	0.004	0.018*
	(0.013)	(0.011)	(0.008)
fixest = rel_year = 2	0.036	0.032	0.038*
	(0.021)	(0.019)	(0.017)
fixest = rel_year = 3	0.027	0.029+	0.010
	(0.018)	(0.016)	(0.011)
fixest = rel_year = 4	0.017	0.011	-0.032**
	(0.018)	(0.018)	(0.011)
fixest = rel_year = 5	0.010	0.006	-0.016
	(0.022)	(0.022)	(0.016)
fixest = rel_year = 6	0.029	0.028	0.008
	(0.027)	(0.029)	(0.022)
fixest = rel_year = 7	0.147*	0.104**	0.029
	(0.059)	(0.027)	(0.021)
fixest = rel_year = 8	0.034	0.030	0.019
	(0.023)	(0.023)	(0.013)
fixest = rel_year = 9	0.012	0.009	0.020+
	(0.015)	(0.016)	(0.010)
fixest = rel_year = 10	-0.016	-0.021	-0.011
	(0.019)	(0.019)	(0.014)
Num.Obs.	33830503	33339651	33339651
R2	0.001	0.0006	0.0003
R2 Adj.	0.001	0.0006	0.0003

Notes: standard errors clustered by RP

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Source: Own elaboration.

Table 40: Dependent variable: Worker at "Simples Nacional" company (Y/n) – second-stage estimation, RP level

	Model 1	Model 2	Model 3
fixest = rel_year = -2	0.018***	0.018***	0.019***
	(0.002)	(0.002)	(0.002)
fixest = rel_year = 0	0.048***	0.048***	0.047***
	(0.003)	(0.003)	(0.003)
fixest = rel_year = 1	0.022***	0.023***	0.033***
	(0.004)	(0.004)	(0.004)
fixest = rel_year = 2	0.039**	0.040**	0.045***
	(0.012)	(0.012)	(0.008)
fixest = rel_year = 3	0.044**	0.045**	0.053***
	(0.012)	(0.012)	(0.006)
fixest = rel_year = 4	0.043**	0.044**	0.057***
	(0.014)	(0.014)	(0.010)
fixest = rel_year = 5	0.033*	0.034*	0.053***
	(0.013)	(0.013)	(0.009)
fixest = rel_year = 6	0.048	0.050	0.070*
	(0.029)	(0.028)	(0.028)
fixest = rel_year = 7	0.035	0.050	0.080*
	(0.035)	(0.029)	(0.028)
fixest = rel_year = 8	0.065*	0.067*	0.083**
	(0.027)	(0.027)	(0.025)
fixest = rel_year = 9	0.032+	0.034*	0.050***
	(0.016)	(0.016)	(0.010)
fixest = rel_year = 10	0.036*	0.038*	0.049***
	(0.014)	(0.014)	(0.008)
Num.Obs.	34334762	33842121	33842121
R2	0.002	0.003	0.006
R2 Adj.	0.002	0.003	0.006

Notes: standard errors clustered by RP

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Source: Own elaboration.

Table 41: Dependent variable: Worker earning under 2 minimum wages/month (Y/n) – second-stage estimation, RP level

	Model 1	Model 2	Model 3
fixest = rel_year = -2	-0.008*	-0.008*	-0.007*
	(0.003)	(0.003)	(0.003)
fixest = rel_year = 0	0.035***	0.034***	0.023**
	(0.006)	(0.006)	(0.006)
fixest = rel_year = 1	0.011	0.011	0.011
	(0.010)	(0.010)	(0.008)
fixest = rel_year = 2	0.010	0.012	0.013
	(0.011)	(0.010)	(0.011)
fixest = rel_year = 3	0.025*	0.025*	0.030**
	(0.010)	(0.009)	(0.008)
fixest = rel_year = 4	0.044***	0.047***	0.063***
	(0.010)	(0.009)	(0.008)
fixest = rel_year = 5	0.045**	0.047**	0.061***
	(0.015)	(0.015)	(0.014)
fixest = rel_year = 6	0.041*	0.043*	0.052**
	(0.018)	(0.019)	(0.017)
fixest = rel_year = 7	-0.026	-0.0003	0.042*
	(0.042)	(0.019)	(0.017)
fixest = rel_year = 8	0.037**	0.040**	0.047***
	(0.012)	(0.012)	(0.011)
fixest = rel_year = 9	0.055***	0.057***	0.058***
	(0.011)	(0.010)	(0.012)
fixest = rel_year = 10	0.061***	0.064***	0.066***
	(0.011)	(0.010)	(0.011)
Num.Obs.	34334762	33842121	33842121
R2	0.002	0.002	0.003
R2 Adj.	0.002	0.002	0.003

Notes: standard errors clustered by RP

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Source: Own elaboration.

Table 42: Dependent variable: Female worker (Y/n) – second-stage estimation, RP level

	Model 1	Model 2	Model 3
fixest = rel_year = -2	0.0008	0.0008	0.002
	(0.001)	(0.002)	(0.002)
fixest = rel_year = 0	-0.002	-0.002	0.002
	(0.008)	(0.008)	(0.008)
fixest = rel_year = 1	0.009	0.010	0.011
	(0.009)	(0.009)	(0.010)
fixest = rel_year = 2	-0.010	-0.010	-0.012
	(0.011)	(0.011)	(0.011)
fixest = rel_year = 3	-0.008	-0.007	-0.011
	(0.013)	(0.013)	(0.012)
fixest = rel_year = 4	-0.012	-0.011	-0.020
	(0.013)	(0.013)	(0.013)
fixest = rel_year = 5	0.009	0.010	0.0004
	(0.015)	(0.014)	(0.013)
fixest = rel_year = 6	-0.014	-0.013	-0.014
	(0.016)	(0.016)	(0.016)
fixest = rel_year = 7	0.024	-0.005	-0.032+
	(0.040)	(0.016)	(0.016)
fixest = rel_year = 8	-0.008	-0.006	-0.014
	(0.013)	(0.013)	(0.013)
fixest = rel_year = 9	0.003	0.004	-0.003
	(0.014)	(0.014)	(0.013)
fixest = rel_year = 10	0.009	0.011	-0.0008
	(0.011)	(0.012)	(0.010)
Num.Obs.	34334762	33842121	33842121
R2	0.0001	0.00008	0.0002
R2 Adj.	0.0001	0.00008	0.0002

Notes: standard errors clustered by RP

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Source: Own elaboration.

Table 43: Dependent variable: Non-white worker (Y/n) – second-stage estimation, RP level

	Model 1	Model 2	Model 3
fixest = rel_year = -2	-0.0002	-0.0001	0.0004
	(0.001)	(0.001)	(0.0006)
fixest = rel_year = 0	0.009***	0.009***	0.001
	(0.001)	(0.002)	(0.002)
fixest = rel_year = 1	0.005*	0.006*	-0.003**
	(0.002)	(0.002)	(0.001)
fixest = rel_year = 2	0.0007	0.0006	-0.004
	(0.002)	(0.002)	(0.003)
fixest = rel_year = 3	0.023***	0.023***	0.023***
	(0.005)	(0.005)	(0.005)
fixest = rel_year = 4	0.009	0.009	0.016+
	(0.007)	(0.007)	(0.009)
fixest = rel_year = 5	0.018+	0.020*	0.016+
	(0.008)	(0.008)	(0.009)
fixest = rel_year = 6	0.032**	0.032**	0.026*
	(0.010)	(0.010)	(0.011)
fixest = rel_year = 7	0.015	0.019+	0.024*
	(0.011)	(0.011)	(0.011)
fixest = rel_year = 8	0.029**	0.030**	0.024*
	(0.008)	(0.008)	(0.010)
fixest = rel_year = 9	0.028**	0.030**	0.019+
	(0.009)	(0.009)	(0.009)
fixest = rel_year = 10	0.023*	0.026*	0.018+
	(0.009)	(0.009)	(0.010)
Num.Obs.	33842121	33842121	33842121
R2	0.0004	0.0004	0.0003
R2 Adj.	0.0004	0.0004	0.0003

Notes: standard errors clustered by RP

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Source: Own elaboration.

