

FUNDAÇÃO GETULIO VARGAS
ESCOLA DE ECONOMIA DE SÃO PAULO

EDUARDO ZYLBERSTAJN

**TRÊS ENSAIOS SOBRE O
MERCADO DE TRABALHO NO BRASIL**

SÃO PAULO
2015

FUNDAÇÃO GETULIO VARGAS
ESCOLA DE ECONOMIA DE SÃO PAULO

EDUARDO ZYLBERSTAJN

**TRÊS ENSAIOS SOBRE O
MERCADO DE TRABALHO NO BRASIL**

Tese apresentada à banca examinadora da
Escola de Economia de São Paulo da Fundação
Getulio Vargas, como requisito para obtenção
do título de Doutor em Economia.

Área de Conhecimento:
Economia do Trabalho

Orientador: Prof. Dr. André Portela Fernandes
de Souza

SÃO PAULO
2015

Zylberstajn, Eduardo.

Três ensaios sobre o mercado de trabalho no Brasil / Eduardo Zylberstajn. - 2015.
80 f.

Orientador: André Portela Fernandes de Souza

Tese (doutorado) - Escola de Economia de São Paulo.

1. Mercado de trabalho - Brasil. 2. Desemprego. 3. Educação e emprego. 4. Terceirização. 5. Salários. 6. Análise de variância. I. Souza, André Portela Fernandes de. II. Tese (doutorado) - Escola de Economia de São Paulo. III. Título.

CDU 331.6(81)

EDUARDO ZYLBERSTAJN

TRÊS ENSAIOS SOBRE O MERCADO DE TRABALHO NO BRASIL

Tese apresentada à Escola de Economia de São Paulo da Fundação Getúlio Vargas, como requisito para obtenção do título de Doutor em Economia.

Área de Conhecimento: Economia do Trabalho

Orientador: Prof. Dr. André Portela Fernandes de Souza

Data de aprovação:

___/___/____.

Banca examinadora:

Prof. Dr. André Portela Fernandes de Souza
(orientador)
FGV - EESP

Prof. Dr. Naércio Aquino Menezes Filho
Insper e FEA-USP

Prof. Dr. Reynaldo Fernandes
FEARP - USP

Profa. Dra. Cristine Campos de Xavier Pinto
FGV - EESP

Prof. Dr. Vladimir Pinheiro Ponczek
FGV - EESP

AGRADECIMENTOS

Ao professor André Portela, pelo apoio, dedicação, paciência e pela inspiração. Aos professores Cristine Pinto, Sergio Firpo e Vladimir Ponczek, pelos ensinamentos e pela disposição em sempre ajudar. Aos professores Enlinson Mattos, Luis Fernando Araujo, Marcio Holland, Marcos Fernandes Gonçalves da Silva, Marcos Hiroyuki Tsuchida, Maria Carolina da S. Leme (*in memoriam*), Paulo Furquim, Paulo Picchetti, Pedro Valls, Vladimir Kuhl Teles, por tudo o que ensinaram ao longo dos últimos oito anos de Mestrado e Doutorado.

Aos meus colegas de doutorado, agradeço pela convivência e troca de experiências.

Agradeço a Fundação Getúlio Vargas, pela oportunidade e infraestrutura oferecidas. Agradeço também pelos amigos que o doutorado me trouxe: Priscilla Tavares, Priscila Ribeiro, Bruno Oliva, Rafael Camelo, Felipe Garcia e Guilherme Stein.

Agradeço a Fipe, por me oferecer as condições de exercer minha nova profissão.

À minha mãe, Berta, pelo amor, carinho e dedicação desde 1982. À minha irmã, o agradecimento é pelo carinho e companhia desde que nasci, e também pelas sobrinhas que ganhei. Ao Hélio, agradeço pelo exemplo de pai, de homem e de economista que é. Aos três, sou profundamente grato por tudo o que me propiciaram na vida.

À minha esposa, Fernanda, devo um agradecimento especial. Os últimos 4 anos foram muito peculiares em diversos aspectos e sou grato por ter conhecido novos limites da mulher mais forte, da mãe mais incrível e da melhor esposa que eu poderia sonhar em encontrar. Agradeço pela compreensão (especialmente nos últimos meses), pelo apoio e pelo amor. Como já lhe disse antes, é minha heroína. Aos meus sogros George e Elza, agradeço por todo o apoio.

Ao Rafael, agradeço por sua ternura e carinho, e também peço desculpas pela ausência que ele tanto sentiu nos últimos meses. À minha filha Tali, agradeço por me mostrar o que é ser forte. As dois, agradeço pela nova pessoa em quem me transformaram.

“Try not to become a man of success.

Rather, become a man of value.”

Albert Einstein

RESUMO

Esta tese é composta por três artigos e uma nota, sendo um em cada capítulo. Todos os capítulos enquadram-se na área de Microeconomia Aplicada e Economia do Trabalho.

O primeiro artigo estende o modelo tradicional de decomposição das flutuações na taxa de desemprego de Shimer (2012), separando o emprego formal do informal. Com essa modificação, os principais resultados da metodologia se alteram e conclui-se que os principais fatores para a queda do desemprego na última década foram (i) a queda na taxa de participação, principalmente pela menor entrada na força de trabalho; (ii) o aumento da formalização, atingido tanto pelo aumento da probabilidade de encontrar um trabalho formal quanto pela probabilidade de deixar a condição de empregado formal.

O segundo capítulo apresenta estimativas para o retorno à educação no Brasil, utilizando uma nova metodologia que não necessita de variáveis de exclusão. A vantagem do método em relação a abordagens que utilizam variáveis instrumentais é a de permitir avaliar o retorno médio para todos os trabalhadores (e não somente os afetados pelos instrumentos) e em qualquer instante do tempo. Face aos resultados, concluímos as estimativas via MQO subestimam o retorno médio. Discute-se possíveis explicações para esse fenômeno.

O terceiro artigo trata da terceirização da mão de obra no Brasil. Mais especificamente, mede-se o diferencial de salários entre os trabalhadores terceirizados e os contratados diretamente. Os resultados de uma comparação não condicional indicam que os terceirizados têm salário médio 17% menor no período 2007 a 2012. Porém, com estimativas que levam em conta o efeito fixo de cada trabalhador, esse diferencial cai para 3,0%. Além disso, o diferencial é bastante heterogêneo entre os tipos de serviços: aqueles que utilizam trabalhadores de baixa qualificação apresentam salário menores, enquanto nas ocupações de maior qualificação os terceirizados têm salários iguais ou maiores do que os diretamente contratados. Mais ainda, as evidências apontam para a diminuição do diferencial ao longo do tempo no período analisado.

Finalmente, a nota que encerra a tese documenta dois aspectos relevantes e pouco conhecidos da Pesquisa Mensal de Emprego do IBGE que podem levar a resultados imprecisos nas pesquisas que utilizam esse painel se não forem tratados adequadamente.

Palavras-chave: Mercado de trabalho, desemprego, transições no mercado de trabalho, retornos a educação, terceirização da mão de obra, diferenciais de salários

ABSTRACT

This dissertation explores in three essays and one short note different issues related to the Brazilian Labor Market.

The first chapter extends the traditional literature on flows into and out of unemployment (as in Shimer, 2012) by accounting for two different types of employment: formal and informal. Results suggest that the standard three-state approach is insufficient for a full understanding of the labor market dynamics witnessed in the past decade in Brazil, since it overstates the contribution of the employment exit rate to the change in unemployment. When disentangling formal and informal work, it is shown that the main drivers of the decline in the unemployment rate between 2003 and 2014 were (in order of relevance): (i) the decline in the participation rate that happened due to a decrease in the entries into the workforce (from inactivity); (ii) the increase in the formal job finding rate and (iii) the decline in the inflow from formal employment into unemployment.