Table 44: Dependent variable: Still at work by end of the year (Y/n) – second-stage estimation, RP level

	Model 1	Model 2	Model 3
fixest = rel_year = -2	-0.003	-0.003	-0.002
	(0.002)	(0.002)	(0.003)
fixest = rel_year = 0	0.008*	0.009**	0.002
	(0.003)	(0.003)	(0.002)
fixest = rel_year = 1	0.017*	0.016*	-0.003
	(0.007)	(0.006)	(0.005)
fixest = rel_year = 2	0.021***	0.019**	0.008
	(0.005)	(0.005)	(0.007)
fixest = rel_year = 3	0.026***	0.025***	0.010*
	(0.005)	(0.005)	(0.004)
fixest = rel_year = 4	0.021***	0.017**	0.006
	(0.004)	(0.004)	(0.004)
fixest = rel_year = 5	0.037***	0.033***	0.005
	(0.005)	(0.004)	(0.006)
fixest = rel_year = 6	0.015*	0.013+	-0.004
	(0.007)	(0.006)	(0.006)
fixest = rel_year = 7	0.103***	0.117***	0.057***
	(0.024)	(0.007)	(0.007)
fixest = rel_year = 8	0.010	0.006	-0.009
	(0.007)	(0.006)	(0.006)
fixest = rel_year = 9	0.0005	-0.004	-0.014+
	(0.005)	(0.005)	(0.007)
fixest = rel_year = 10	0.007	0.002	-0.011+
	(0.005)	(0.005)	(0.005)
Num.Obs.	34334762	33842121	33842121
R2	0.002	0.002	0.0005
R2 Adj.	0.002	0.002	0.0005

Notes: standard errors clustered by RP

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Source: Own elaboration.

DiD first-stage results - AP & RP level

Table 45: Dependent variable: Log Average Income (deflated) – first-stage estimation, AP level

	Model 1	Model 2
race_id	-0.342***	-0.118***
	(0.013)	(0.007)
gender_id	-0.147***	-0.306***
	(0.013)	(0.006)
under_40	-0.376***	-0.279***
	(0.025)	(0.019)
public_work		0.177**
		(0.022)
educ_id2		0.073
		(0.049)
educ_id3		-0.015+
		(0.006)
educ_id4		-0.021
		(0.028)
educ_id5		0.081
		(0.068)
educ_id6		0.172+
		(0.080)
educ_id7		0.319**
		(0.053)
educ_id8		0.664***
		(0.019)
educ_id9		1.178***
		(0.016)
educ_id10		1.646***
		(0.176)
educ_id11		1.707***
		(0.027)
tamanho_estabelecimento01		-0.274**
		(0.033)
tamanho_estabelecimento02		-0.431***
		(0.018)
tamanho_estabelecimento03		-0.422***
		(0.005)
tamanho_estabelecimento04		-0.385***
		(0.026)
tamanho_estabelecimento05		-0.331**
		(0.053)
tamanho_estabelecimento06		-0.240*
		(0.057)
tamanho_estabelecimento07		-0.121*
		(0.037)
tamanho_estabelecimento08		-0.077***
		(0.007)
tamanho_estabelecimento010		-0.129+
		(0.048)
Num.Obs.	20882644	20882644
R2	0.128	0.412
R2 Adj.	0.128	0.412
R2 Within	0.090	0.387

Notes: standard errors clustered by AP

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Source: Own elaboration.

Table 46: Dependent variable: Log Average Income (deflated) – first-stage estimation, RP level

	Model 1	Model 2
race_id	-0.328***	-0.115***
	(0.021)	(0.007)
gender_id	-0.148***	-0.303***
	(0.009)	(0.007)
under_40	-0.364***	-0.271***
	(0.029)	(0.019)
public_work		0.190***
		(0.029)
educ_id2		0.062
		(0.043)
educ_id3		-0.006
		(0.015)
educ_id4		-0.007
		(0.026)
educ_id5		0.082
		(0.053)
educ_id6		0.160*
		(0.064)
educ_id7		0.311***
		(0.043)
educ_id8		0.667***
		(0.023)
educ_id9		1.173***
		(0.028)
educ_id10		1.618***
		(0.143)
educ_id11		1.664***
		(0.067)
tamanho_estabelecimento1		-0.273***
		(0.033)
tamanho_estabelecimento2		-0.433***
		(0.018)
tamanho_estabelecimento3		-0.416***
		(0.020)
tamanho_estabelecimento4		-0.375***
		(0.031)
tamanho_estabelecimento5		-0.331***
		(0.043)
tamanho_estabelecimento6		-0.239***
		(0.043)
tamanho_estabelecimento7		-0.127***
		(0.028)
tamanho_estabelecimento8		-0.074***
		(0.009)
tamanho_estabelecimento10		-0.113*
		(0.045)
Num.Obs.	25026393	25026393
R2	0.135	0.419
R2 Adj.	0.135	0.419
R2 Within	0.087	0.386

Notes: standard errors clustered by RP
+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Source: Own elaboration.

Table 47: Dependent variable: Worker at "Simples Nacional" company (Y/n) – first-stage estimation, AP level

	Model 1	Model 2
race_id	0.004	-0.009*
	(0.003)	(0.003)
gender_id	0.030	0.042+
	(0.022)	(0.015)
under_40	0.049**	0.018*
	(0.008)	(0.005)
public_work		0.021**
		(0.002)
educ_id2		-0.023
		(0.030)
educ_id3		0.059**
		(0.013)
educ_id4		0.072***
		(0.004)
educ_id5		0.129**
		(0.021)
educ_id6		0.099**
		(0.013)
educ_id7		0.056**
		(0.010)
educ_id8		0.007
		(0.003)
educ_id9		0.013*
		(0.004)
educ_id10		0.005
		(0.004)
educ_id11		0.006
		(0.006)
tamanho_estabelecimento1		-0.423**
		(0.055)
tamanho_estabelecimento2		-0.308*
		(0.098)
tamanho_estabelecimento3		-0.230+
		(0.090)
tamanho_estabelecimento4		-0.275*
		(0.083)
tamanho_estabelecimento5		-0.373**
		(0.073)
tamanho_estabelecimento6		-0.544**
		(0.085)
tamanho_estabelecimento7		-0.621**
		(0.095)
tamanho_estabelecimento8		-0.642**
		(0.091)
tamanho_estabelecimento9		-0.644**
		(0.093)
tamanho_estabelecimento10		-0.637**
		(0.096)
Num.Obs.	21147314	21147314
R2	0.030	0.233
R2 Adj.	0.030	0.233
R2 Within	0.007	0.216

Notes: standard errors clustered by AP

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Source: Own elaboration.

Table 48: Dependent variable: Worker at "Simples Nacional" company (Y/n) – first-stage estimation, RP level

	Model 1	Model 2
race_id	0.001	-0.010***
	(0.003)	(0.002)
gender_id	0.037+	0.045**
	(0.018)	(0.012)
under_40	0.052***	0.020***
	(0.007)	(0.004)
public_work		0.021***
		(0.004)
educ_id2		-0.016
		(0.023)
educ_id3		0.059*
		(0.021)
educ_id4		0.075**
		(0.020)
educ_id5		0.138***
		(0.031)
educ_id6		0.107***
		(0.025)
educ_id7		0.063*
		(0.023)
educ_id8		0.011
		(0.018)
educ_id9		0.018
		(0.017)
educ_id10		0.006
		(0.019)
educ_id11		0.013
		(0.021)
tamanho_estabelecimento1		1.412
		(3.954)
tamanho_estabelecimento2		1.535
		(3.935)
tamanho_estabelecimento3		1.614
		(3.938)
tamanho_estabelecimento4		1.561
		(3.938)
tamanho_estabelecimento5		1.461
		(3.942)
tamanho_estabelecimento6		1.291
		(3.936)
tamanho_estabelecimento7		1.206
		(3.926)
tamanho_estabelecimento8		1.183
		(3.927)
tamanho_estabelecimento9		1.177
		(3.923)
tamanho_estabelecimento10		1.184
		(3.920)
Num.Obs.	25372709	25372709
R2	0.036	0.245
R2 Adj.	0.036	0.245
R2 Within	0.008	0.223

Notes: standard errors clustered by RP

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Source: Own elaboration.

Table 49: Dependent variable: Worker earning under 2 minimum wages/month (Y/n) – first-stage estimation, AP level

	Model 1	Model 2
race_id	0.126***	0.043***
	(0.004)	(0.002)
gender_id	0.074**	0.142***
	(0.011)	(0.010)
under_40	0.151***	0.103***
	(0.006)	(0.004)
public_work		-0.156**
		(0.022)
educ_id2		-0.038+
		(0.015)
educ_id3		-0.022
		(0.025)
educ_id4		0.025
		(0.040)
educ_id5		-0.036
		(0.074)
educ_id6		-0.052
		(0.070)
educ_id7		-0.143*
		(0.051)
educ_id8		-0.349***
		(0.036)
educ_id9		-0.409***
		(0.028)
educ_id10		-0.503***
		(0.057)
educ_id11		-0.504***
		(0.038)
tamanho_estabelecimento1		-0.397**
		(0.048)
tamanho_estabelecimento2		-0.310***
		(0.005)
tamanho_estabelecimento3		-0.297***
		(0.020)
tamanho_estabelecimento4		-0.317**
		(0.037)
tamanho_estabelecimento5		-0.349**
		(0.055)
tamanho_estabelecimento6		-0.410**
		(0.051)
tamanho_estabelecimento7		-0.475***
		(0.036)
tamanho_estabelecimento8		-0.496***
		(0.027)
tamanho_estabelecimento9		-0.522***
		(0.018)
tamanho_estabelecimento10		-0.457***
		(0.017)
Num.Obs.	21147314	21147314
R2	0.087	0.250
R2 Adj.	0.087	0.250
R2 Within	0.049	0.219

Notes: standard errors clustered by AP

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Source: Own elaboration.

Table 50: Dependent variable: Worker earning under 2 minimum wages/month (Y/n) – first-stage estimation, RP level

	Model 1	Model 2
race_id	0.123***	0.042***
	(0.005)	(0.001)
gender_id	0.081***	0.146***
	(0.012)	(0.009)
under_40	0.150***	0.105***
	(0.007)	(0.003)
public_work		-0.156***
		(0.016)
educ_id2		-0.038*
		(0.014)
educ_id3		-0.025
		(0.024)
educ_id4		0.015
		(0.030)
educ_id5		-0.033
		(0.056)
educ_id6		-0.048
		(0.053)
educ_id7		-0.140**
		(0.039)
educ_id8		-0.355***
		(0.030)
educ_id9		-0.416***
		(0.026)
educ_id10		-0.506***
		(0.049)
educ_id11		-0.488***
		(0.042)
tamanho_estabelecimento1		0.309
		(0.429)
tamanho_estabelecimento2		0.405
		(0.431)
tamanho_estabelecimento3		0.408
		(0.426)
tamanho_estabelecimento4		0.382
		(0.425)
tamanho_estabelecimento5		0.353
		(0.427)
tamanho_estabelecimento6		0.292
		(0.428)
tamanho_estabelecimento7		0.230
		(0.429)
tamanho_estabelecimento8		0.206
		(0.427)
tamanho_estabelecimento9		0.173
		(0.416)
tamanho_estabelecimento10		0.234
		(0.419)
Num.Obs.	25372709	25372709
R2	0.100	0.261
R2 Adj.	0.100	0.261
R2 Within	0.048	0.218

Notes: standard errors clustered by RP

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Source: Own elaboration.

Table 51: Dependent variable: Female worker (Y/n) – first-stage estimation, AP level

	Model 1	Model 2
race_id	-0.051***	-0.0005
	(0.006)	(0.002)
under_40	-0.002	0.003
	(0.020)	(0.004)
public_work		0.108**
		(0.016)
educ_id2		-0.036+
		(0.015)
educ_id3		0.010
		(0.011)
educ_id4		0.055*
		(0.017)
educ_id5		0.045
		(0.055)
educ_id6		0.131*
		(0.040)
educ_id7		0.255**
		(0.049)
educ_id8		0.314**
		(0.038)
educ_id9		0.290**
		(0.056)
educ_id10		0.245*
		(0.054)
educ_id11		0.234*
		(0.068)
tamanho_estabelecimento1		0.437**
		(0.052)
tamanho_estabelecimento2		0.446***
		(0.030)
tamanho_estabelecimento3		0.451***
		(0.038)
tamanho_estabelecimento4		0.458***
		(0.046)
tamanho_estabelecimento5		0.425***
		(0.045)
tamanho_estabelecimento6		0.392**
		(0.049)
tamanho_estabelecimento7		0.365**
		(0.049)
tamanho_estabelecimento8		0.365***
		(0.034)
tamanho_estabelecimento9		0.341***
		(0.036)
tamanho_estabelecimento10		0.412***
		(0.025)
Num.Obs.	21147314	21147314
R2	0.005	0.067
R2 Adj.	0.005	0.067
R2 Within	0.003	0.065

Notes: standard errors clustered by AP

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Source: Own elaboration.

Table 52: Dependent variable: Female worker (Y/n) – first-stage estimation, RP level

	Model 1	Model 2
race_id	-0.053***	-0.002
	(0.004)	(0.001)
under_40	0.006	0.006
	(0.017)	(0.005)
public_work		0.108***
		(0.015)
educ_id2		-0.027+
		(0.014)
educ_id3		0.017
		(0.014)
educ_id4		0.062*
		(0.024)
educ_id5		0.061
		(0.047)
educ_id6		0.142**
		(0.037)
educ_id7		0.265***
		(0.041)
educ_id8		0.325***
		(0.033)
educ_id9		0.306***
		(0.049)
educ_id10		0.264***
		(0.049)
educ_id11		0.256***
		(0.059)
tamanho_estabelecimento1		0.161
		(1.018)
tamanho_estabelecimento2		0.178
		(1.023)
tamanho_estabelecimento3		0.186
		(1.019)
tamanho_estabelecimento4		0.186
		(1.017)
tamanho_estabelecimento5		0.149
		(1.018)
tamanho_estabelecimento6		0.115
		(1.017)
tamanho_estabelecimento7		0.089
		(1.016)
tamanho_estabelecimento8		0.087
		(1.025)
tamanho_estabelecimento9		0.063
		(1.030)
tamanho_estabelecimento10		0.134
		(1.039)
Num.Obs.	25372709	25372709
R2	0.005	0.068
R2 Adj.	0.005	0.068
R2 Within	0.003	0.065

Notes: standard errors clustered by RP

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Source: Own elaboration.

Table 53: Dependent variable: Non-white worker (Y/n) – first-stage estimation, AP level

	Model 1	Model 2
gender_id	-0.050**	-0.0004
	(0.006)	(0.002)
under_40	0.048**	0.036**
	(0.010)	(0.005)
public_work		-0.074***
		(0.007)
educ_id2		-0.009
		(0.023)
educ_id3		-0.017
		(0.021)
educ_id4		0.009
		(0.021)
educ_id5		-0.056*
		(0.017)
educ_id6		-0.044*
		(0.016)
educ_id7		-0.128**
		(0.018)
educ_id8		-0.258***
		(0.022)
educ_id9		-0.344***
		(0.024)
educ_id10		-0.424***
		(0.006)
educ_id11		-0.440***
		(0.034)
tamanho_estabelecimento1		0.068
		(0.040)
tamanho_estabelecimento2		0.056
		(0.041)
tamanho_estabelecimento3		0.051
		(0.041)
tamanho_estabelecimento4		0.052
		(0.045)
tamanho_estabelecimento5		0.051
		(0.044)
tamanho_estabelecimento6		0.073
		(0.044)
tamanho_estabelecimento7		0.086
		(0.054)
tamanho_estabelecimento8		0.102+
		(0.045)
tamanho_estabelecimento9		0.110+
		(0.042)
tamanho_estabelecimento10		0.140*
		(0.040)
Num.Obs.	21147314	21147314
R2	0.007	0.071
R2 Adj.	0.007	0.071
R2 Within	0.005	0.069

Notes: standard errors clustered by AP

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Source: Own elaboration.

Table 54: Dependent variable: Non-white worker (Y/n) – first-stage estimation, RP level

	Model 1	Model 2
gender_id	-0.052***	-0.002
	(0.005)	(0.001)
under_40	0.047***	0.037***
	(0.008)	(0.004)
public_work		-0.074***
		(0.008)
educ_id2		-0.003
		(0.015)
educ_id3		-0.017
		(0.016)
educ_id4		0.007
		(0.014)
educ_id5		-0.057***
		(0.012)
educ_id6		-0.047***
		(0.011)
educ_id7		-0.128***
		(0.012)
educ_id8		-0.258***
		(0.015)
educ_id9		-0.343***
		(0.016)
educ_id10		-0.429***
		(0.006)
educ_id11		-0.446***
		(0.024)
tamanho_estabelecimento1		-0.312
		(2.440)
tamanho_estabelecimento2		-0.324
		(2.439)
tamanho_estabelecimento3		-0.328
		(2.438)
tamanho_estabelecimento4		-0.325
		(2.437)
tamanho_estabelecimento5		-0.325
		(2.438)
tamanho_estabelecimento6		-0.306
		(2.437)
tamanho_estabelecimento7		-0.288
		(2.434)
tamanho_estabelecimento8		-0.272
		(2.438)
tamanho_estabelecimento9		-0.266
		(2.443)
tamanho_estabelecimento10		-0.239
		(2.447)
Num.Obs.	25372709	25372709
R2	0.010	0.070
R2 Adj.	0.010	0.070
R2 Within	0.005	0.066

Notes: standard errors clustered by RP

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Source: Own elaboration.