The second chapter investigates the causal effect of education on earnings in Brazil by employing a new method that does not need first order exclusion restrictions. This approach yields unbiased estimates in the absence of instruments and allows for interpretation of the coefficients that is not confined to local average treatment effects. Results indicate that the return to schooling is downward biased in OLS estimates.

The third chapter investigates wage differentials between outsourced and directly hired employees. The findings show that unobservable characteristics account for most of the difference and that low skill occupations tend to impose higher penalties on wages for outsourced workers.

Finally, chapter four presents a short note documenting two relevant aspects of the Monthly Employment Survey (*Pesquisa Mensal de Emprego/IBGE*) that are usually neglected by researchers but may lead to imprecise results if not properly accounted for.

Keywords: Labor market, unemployment, employment flows, transition rates, returns to education, outsourcing, wage differentials

LISTA DE TABELAS

Table 1 – Labor status distribution after correcting for response error	17
Table 2 – Level and volatility of transition probabilities	17
Table 3 – Decomposition of the change in the unemployment rate	24
Table 4 – Accounting for informality in the decomposition of the change in the unemployment rate	30
Table 5 – Decomposition of the change in the participation rate	31
Table 6 – Decomposition of the change in the formality rate	33
Table 7 – Variables description.....	42
Table 8 – Descriptive statistics.....	42
Table 9 – OLS Estimates: Education equation (2013)	45
Table 10 – OLS and Control function estimates: Wage equation (2013).....	46
Table 11 – Cross tabulation of positive and negative residuals (2013).....	50
Table 12 - Control function estimates with alternative specifications: Wage equation (2013)	51
Table 13 – Heteroskedasticity test: Education equation (2013)	54
Table 14 – Heteroskedasticity test: Wage equation (2013).....	54
Table 15 –Heteroskedasticity functions: education and wage equations (2013).....	55

Tabela 1 – Serviços consideradas para a análise da terceirização e suas participações no emprego formal do Brasil (2007/2012)	61
Tabela 2 – Estatísticas descritivas	62
Tabela 3 – Transições de e para a terceirização	64
Tabela 4 – Efeitos da terceirização nos salários	65
Tabela 5 – Efeitos da terceirização nos salários de acordo com as ocupações terceirizadas	66
Tabela 6 – Efeitos da terceirização nos salários em subamostras (2007-2008, 2009-2010 e 2011-2012)	67
Tabela 7 – Viés de grupo rotacional na taxa de desemprego	75
Tabela 8 – Distribuição do total de entrevistas completadas na PME	77
Tabela 9 – Matriz de transição mensal sem correção (PME/IBGE)	77
Tabela 10 – Matriz de transição mensal após a correção das trajetórias (PME/IBGE)	78

LISTA DE FIGURAS

Figure 1 – Unemployment and participation rates in Brazil	14
Figure 2 – Monthly transition rates: unemployment, employment and inactivity.....	18
Figure 3 – Monthly transition rates: unemployment, formal/informal employment and inactivity.....	19
Figure 4 – Unemployment rate	21
Figure 5 – Contribution of fluctuations in transition rates to the unemployment rate.....	22
Figure 6 – Unemployment rate	26
Figure 7 – Accounting for informality to assess the contribution of fluctuations in transition rates to the unemployment rate.....	27
Figure 8 – Formality rate	32
Figure 9 – Education equation: graphical analysis of heteroskedasticity	43
Figure 10 – Wage equation: graphical analysis of heteroskedasticity.....	44
Figure 11 – Returns to education in Brazil: 1995-2013	47
Figure 12 – Residuals analysis: education and wage equations (2013)	49

Figura 1 – Taxa de desemprego mensal da PME para os indivíduos na primeira e oitava entrevistas	75
Figura 2 – Taxa de desemprego (média anual) da PME para segundo o painel rotacional.....	76