Table 55: Dependent variable: Still at work by end of the year (Y/n) – first-stage estimation, AP level

	Model 1	Model 2
race_id	-0.037*	-0.018**
	(0.010)	(0.003)
gender_id	0.016	-0.010*
	(0.012)	(0.003)
under_40	-0.137***	-0.079***
	(0.007)	(0.007)
public_work		0.187***
		(0.014)
educ_id2		-0.006
		(0.010)
educ_id3		0.009
		(0.013)
educ_id4		-0.020
		(0.016)
educ_id5		-0.012
		(0.026)
educ_id6		-0.017
		(0.031)
educ_id7		-0.028
		(0.027)
educ_id8		-0.014
		(0.029)
educ_id9		0.034
		(0.026)
educ_id10		0.081+
		(0.033)
educ_id11		0.082*
		(0.025)
tamanho_estabelecimento1		-0.605***
		(0.022)
tamanho_estabelecimento2		0.064***
		(0.003)
tamanho_estabelecimento3		0.082*
		(0.020)
tamanho_estabelecimento4		0.078*
		(0.021)
tamanho_estabelecimento5		0.081*
		(0.018)
tamanho_estabelecimento6		0.063**
		(0.010)
tamanho_estabelecimento7		0.065**
		(0.008)
tamanho_estabelecimento8		0.037+
		(0.015)
tamanho_estabelecimento9		0.104***
		(0.010)
tamanho_estabelecimento10		0.157***
		(0.007)
Num.Obs.	21147314	21147314
R2	0.040	0.140
R2 Adj.	0.040	0.140
R2 Within	0.027	0.128

Notes: standard errors clustered by AP

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Source: Own elaboration.

Table 56: Dependent variable: Still at work by end of the year (Y/n) – first-stage estimation, RP level

	Model 1	Model 2
race_id	-0.037***	-0.018***
	(0.007)	(0.002)
gender_id	0.016+	-0.007
	(0.008)	(0.004)
under_40	-0.133***	-0.081***
	(0.009)	(0.005)
public_work		0.191***
		(0.012)
educ_id2		-0.007
		(0.008)
educ_id3		0.006
		(0.009)
educ_id4		-0.023+
		(0.012)
educ_id5		-0.017
		(0.018)
educ_id6		-0.023
		(0.022)
educ_id7		-0.031+
		(0.017)
educ_id8		-0.015
		(0.019)
educ_id9		0.034+
		(0.017)
educ_id10		0.082**
		(0.022)
educ_id11		0.068**
		(0.022)
tamanho_estabelecimento1		-0.497
		(1.005)
tamanho_estabelecimento2		0.167
		(1.011)
tamanho_estabelecimento3		0.183
		(1.009)
tamanho_estabelecimento4		0.182
		(1.007)
tamanho_estabelecimento5		0.184
		(1.004)
tamanho_estabelecimento6		0.168
		(1.004)
tamanho_estabelecimento7		0.169
		(1.005)
tamanho_estabelecimento8		0.140
		(1.011)
tamanho_estabelecimento9		0.200
		(1.016)
tamanho_estabelecimento10		0.251
		(1.018)
Num.Obs.	25372709	25372709
R2	0.041	0.133
R2 Adj.	0.041	0.133
R2 Within	0.025	0.118

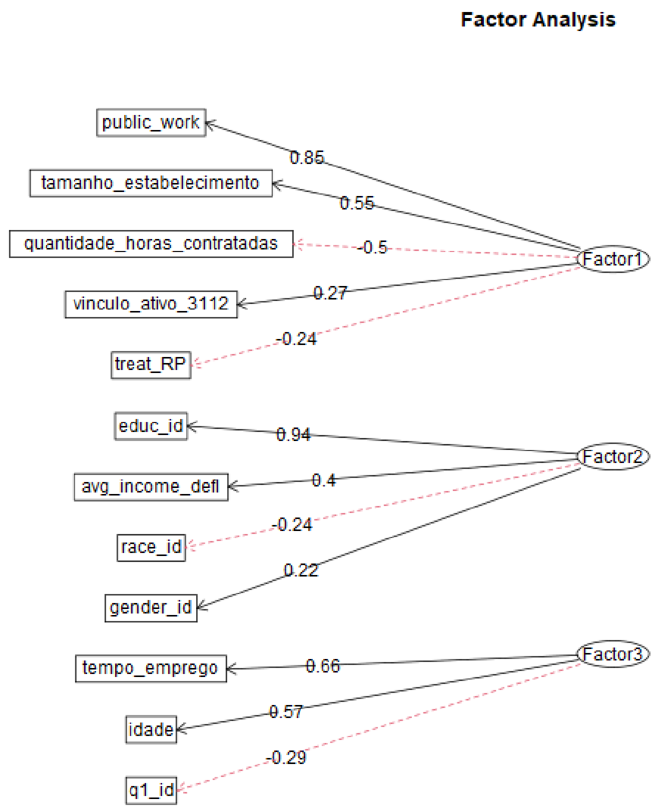
Notes: standard errors clustered by RP

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Source: Own elaboration.

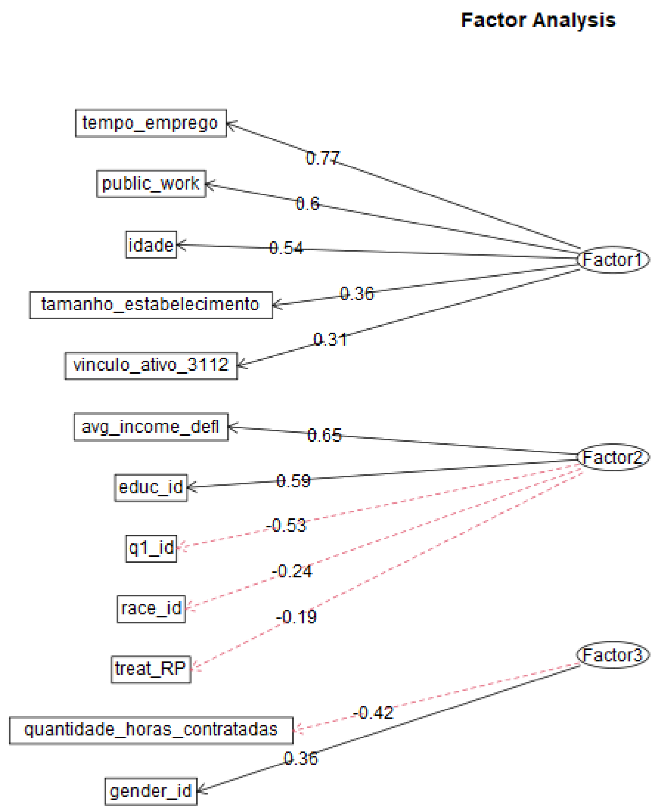
PCA results - RAIS “vínculos” from 2000 to 2015

Figure 43: Factor decomposition for RP treatment in 2000, in RAIS "vínculos" database



Source: Own elaboration.

Figure 44: Factor decomposition for RP treatment in 2015, in RAIS "vínculos" database



Source: Own elaboration.

Table 57: Factor loadings for RAIS "vínculos" at the AP level, 2000 to 2005

	2000			2001		
	Factor1	*Factor2*	*Factor3*	*Factor1*	*Factor2*	*Factor3*
avg_income_defl	0.025	0.427	0.369	0.038	0.393	0.325
vinculo_ativo_3112	0.284	0.041	0.177	0.287	0.032	0.176
tempo_emprego	0.482	0.087	0.634	0.517	0.077	0.634
quantidade_horas_contratadas	-0.513	-0.23	-0.065	-0.482	-0.253	-0.073
idade	0.206	-0.016	0.555	0.203	-0.022	0.586
tamanho_estabelecimento	0.554	0.115	0.216	0.561	0.134	0.163
race_id	-0.064	-0.252	-0.047	-0.055	-0.256	-0.051
educ_id	0.15	0.9	-0.049	0.114	0.888	-0.025
gender_id	0.199	0.214	-0.157	0.141	0.218	-0.108
treat_AP	-0.159	-0.109	-0.109	-0.155	-0.107	-0.104
public_work	0.832	0.09	0.189	0.848	0.103	0.173
q1_id	-0.168	-0.274	-0.29	-0.21	-0.294	-0.274
	2002			**2003**		
	Factor1	*Factor2*	*Factor3*	*Factor1*	*Factor2*	*Factor3*
avg_income_defl	0.052	0.424	0.337	0.046	0.514	0.244
vinculo_ativo_3112	0.276	0.042	0.195	0.273	0.063	0.162
tempo_emprego	0.481	0.085	0.653	0.515	0.212	0.58
quantidade_horas_contratadas	-0.479	-0.243	-0.076	-0.457	-0.268	0.014
idade	0.182	-0.003	0.589	0.208	0.098	0.592
tamanho_estabelecimento	0.569	0.137	0.19	0.565	0.189	0.122
race_id	-0.052	-0.271	-0.054	-0.07	-0.281	-0.006
educ_id	0.137	0.856	-0.062	0.225	0.783	-0.209
gender_id	0.154	0.199	-0.125	0.147	0.128	-0.163
treat_AP	-0.14	-0.117	-0.115	-0.156	-0.193	-0.067
public_work	0.829	0.09	0.211	0.837	0.132	0.136
q1_id	-0.216	-0.324	-0.3	-0.241	-0.419	-0.215
	2004			**2005**		
	Factor1	*Factor2*	*Factor3*	*Factor1*	*Factor2*	*Factor3*
avg_income_defl	0.04	0.531	0.263	0.209	0.573	-0.038
vinculo_ativo_3112	0.254	0.058	0.197	0.299	0.043	0.109
tempo_emprego	0.463	0.178	0.635	0.765	0.202	0.074
quantidade_horas_contratadas	-0.485	-0.207	-0.043	-0.195	-0.19	-0.353
idade	0.151	0.09	0.606	0.597	0.141	-0.162
tamanho_estabelecimento	0.537	0.18	0.179	0.416	0.149	0.372
race_id	-0.077	-0.281	-0.014	-0.027	-0.279	-0.108
educ_id	0.311	0.746	-0.197	-0.051	0.661	0.482
gender_id	0.19	0.109	-0.155	-0.029	0.035	0.298
treat_AP	-0.279	-0.189	-0.094	-0.232	-0.139	-0.254
public_work	0.815	0.08	0.245	0.639	0.054	0.551
q1_id	-0.213	-0.42	-0.245	-0.294	-0.475	-0.088

Source: Own elaboration.

Table 58: Factor loadings for RAIS "vínculos" at the AP level, 2006 to 2011

	2006			**2007**		
	Factor1	*Factor2*	*Factor3*	*Factor1*	*Factor2*	*Factor3*
avg_income_defl	0.041	0.29	0.493	0.093	0.29	0.491
vinculo_ativo_3112	0.198	0.25	0.05	0.224	0.219	0.05
tempo_emprego	0.337	0.703	0.144	0.389	0.689	0.107
quantidade_horas_contratadas	-0.401	-0.075	-0.255	-0.442	-0.051	-0.186
idade	0.058	0.64	0.064	0.088	0.644	0.065
tamanho_estabelecimento	0.501	0.25	0.143	0.527	0.168	0.116
race_id	-0.069	-0.042	-0.27	-0.044	-0.043	-0.269
educ_id	0.294	-0.168	0.792	0.235	-0.219	0.751
gender_id	0.206	-0.113	0.123	0.197	-0.129	0.134
treat_AP	-0.321	-0.144	-0.131	-0.334	-0.099	-0.125
public_work	0.823	0.349	0.034	0.785	0.343	0.022
q1_id	-0.187	-0.311	-0.428	-0.247	-0.297	-0.418
	2008			**2009**		
	Factor1	*Factor2*	*Factor3*	*Factor1*	*Factor2*	*Factor3*
avg_income_defl	0.245	0.536	0.007	0.051	0.293	0.503
vinculo_ativo_3112	0.299	0.067	0.112	0.211	0.231	0.069
tempo_emprego	0.756	0.171	0.127	0.341	0.655	0.139
quantidade_horas_contratadas	-0.176	-0.19	-0.414	-0.43	-0.049	-0.235
idade	0.61	0.137	-0.143	0.072	0.626	0.076
tamanho_estabelecimento	0.361	0.102	0.386	0.482	0.21	0.093
race_id	-0.04	-0.262	-0.071	-0.066	-0.042	-0.267
educ_id	-0.148	0.692	0.455	0.315	-0.216	0.769
gender_id	-0.056	0.059	0.265	0.218	-0.145	0.112
treat_AP	-0.171	-0.16	-0.211	-0.321	-0.11	-0.145
public_work	0.604	0.003	0.613	0.802	0.327	0.046
q1_id	-0.331	-0.463	-0.13	-0.186	-0.327	-0.446
	2010			**2011**		
	Factor1	*Factor2*	*Factor3*	*Factor1*	*Factor2*	*Factor3*
avg_income_defl	0.175	0.56	0.123	0.159	0.602	0.087
vinculo_ativo_3112	0.267	0.03	0.008	0.307	0.134	0.075
tempo_emprego	0.821	0.077	0.064	0.66	0.268	0.328
quantidade_horas_contratadas	-0.062	-0.073	-0.333	-0.39	-0.173	0.285
idade	0.478	0.118	-0.137	0.378	0.204	0.446
tamanho_estabelecimento	0.021	0.233	0.03	0.483	0.117	-0.123
race_id	-0.07	-0.131	-0.194	-0.058	-0.264	0.045
educ_id	-0.136	0.396	0.754	0.104	0.622	-0.532
gender_id	-0.028	-0.172	0.388	0.089	0.019	-0.269
treat_AP	-0.002	-0.137	-0.152	-0.325	-0.131	0.125
public_work	0.167	0.047	0.075	0.822	0.079	-0.161
q1_id	-0.185	-0.64	-0.069	-0.275	-0.521	-0.05

Source: Own elaboration.

Table 59: Factor loadings for RAIS "vínculos" at the AP level, 2012 to 2015

	2012			2013		
	Factor1	*Factor2*	*Factor3*	*Factor1*	*Factor2*	*Factor3*
avg_income_defl	0.205	0.642	-0.077	0.211	0.668	-0.097
vinculo_ativo_3112	0.321	0.115	0.069	0.319	0.107	0.084
tempo_emprego	0.755	0.198	-0.033	0.759	0.189	-0.005
quantidade_horas_contratadas	-0.294	-0.111	-0.507	-0.249	-0.135	-0.509
idade	0.523	0.124	-0.206	0.528	0.093	-0.174
tamanho_estabelecimento	0.391	0.134	0.26	0.368	0.179	0.265
race_id	-0.058	-0.24	-0.086	-0.062	-0.233	-0.075
educ_id	-0.044	0.6	0.558	-0.054	0.587	0.516
gender_id	-0.03	-0.005	0.325	-0.038	-0.025	0.35
treat_AP	-0.269	-0.15	-0.214	-0.247	-0.182	-0.203
public_work	0.679	0.081	0.424	0.638	0.138	0.411
q1_id	-0.307	-0.534	-0.019	-0.298	-0.528	-0.025
	2014			**2015**		
	Factor1	*Factor2*	*Factor3*	*Factor1*	*Factor2*	*Factor3*
avg_income_defl	0.213	0.659	-0.08	0.226	0.633	-0.053
vinculo_ativo_3112	0.327	0.115	0.058	0.315	0.125	0.059
tempo_emprego	0.752	0.206	-0.016	0.746	0.222	-0.051
quantidade_horas_contratadas	-0.191	-0.137	-0.413	-0.191	-0.141	-0.381
idade	0.511	0.101	-0.189	0.51	0.112	-0.225
tamanho_estabelecimento	0.37	0.169	0.232	0.395	0.164	0.254
race_id	-0.085	-0.236	-0.081	-0.094	-0.254	-0.066
educ_id	-0.038	0.576	0.545	-0.043	0.591	0.519
gender_id	-0.036	-0.047	0.363	-0.027	-0.026	0.333
treat_AP	-0.246	-0.183	-0.197	-0.28	-0.172	-0.197
public_work	0.638	0.162	0.353	0.658	0.145	0.375
q1_id	-0.294	-0.523	-0.016	-0.298	-0.508	-0.028

Source: Own elaboration.

Table 60: Factor loadings for RAIS "vínculos" at the RP level, 2000 to 2005

	2000			2001		
	Factor1	*Factor2*	*Factor3*	*Factor1*	*Factor2*	*Factor3*
avg_income_defl	0.035	0.401	0.36	0.046	0.378	0.317
vinculo_ativo_3112	0.271	0.041	0.191	0.28	0.032	0.185
tempo_emprego	0.463	0.076	0.655	0.504	0.073	0.643
quantidade_horas_contratadas	-0.5	-0.223	-0.089	-0.477	-0.243	-0.085
idade	0.187	-0.022	0.567	0.19	-0.021	0.595
tamanho_estabelecimento	0.547	0.103	0.228	0.558	0.123	0.174
race_id	-0.066	-0.24	-0.054	-0.059	-0.246	-0.053
educ_id	0.151	0.936	-0.03	0.121	0.917	-0.023
gender_id	0.193	0.22	-0.134	0.141	0.221	-0.098
treat_RP	-0.243	-0.107	-0.083	-0.244	-0.112	-0.072
public_work	0.85	0.079	0.199	0.858	0.092	0.182
q1_id	-0.169	-0.255	-0.291	-0.211	-0.281	-0.275
	2002			**2003**		
	Factor1	*Factor2*	*Factor3*	*Factor1*	*Factor2*	*Factor3*
avg_income_defl	0.068	0.407	0.333	0.052	0.473	0.278
vinculo_ativo_3112	0.272	0.04	0.198	0.257	0.051	0.185
tempo_emprego	0.484	0.075	0.651	0.481	0.163	0.62
quantidade_horas_contratadas	-0.476	-0.234	-0.079	-0.451	-0.259	-0.037
idade	0.183	-0.007	0.593	0.172	0.057	0.614
tamanho_estabelecimento	0.567	0.124	0.189	0.555	0.163	0.171
race_id	-0.056	-0.263	-0.057	-0.075	-0.271	-0.034
educ_id	0.142	0.88	-0.055	0.244	0.822	-0.143
gender_id	0.15	0.205	-0.117	0.15	0.15	-0.139
treat_RP	-0.242	-0.118	-0.083	-0.241	-0.142	-0.064
public_work	0.845	0.077	0.2	0.85	0.099	0.183
q1_id	-0.224	-0.309	-0.299	-0.238	-0.384	-0.26
	2004			**2005**		
	Factor1	*Factor2*	*Factor3*	*Factor1*	*Factor2*	*Factor3*
avg_income_defl	0.024	0.529	0.266	0.208	0.577	-0.036
vinculo_ativo_3112	0.24	0.057	0.215	0.298	0.042	0.114
tempo_emprego	0.417	0.177	0.666	0.765	0.203	0.085
quantidade_horas_contratadas	-0.48	-0.205	-0.078	-0.191	-0.187	-0.36
idade	0.108	0.089	0.615	0.598	0.142	-0.153
tamanho_estabelecimento	0.523	0.176	0.218	0.411	0.149	0.377
race_id	-0.077	-0.28	-0.021	-0.025	-0.278	-0.11
educ_id	0.328	0.748	-0.174	-0.057	0.657	0.484
gender_id	0.202	0.11	-0.141	-0.032	0.031	0.3
treat_RP	-0.203	-0.166	-0.076	-0.155	-0.128	-0.198
public_work	0.796	0.076	0.302	0.631	0.056	0.555
q1_id	-0.196	-0.417	-0.261	-0.291	-0.477	-0.092