SUMÁRIO

Capítulo 1: The Ins and Outs of Unemployment in a Dual Labor Market.....	12
1. Introduction.....	13
2. Data.....	14
3. Using flows to assess unemployment dynamics	20
4. Accounting for informality	24
5. Participation and formalization rates.....	30
6. Conclusion.....	33
References.....	35
Capítulo 2: Estimating the Returns to Education in Brazil: a parametric control function approach.....	36
1. Introduction.....	37
2. Empirical exercise.....	38
2.1 Econometric model and identification.....	38
2.2 Implementation.....	40
3. Data.....	41
4. Results.....	44
5. Discussion: the OLS-IV gap.....	48
6. Robustness check.....	50
7. Conclusion.....	51
References.....	52
Capítulo 3: Diferencial de salários da mão de obra terceirizada no Brasil.....	56
1. Introdução.....	57
2. Dados.....	59
3. Metodologia e modelo econométrico.....	62
4. Resultados.....	64
5. Testes de robustez.....	66
6. Conclusões e considerações finais.....	68
Referências	69
Anexo	70
Capítulo 4: Painéis de rotação e trajetórias inconsistentes: Fontes de viés na Pesquisa Mensal de Emprego do IBGE.....	73
1. Introdução.....	74
2. Viés do painel rotacional.....	74
3. Trajetórias inconsistentes no mercado de trabalho.....	77
4. Considerações finais.....	79
Referências	80

Capítulo 1: The Ins and Outs of Unemployment in a Dual Labor Market

Abstract

This paper extends the traditional literature on flows into and out of unemployment by accounting for two different types of employment. We develop a four-state model and apply it to Brazil, where formal and informal work are both relevant in the labor market. Our findings suggest that the standard three-state approach is insufficient for a full understanding of the labor market dynamics witnessed in the past decade, since it overstates the contribution of the employment exit rate to the change in unemployment. When disentangling formal and informal work, we find that the main drivers of the decline in the unemployment rate between 2003 and 2014 were (in order of relevance): (i) the decline in the participation rate that happened due to a decrease in the entries into the workforce (from inactivity); (ii) the increase in the formal job finding rate and (iii) the decline in the inflow from formal employment into unemployment.

Keywords: unemployment, employment flows, transition rates, participation rate.

JEL Classification: J6, E24, E32

1. Introduction

What drives the unemployment dynamics: inflows or outflows? Many authors raised this question in the past decade, and still no consensus has emerged: answers change according to the countries where the analysis is made and also depend on the methodology adopted in each study. Understanding the unemployment rate dynamics is extremely relevant for macroeconomic and labor market policies, since the absence or presence of cyclicity in separations/admissions should drive employment protection policies and researchers' attention.

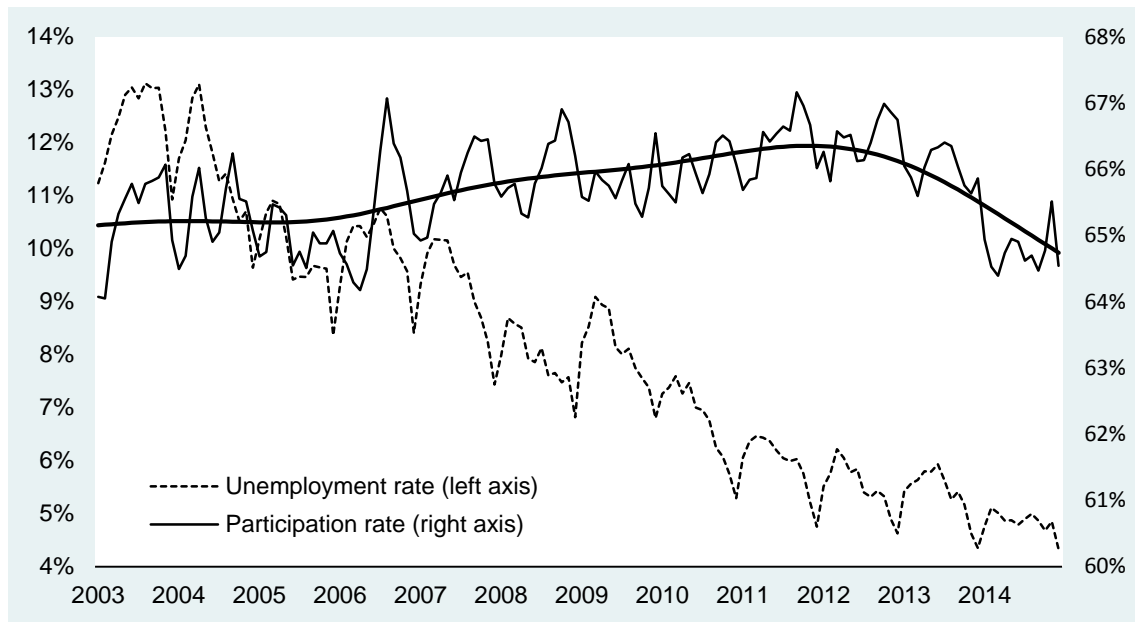
While Hall (2005) and Shimer (2007 and 2012) found that in the US outflows dominates the unemployment rate's dynamics, Fujita and Ramey (2009) and Elsby et al. (2009) have argued the opposite. Petrongolo and Pissarides (2008) found that in the UK and in Spain the inflow rate explains roughly half of the changes in unemployment dynamics, whereas in France the outflows are more relevant. Smith (2011) contributed by developing a slightly different framework that accounts for deviations in the equilibrium unemployment rate. We make a different, but in our view also relevant contributions to the theory developed by these authors.

The contributions of this paper are twofold. First, we extend the traditional framework that calculates the relative importance of changes in inflow and outflow rates to equilibrium unemployment dynamics by incorporating a second employment status. This is particularly relevant for economies with large informal sectors, but even in the developed world this might be useful. Particularly after the Great Recession, structural changes in the labor markets are occurring throughout the world, with increases in the share of part time work and contingent work. Attachment to employment or the labor force might be different depending on the type of employment relationship, and thus accounting for these two different states might be relevant.

Second, we investigate three relevant changes in the Brazilian labor market that occurred in the past decade: the decline of roughly seven percentage points in the unemployment rate (see Figure 1), the increase of more than 10 p.p. in the share of formal employment from and the decline of roughly 1.5 p.p since 2012 in the share of the working age population that engages in the labor market (the participation rate).

Menezes Filho et al. (2014) documented that the separation rate declined and the job finding rate increased in Brazil in recent years and that both movements contributed to the decline in the unemployment rate, but they argue that the latter explained almost entirely the unemployment variation between 2002 and 2009. Silva and Pires (2014), using the standard framework to assess which flows are more relevant to explain the unemployment dynamics in Brazil, find the opposite and argue that more than 80% of the unemployment rate variation is explained by changes in the separation rate (inflows). Attuy (2012) find similar results, even though his focus is on the relationship between unemployment, transitions and the business cycle. However, the last two papers do not pay proper attention to movements in and out of the labor force, even if their modeling strategy allows for it; also, there seems to be a confusion in both papers regarding what they are actually measuring. The authors sometimes refer to the contributions of changes in the transitions rates' volatility to the unemployment rate's volatility as if they were contributions of changes in the level of each rate to the level of unemployment.

Figure 1 – Unemployment and participation rates in Brazil



Note: Unemployment and participation rate are calculated based on PME (IBGE) data for individuals aged 15 years of age or more. Thick lines depict each series' HP filtered trend, using a smoothing parameter of 14,400.

Still, our findings show that in the traditional three state world, indeed it was the decline of the inflows into unemployment that explain almost half of the decline in unemployment in Brazil between 2003 and 2014. However, when disentangling informal and formal employment (and thus allowing for a fourth state in the labor market), the most relevant change was the decline in the likelihood of entering the workforce. If analyzing employment types separately, the outflows of unemployment into formal employment explained slightly more than the inflows (i.e. formal job finding rates dominate), whereas regarding informal employment and unemployment, inflows and outflows almost offset each other. It seems relevant, thus, to account for different labor market statuses in this type of analysis.