Source: Own elaboration.

Table 61: Factor loadings for RAIS "vínculos" at the RP level, 2006 to 2011

	2006			2007		
	Factor1	*Factor2*	*Factor3*	*Factor1*	*Factor2*	*Factor3*
avg_income_defl	0.283	0.038	0.498	0.092	0.295	0.484
vinculo_ativo_3112	0.261	0.182	0.05	0.212	0.227	0.05
tempo_emprego	0.721	0.297	0.148	0.37	0.699	0.102
quantidade_horas_contratadas	-0.096	-0.402	-0.246	-0.435	-0.067	-0.186
idade	0.642	0.02	0.074	0.068	0.649	0.061
tamanho_estabelecimento	0.278	0.486	0.135	0.511	0.187	0.114
race_id	-0.041	-0.074	-0.269	-0.047	-0.045	-0.267
educ_id	-0.162	0.325	0.778	0.245	-0.209	0.758
gender_id	-0.102	0.215	0.115	0.198	-0.122	0.138
treat_RP	-0.104	-0.233	-0.125	-0.256	-0.076	-0.134
public_work	0.398	0.804	0.017	0.793	0.361	0.01
q1_id	-0.314	-0.179	-0.43	-0.239	-0.308	-0.412

	2008			2009		
	Factor1	*Factor2*	*Factor3*	*Factor1*	*Factor2*	*Factor3*
avg_income_defl	0.244	0.534	0.007	0.042	0.289	0.505
vinculo_ativo_3112	0.298	0.07	0.111	0.192	0.246	0.068
tempo_emprego	0.754	0.176	0.128	0.293	0.678	0.141
quantidade_horas_contratadas	-0.173	-0.196	-0.409	-0.426	-0.079	-0.23
idade	0.611	0.139	-0.143	0.025	0.631	0.082
tamanho_estabelecimento	0.357	0.105	0.385	0.464	0.244	0.086
race_id	-0.04	-0.262	-0.07	-0.069	-0.044	-0.267
educ_id	-0.152	0.697	0.448	0.344	-0.198	0.758
gender_id	-0.058	0.064	0.262	0.228	-0.129	0.107
treat_RP	-0.116	-0.157	-0.166	-0.243	-0.087	-0.15
public_work	0.603	0.007	0.624	0.785	0.385	0.031
q1_id	-0.328	-0.464	-0.129	-0.171	-0.334	-0.447

	2010			2011		
	Factor1	*Factor2*	*Factor3*	*Factor1*	*Factor2*	*Factor3*
avg_income_defl	0.178	0.56	0.133	0.251	0.022	0.58
vinculo_ativo_3112	0.269	0.03	0.004	0.295	0.143	0.099
tempo_emprego	0.815	0.076	0.052	0.743	0.184	0.179
quantidade_horas_contratadas	-0.068	-0.064	-0.336	-0.128	-0.47	-0.168
idade	0.48	0.12	-0.141	0.599	-0.09	0.124
tamanho_estabelecimento	0.022	0.229	0.034	0.295	0.405	0.092
race_id	-0.073	-0.128	-0.197	-0.046	-0.067	-0.262
educ_id	-0.124	0.38	0.761	-0.188	0.453	0.657
gender_id	-0.022	-0.183	0.387	-0.106	0.263	0.036
treat_RP	0.023	-0.117	-0.153	-0.113	-0.241	-0.139
public_work	0.169	0.044	0.074	0.518	0.658	0.034
q1_id	-0.187	-0.636	-0.082	-0.303	-0.129	-0.493

Source: Own elaboration.

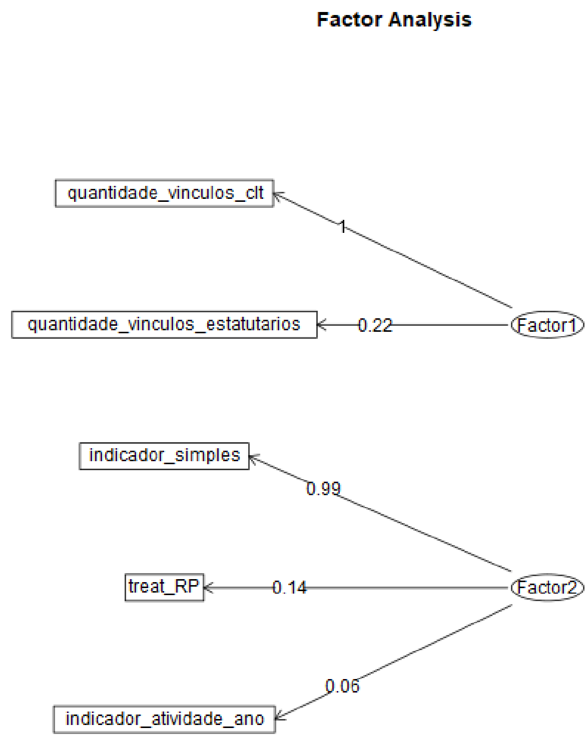
Table 62: Factor loadings for RAIS "vínculos" at the RP level, 2012 to 2015

	2012			2013		
	Factor1	*Factor2*	*Factor3*	*Factor1*	*Factor2*	*Factor3*
avg_income_defl	0.207	0.648	-0.073	0.204	0.675	-0.101
vinculo_ativo_3112	0.317	0.114	0.092	0.316	0.11	0.098
tempo_emprego	0.762	0.195	0.019	0.766	0.195	0.022
quantidade_horas_contratadas	-0.26	-0.114	-0.531	-0.231	-0.141	-0.525
idade	0.54	0.116	-0.167	0.539	0.091	-0.151
tamanho_estabelecimento	0.368	0.14	0.282	0.349	0.191	0.272
race_id	-0.053	-0.24	-0.089	-0.059	-0.233	-0.076
educ_id	-0.078	0.601	0.544	-0.073	0.586	0.505
gender_id	-0.049	-0.005	0.325	-0.045	-0.025	0.353
treat_RP	-0.169	-0.163	-0.194	-0.152	-0.184	-0.173
public_work	0.64	0.093	0.461	0.61	0.158	0.421
q1_id	-0.302	-0.536	-0.034	-0.287	-0.533	-0.029
	2014			**2015**		
	Factor1	*Factor2*	*Factor3*	*Factor1*	*Factor2*	*Factor3*
avg_income_defl	0.207	0.662	-0.088	0.217	0.653	-0.073
vinculo_ativo_3112	0.324	0.118	0.072	0.309	0.126	0.086
tempo_emprego	0.764	0.207	0.016	0.765	0.216	0.015
quantidade_horas_contratadas	-0.177	-0.144	-0.428	-0.167	-0.144	-0.419
idade	0.526	0.092	-0.159	0.539	0.087	-0.161
tamanho_estabelecimento	0.348	0.187	0.232	0.355	0.203	0.257
race_id	-0.083	-0.234	-0.085	-0.091	-0.244	-0.078
educ_id	-0.06	0.58	0.53	-0.079	0.589	0.493
gender_id	-0.043	-0.049	0.374	-0.042	-0.035	0.358
treat_RP	-0.149	-0.194	-0.174	-0.164	-0.192	-0.172
public_work	0.605	0.189	0.358	0.596	0.195	0.383
q1_id	-0.282	-0.533	-0.012	-0.281	-0.527	-0.023

Source: Own elaboration.

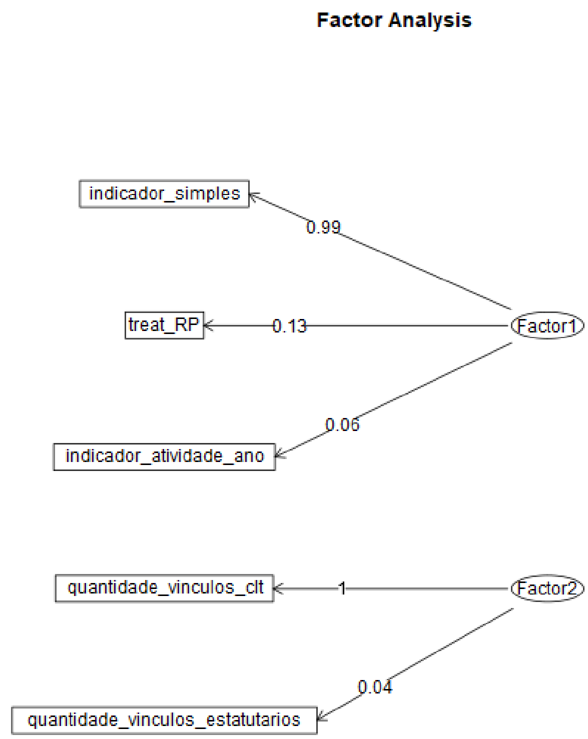
PCA results - RAIS “estabelecimentos” from 2003 to 2015

Figure 45: Factor decomposition for RP treatment in 2003, in RAIS "estabelecimentos" database



Source: Own elaboration.

Figure 46: Factor decomposition for RP treatment in 2015, in RAIS "estabelecimentos" database



Source: Own elaboration.

Table 63: Factor loadings for RAIS "estabelecimentos" at the AP level, 2003 to 2015

	2003		**2004**		**2005**	
	Factor1	*Factor2*	*Factor1*	*Factor2*	*Factor1*	*Factor2*
quantidade_vinculos_clt	0.997	0.02	0.995	-0.07	0.997	0.01
quantidade_vinculos_estatutarios	0.216	0.008	0.282	-0.011	0.264	0.006
indicador_simples	-0.066	0.995	-0.012	0.456	-0.043	0.997
indicador_atividade_ano	0.023	0.064	0.021	-0.127	0.025	0.035
treat_AP	-0.014	0.136	0.015	0.259	-0.009	0.181

	2006		**2007**		**2008**	
	Factor1	*Factor2*	*Factor1*	*Factor2*	*Factor1*	*Factor2*
quantidade_vinculos_clt	0.997	-0.025	0	0.241	0.006	0.626
quantidade_vinculos_estatutarios	0.053	-0.009	0.007	0.237	0.005	0.19
indicador_simples	-0.029	0.556	0.996	-0.05	0.996	-0.055
indicador_atividade_ano	0.016	-0.055	0.07	0.024	0.106	0.015
treat_AP	0.003	0.199	0.171	-0.029	0.165	-0.01

	2009		**2010**		**2011**	
	Factor1	*Factor2*	*Factor1*	*Factor2*	*Factor1*	*Factor2*
quantidade_vinculos_clt	-0.007	0.334	-0.03	0.011	0.018	0.589
quantidade_vinculos_estatutarios	0.003	0.14	-0.006	0.004	0	0.069
indicador_simples	0.996	-0.053	0.982	0.178	0.994	-0.082
indicador_atividade_ano	0.113	0.057	-0.031	0.77	0.109	0.058
treat_AP	0.142	-0.026	0.149	-0.017	0.155	-0.015

	2012		**2013**		**2014**	
	Factor1	*Factor2*	*Factor1*	*Factor2*	*Factor1*	*Factor2*
quantidade_vinculos_clt	0.026	0.75	0.049	0.996	0.034	0.997
quantidade_vinculos_estatutarios	0.001	0.072	0	0.065	-0.002	0.05
indicador_simples	0.994	-0.081	0.994	-0.087	0.995	-0.072
indicador_atividade_ano	0.087	0.045	0.081	0.036	0.056	0.042
treat_AP	0.132	-0.012	0.136	-0.009	0.139	-0.006

	2015	
	Factor1	*Factor2*
quantidade_vinculos_clt	0.058	0.981
quantidade_vinculos_estatutarios	-0.002	0.037
indicador_simples	0.992	-0.102
indicador_atividade_ano	0.062	0.041
treat_AP	0.119	-0.007

Source: Own elaboration.

Table 64: Factor loadings for RAIS "estabelecimentos" at the RP level, 2003 to 2015

	2003		2004		2005	
	Factor1	*Factor2*	*Factor1*	*Factor2*	*Factor1*	*Factor2*
quantidade_vinculos_clt	0.997	0.033	0.996	-0.045	0.997	0.015
quantidade_vinculos_estatutarios	0.216	0.011	0.281	-0.011	0.264	0.008
indicador_simples	-0.079	0.994	-0.034	0.228	-0.047	0.996
indicador_atividade_ano	0.022	0.064	0.018	-0.253	0.025	0.035
treat_RP	-0.007	0.136	0.02	0.588	-0.008	0.197

	2006		2007		2008	
	Factor1	*Factor2*	*Factor1*	*Factor2*	*Factor1*	*Factor2*
quantidade_vinculos_clt	0.997	0.001	-0.011	0.019	0.028	0.997
quantidade_vinculos_estatutarios	0.053	-0.012	-0.004	0.018	0.002	0.119
indicador_simples	-0.043	0.31	0.996	-0.048	0.996	-0.056
indicador_atividade_ano	0.017	-0.099	0.085	0.349	0.106	0.007
treat_RP	0.001	0.298	0.161	-0.214	0.161	-0.003

	2009		2010		2011	
	Factor1	*Factor2*	*Factor1*	*Factor2*	*Factor1*	*Factor2*
quantidade_vinculos_clt	0.049	0.996	0.007	-0.029	-0.03	0.048
quantidade_vinculos_estatutarios	-0.001	0.048	0.003	-0.006	-0.006	0.006
indicador_simples	0.995	-0.073	0.179	0.981	0.997	0
indicador_atividade_ano	0.111	0.013	0.995	-0.072	0.104	0.812
treat_RP	0.152	-0.006	-0.012	0.161	0.153	-0.031

	2012		2013		2014	
	Factor1	*Factor2*	*Factor1*	*Factor2*	*Factor1*	*Factor2*
quantidade_vinculos_clt	0.067	0.995	0.06	0.996	0.043	0.997
quantidade_vinculos_estatutarios	0	0.055	0.001	0.065	-0.001	0.05
indicador_simples	0.992	-0.101	0.993	-0.099	0.994	-0.081
indicador_atividade_ano	0.087	0.031	0.081	0.035	0.056	0.041
treat_RP	0.148	-0.007	0.156	-0.007	0.171	-0.004

	2015	
	Factor1	*Factor2*
quantidade_vinculos_clt	0.069	0.995
quantidade_vinculos_estatutarios	-0.002	0.036
indicador_simples	0.991	-0.112
indicador_atividade_ano	0.063	0.04
treat_RP	0.133	-0.005

Source: Own elaboration.

PCA results - PNAD from 2001 to 2015

Table 65: Factor loadings for PNAD, 2001 to 2006

	2001				2002			
	Factor1	*Factor2*	*Factor3*	*Factor4*	*Factor1*	*Factor2*	*Factor3*	*Factor4*
q1_id	0.137	-0.272	-0.053	0.013	0.139	-0.312	-0.049	0.025
idade	-0.381	0.09	0.294	0.064	-0.381	0.101	0.315	0.088
anos_estudo	0.245	0.304	0.068	-0.039	0.258	0.355	0.062	-0.077
gender_id	-0.337	-0.101	-0.03	-0.093	-0.34	-0.1	-0.052	-0.089
race_id	0.031	-0.161	-0.064	-0.03	0.045	-0.166	-0.056	0.001
trabalhou_semana	0.933	0.026	-0.136	0.153	0.932	0.006	-0.164	0.152
possui_carteira_assinada	0.569	-0.03	-0.09	-0.472	0.569	-0.007	-0.112	-0.515
horas_trabalhadas_todos_trabalhos	0.911	0.034	-0.134	0.105	0.913	0.015	-0.16	0.106
entrep	0.271	0.049	-0.03	0.871	0.277	-0.008	-0.028	0.797
renda_aposentadoria_pensao	-0.106	0.236	0.963	-0.031	-0.1	0.227	0.958	-0.024
renda_mensal_ocupacao_principal	0.324	0.911	-0.236	0.062	0.36	0.871	-0.247	0.118
renda_mensal_todas_fontes	0.188	0.89	0.223	0.027	0.251	0.907	0.302	0.084