This paper is organized as follows. Section 2 describes the data used throughout this paper. Section 3 describes the standard three-state model used in the literature to assess changes in the unemployment rate and describe the results of such model when applied to the Brazilian case. Section 4 extends this model to account for two different employment types and shows how results change with this improvement. Finally, section 5 extends the analysis for the formality and participation rates.

2. Data

This paper uses the *Pesquisa Mensal de Emprego* (Monthly Employment Survey, henceforth PME), a monthly rotating panel of households in six major metropolitan areas in Brazil (São Paulo, Rio de Janeiro, Belo Horizonte, Salvador, Porto Alegre and Recife) surveyed by IBGE (Brazilian Institute of Geography and Statistics). The six regions represent roughly 25% of the country's population, providing a reach set of information on employment in

Metropolitan Areas in Brazil. The paper covers the period from January, 2003 to December, 2014¹.

The survey investigates schooling, labor force status, demographic attributes, and labor income of each dweller aged 10 or more that lives on the surveyed households. This yields approximately 100,000 individuals from 35,000 households every month. We restrain our sample to individuals aged between 15 and 70 years.

The rotating scheme is as follows. Each household is surveyed during the first four consecutive months. After this initial period, the household leaves the sample for eight months and then returns to the panel and is interviewed again for another four consecutive months. After the eighth interview the household is permanently excluded from the sample. Households are divided into 4 rotating groups in order to ensure that 75% of the sample is repeated in two consecutive months.

Shimer (2012) argues that using survey data to measure the transition rates might yield biased estimates due to time aggregation. In fact, several countries carry labor force surveys on a quarterly basis, and both Shimer (2012) and Gomes (2015) show that quarterly data underestimate the probability of transiting between states, because some individuals can move more than once during this relatively long time span. Shimer (2012) proposes a method to recover the instantaneous transitions, which he used in his study of the US case. However, Petrongolo and Pissarides (2008) argue and Gomes (2015) shows that for monthly surveys, time aggregation does not influence their results, whereas calculations based on quarterly surveys are quite significantly affected. Based on Petrongolo and Pissarides, we have used discrete monthly transitions rates to obtain the results described later in this paper.

One important caveat of PME survey is that it does not identify individuals directly. Only their households are identified, making it difficult to match each person over time. Also important to highlight is the fact that in each household only one individual answers the questions for all members in each round. Thus, attrition and reporting errors occur with a high frequency. For instance, it is common to have more than two years of difference in age or schooling in two consecutive months for the same individual. We follow the matching algorithm proposed by Ribas and Soares (2008), which corrects for misinformation on individuals' characteristics and reduces attrition. After pairing each individuals and restraining the sample to the working age population (between 15 and 70 years of age), our database contains 11,623,365 observations.

Because we are interested in transition rates, we only keep the observations for which we have information for two consecutive months. To avoid attrition bias (which still remains even after pairing individuals as shown by Ribas and Soares, 2008) we only use the first four months of each household's interviews. For instance, for any given individual, if s/he participated in all four interviews, she will contribute with three observations: the pairs of interviews (1, 2), (2,3) and (3,4). If there was attrition and she did not complete the four first interviews, we will have less contributions. We then pool all pairs of information into

¹ We will focus on transition rates between employment statuses considering two consecutive months. Thus, to assess the period from January, 2003 to December, 2014 we used observations since December, 2002.

one single dataset and discard individuals with missing employment status information on one of both sequential periods. This leaves us with 5,505,985 observations in the dataset.

For the objective of this paper, we must also examine carefully the impact of response errors. To put it shortly, response errors occur whenever the interviewed member of the household misreports the labor status of any of the other inhabitants of the household (see Abowd and Zellner, 1985; Smith, 2011; Gomes, 2015). As Smith (2011) and Gomes (2015) point out, this error can be particularly relevant for our study, since they create spurious transitions. In our case, the most relevant cases of misreporting are answers regarding to formal and informal work and out of labor force and unemployed status. For instance, it is common to find in the data individuals who are reported as informal workers in a given month, than in formal employment in the subsequent month and back to informal in the next period, but on the same time with long employment spells recorded in every interview (and thus with an inconsistent employment history). Because of the way formal employment is defined (the employer must sign each workers *carteira de trabalho*; if she does not sign, than the job relation is considered informal), the interviewed member might not have perfect information on the member of interest's *carteira* status (remember that in each round the respondent might change) and be confused or mistaken when answering the survey. Thus, the likelihood of misreporting formal/informal employment might be high. The same holds for every other employment status, including unemployment and non-participation. In Chapter 4 the sources of this error are more deeply explored and a correction algorithm is proposed.