	2003				2004			
	Factor1	*Factor2*	*Factor3*	*Factor4*	*Factor1*	*Factor2*	*Factor3*	*Factor4*
q1_id	0.188	-0.295	-0.05	0.008	0.18	-0.336	-0.048	0.024
idade	-0.385	0.091	0.294	0.069	-0.406	0.088	0.299	0.079
anos_estudo	0.248	0.335	0.074	-0.062	0.251	0.348	0.047	-0.065
gender_id	-0.315	-0.115	-0.039	-0.108	-0.297	-0.105	-0.043	-0.098
race_id	0.022	-0.161	-0.063	-0.02	0.026	-0.166	-0.051	-0.013
trabalhou_semana	0.933	0.059	-0.137	0.164	0.932	0.028	-0.129	0.157
possui_carteira_assinada	0.59	0	-0.082	-0.468	0.576	-0.012	-0.078	-0.503
horas_trabalhadas_todos_trabalhos	0.911	0.06	-0.136	0.137	0.917	0.048	-0.128	0.107
entrep	0.251	0.039	-0.024	0.841	0.27	0.005	-0.018	0.812
renda_aposentadoria_pensao	-0.128	0.258	0.954	-0.031	-0.127	0.244	0.959	-0.02
renda_mensal_ocupacao_principal	0.312	0.899	-0.251	0.071	0.353	0.846	-0.226	0.078
renda_mensal_todas_fontes	0.18	0.914	0.317	0.038	0.189	0.857	0.362	0.044

	2005				2006			
	Factor1	*Factor2*	*Factor3*	*Factor4*	*Factor1*	*Factor2*	*Factor3*	*Factor4*
q1_id	0.172	-0.324	-0.042	0.024	0.19	-0.252	-0.091	0.022
idade	-0.402	0.108	0.295	0.075	-0.389	0.084	0.313	0.062
anos_estudo	0.251	0.332	0.068	-0.065	0.261	0.265	0.082	-0.05
gender_id	-0.328	-0.101	-0.045	-0.096	-0.323	-0.086	-0.029	-0.076
race_id	0.034	-0.178	-0.067	-0.017	0.017	-0.129	-0.068	-0.017
trabalhou_semana	0.923	0.027	-0.131	0.158	0.91	0.036	-0.16	0.168
possui_carteira_assinada	0.593	-0.014	-0.086	-0.518	0.597	0	-0.104	-0.49
horas_trabalhadas_todos_trabalhos	0.913	0.042	-0.129	0.101	0.901	0.046	-0.154	0.111
entrep	0.273	0.026	-0.022	0.817	0.246	0.035	-0.023	0.823
renda_aposentadoria_pensao	-0.123	0.255	0.956	-0.024	-0.108	0.236	0.963	-0.027
renda_mensal_ocupacao_principal	0.365	0.877	-0.22	0.101	0.25	0.928	-0.203	0.042
renda_mensal_todas_fontes	0.21	0.88	0.386	0.062	0.16	0.945	0.223	0.024

Source: Own elaboration.

Table 66: Factor loadings for PNAD, 2007 to 2013

	2007				**2008**			
	Factor1	*Factor2*	*Factor3*	*Factor4*	*Factor1*	*Factor2*	*Factor3*	*Factor4*
q1_id	0.161	-0.222	-0.072	0.029	0.201	-0.323	-0.046	-0.005
idade	-0.395	0.094	0.315	0.06	-0.41	0.081	0.293	0.098
anos_estudo	0.294	0.237	0.065	-0.061	0.239	0.313	0.048	-0.058
gender_id	-0.313	-0.074	-0.043	-0.089	-0.324	-0.082	-0.023	-0.053
race_id	0.026	-0.123	-0.078	0.001	0.031	-0.162	-0.067	0.001
trabalhou_semana	0.904	0.003	-0.175	0.192	0.918	0.029	-0.13	0.176
possui_carteira_assinada	0.601	-0.021	-0.113	-0.481	0.596	-0.002	-0.085	-0.52
horas_trabalhadas_todos_trabalhos	0.9	0.016	-0.171	0.12	0.897	0.048	-0.137	0.125
entrep	0.229	0.008	-0.034	0.82	0.237	0.016	-0.018	0.787
renda_aposentadoria_pensao	-0.075	0.188	0.976	-0.025	-0.13	0.248	0.957	-0.022
renda_mensal_ocupacao_principal	0.268	0.916	-0.209	0.075	0.334	0.877	-0.252	0.08
renda_mensal_todas_fontes	0.201	0.933	0.253	0.056	0.193	0.906	0.303	0.052
	2009				**2011**			
	Factor1	*Factor2*	*Factor3*	*Factor4*	*Factor1*	*Factor2*	*Factor3*	*Factor4*
q1_id	0.175	-0.292	-0.061	0.017	0.174	-0.295	-0.038	-0.009
idade	-0.426	0.1	0.305	0.094	-0.414	0.101	0.296	0.091
anos_estudo	0.275	0.269	0.062	-0.071	0.27	0.275	0.046	-0.082
gender_id	-0.316	-0.091	-0.029	-0.063	-0.301	-0.08	-0.034	-0.078
race_id	0.013	-0.145	-0.075	-0.012	0.023	-0.156	-0.066	-0.004
trabalhou_semana	0.913	0.011	-0.136	0.184	0.934	0.022	-0.147	0.137
possui_carteira_assinada	0.615	-0.025	-0.093	-0.517	0.611	-0.026	-0.094	-0.54
horas_trabalhadas_todos_trabalhos	0.91	0.039	-0.144	0.132	0.9	0.019	-0.143	0.117
entrep	0.22	0.035	-0.019	0.789	0.248	0.003	-0.025	0.819
renda_aposentadoria_pensao	-0.114	0.219	0.966	-0.029	-0.125	0.232	0.962	-0.024
renda_mensal_ocupacao_principal	0.305	0.921	-0.22	0.075	0.336	0.903	-0.224	0.05
renda_mensal_todas_fontes	0.174	0.889	0.211	0.053	0.212	0.901	0.314	0.026
	2012				**2013**			
	Factor1	*Factor2*	*Factor3*	*Factor4*	*Factor1*	*Factor2*	*Factor3*	*Factor4*
q1_id	0.142	-0.256	-0.052	-0.02	0.176	-0.316	-0.036	0
idade	-0.421	0.116	0.321	0.073	-0.405	0.088	0.328	0.056
anos_estudo	0.286	0.239	0.056	-0.054	0.271	0.275	0.043	-0.063
gender_id	-0.302	-0.073	-0.038	-0.087	-0.291	-0.071	-0.017	-0.082
race_id	0.027	-0.163	-0.082	-0.012	0.021	-0.163	-0.07	-0.012
trabalhou_semana	0.927	0.003	-0.18	0.159	0.925	0.034	-0.165	0.151
possui_carteira_assinada	0.625	-0.061	-0.12	-0.495	0.626	-0.037	-0.102	-0.467
horas_trabalhadas_todos_trabalhos	0.905	0.01	-0.174	0.127	0.881	0.033	-0.152	0.128
entrep	0.218	0.04	-0.041	0.82	0.221	0.04	-0.028	0.894
renda_aposentadoria_pensao	-0.108	0.202	0.97	-0.042	-0.13	0.21	0.966	-0.031
renda_mensal_ocupacao_principal	0.306	0.913	-0.212	0.046	0.331	0.899	-0.218	0.072
renda_mensal_todas_fontes	0.23	0.932	0.219	0.025	0.229	0.918	0.255	0.046

Source: Own elaboration.

Table 67: Factor loadings for PNAD, 2014 and 2015

	2014				**2015**			
	Factor1	*Factor2*	*Factor3*	*Factor4*	*Factor1*	*Factor2*	*Factor3*	*Factor4*
q1_id	0.129	-0.357	-0.048	-0.011	0.091	-0.265	-0.058	0.007
idade	-0.431	0.097	0.294	0.078	-0.413	0.112	0.31	0.062
anos_estudo	0.272	0.297	0.039	-0.065	0.293	0.225	0.053	-0.045
gender_id	-0.297	-0.07	-0.025	-0.068	-0.282	-0.062	-0.018	-0.077
race_id	0.05	-0.189	-0.072	-0.006	0.011	-0.162	-0.067	-0.008
trabalhou_semana	0.917	0.007	-0.154	0.171	0.921	-0.011	-0.165	0.169
possui_carteira_assinada	0.651	-0.061	-0.092	-0.517	0.63	-0.053	-0.102	-0.499
horas_trabalhadas_todos_trabalhos	0.904	0.026	-0.148	0.141	0.881	0.021	-0.156	0.116
entrep	0.219	0.037	-0.029	0.808	0.23	0.017	-0.03	0.865
renda_aposentadoria_pensao	-0.14	0.215	0.964	-0.032	-0.123	0.224	0.964	-0.027
renda_mensal_ocupacao_principal	0.35	0.856	-0.216	0.085	0.298	0.91	-0.227	0.057
renda_mensal_todas_fontes	0.243	0.899	0.228	0.067	0.215	0.931	0.172	0.036

Source: Own elaboration.

FACULDADE DE **ECONOMIA**