By adopting the procedure described in Chapter 4, we corrected 7.92% of individuals' labor status (a total of 435,785 observations). This apparently small share of response error actually decreased average transitions significantly, and has caused a decrease of approximately 1 percentage point in the average unemployment rate for 2003:01 to 2014:12 (see Table 1 and Table 2). Smith (2011), also faces a similar pattern depending on the method adopted for the correction. We refer the reader to Chapter 4 for details on the correction procedure. Whenever we present results throughout the paper, we will always show the results for both the corrected and original data. Often, the results do not change significantly. What is most important is that there is a substantial commonality across the corrected and original series.

We are interested in the information on each individual's employment status. Originally, we would observe in the data seven possible classifications: out of the labor force (OLF), unemployed, unpaid worker, self-employed, employer, informal employee and formal employee. We grouped the categories by reclassifying the unpaid workers as OLF, self-employed and informal employees as informal employment and employers and formal employees as formal employment. This is important for the four-state model we describe in section 4, which accounts for OLF, unemployment and formal and informal employment only.

Table 1 – Labor status distribution after correcting for response error

	Original		Corrected	
	Obs.	%	Obs.	%
Out of labor force	1,915,093	34.8%	1,946,637	35.4%
Unemployed	318,034	5.8%	268,896	4.9%
Self employed	620,677	11.3%	624,815	11.4%
Employer	146,252	2.7%	139,505	2.5%
Informal employment	624,317	11.3%	594,005	10.8%
Formal employment	1,880,985	34.2%	1,931,500	35.1%
Total	5,505,985	100.0%	5,505,985	100.0%

Source: Author's calculations based on PME (IBGE) for the period 2003:01 to 2014:12.

Note: unpaid workers are considered out of the labor force, self-employed and informal employees are grouped into informal employment and employers and formal employees are grouped into formal employment.

An interesting test is to compare the average transition rates for Brazil with those reported by Gomes (2015) for the US, based on CPS data. Results for the transitions involving employment (E, considering both formal and informal as one status), unemployment (U) and inactivity (I) are shown in Table 2. It is interesting to note that after correcting for response error we have obtained smaller values and volatilities for the transition probabilities (as would be expected), and the US serves as a good benchmark (remember that there is no response error correction in the US estimates).

Table 2 – Level and volatility of transition probabilities

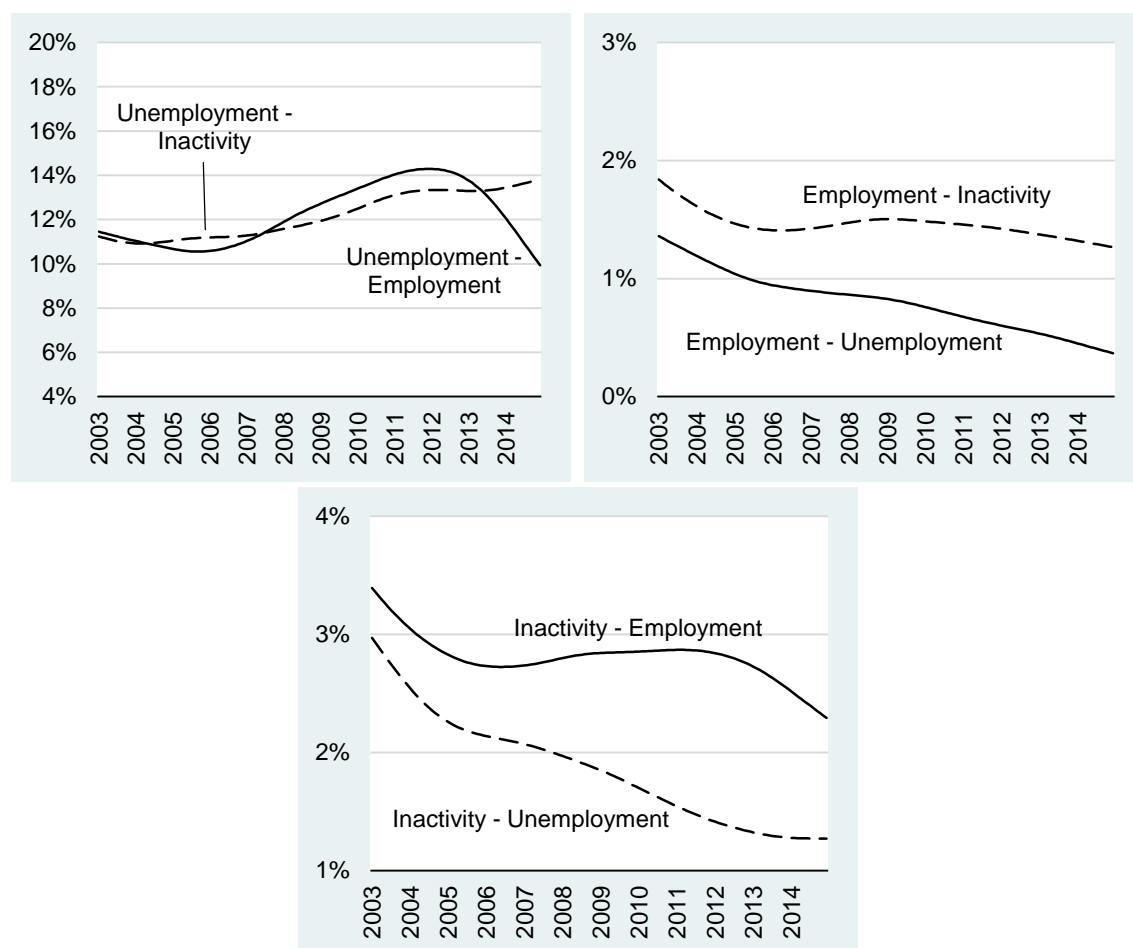
	US (CPS)	Brazil (original)	Brazil (corrected)
<i>Mean</i>			
E → U	0.015	0.012	0.008
E → I	0.029	0.040	0.015
U → E	0.260	0.162	0.122
U → I	0.221	0.294	0.121
I → E	0.046	0.059	0.028
I → U	0.026	0.042	0.018
<i>Volatility</i>			
E → U	0.0025	0.0042	0.0027
E → I	0.0026	0.0038	0.0015
U → E	0.0405	0.0184	0.0198
U → I	0.0240	0.0311	0.0176
I → E	0.0040	0.0057	0.0027
I → U	0.0035	0.0118	0.0050

Note: Average transition rates and standard deviation from 2003:1 to 2014:12 for Brazil. Correction for response errors follow the algorithm proposed in Chapter 4. For the US, transitions rates and standard deviation from 1976:2 to 2011:6.

Sources: Author's calculations based on PME (IBGE) for the Brazilian data and Gomes (2015) for the US.

Figure 2 displays the observed separation and finding rates and the transition rates into and out of inactivity. The top left panel shows that the transition rate $U \rightarrow I$ has been rising since 2005, while the job finding rate has risen between 2006 and 2012, and declined afterwards. Other important trends are the continuous decline in the separation rates (i.e. the transitions $E \rightarrow U$ and $E \rightarrow I$) and the sharp fall in the transitions rates $I \rightarrow U$ during the entire analyzed period. Finally, the decline in transitions from inactivity into employment after 2012 is also worth noting. We can state that 2012 seems to be a turning point in the movements of some relevant transition rates.

Figure 2 – Monthly transition rates: unemployment, employment and inactivity



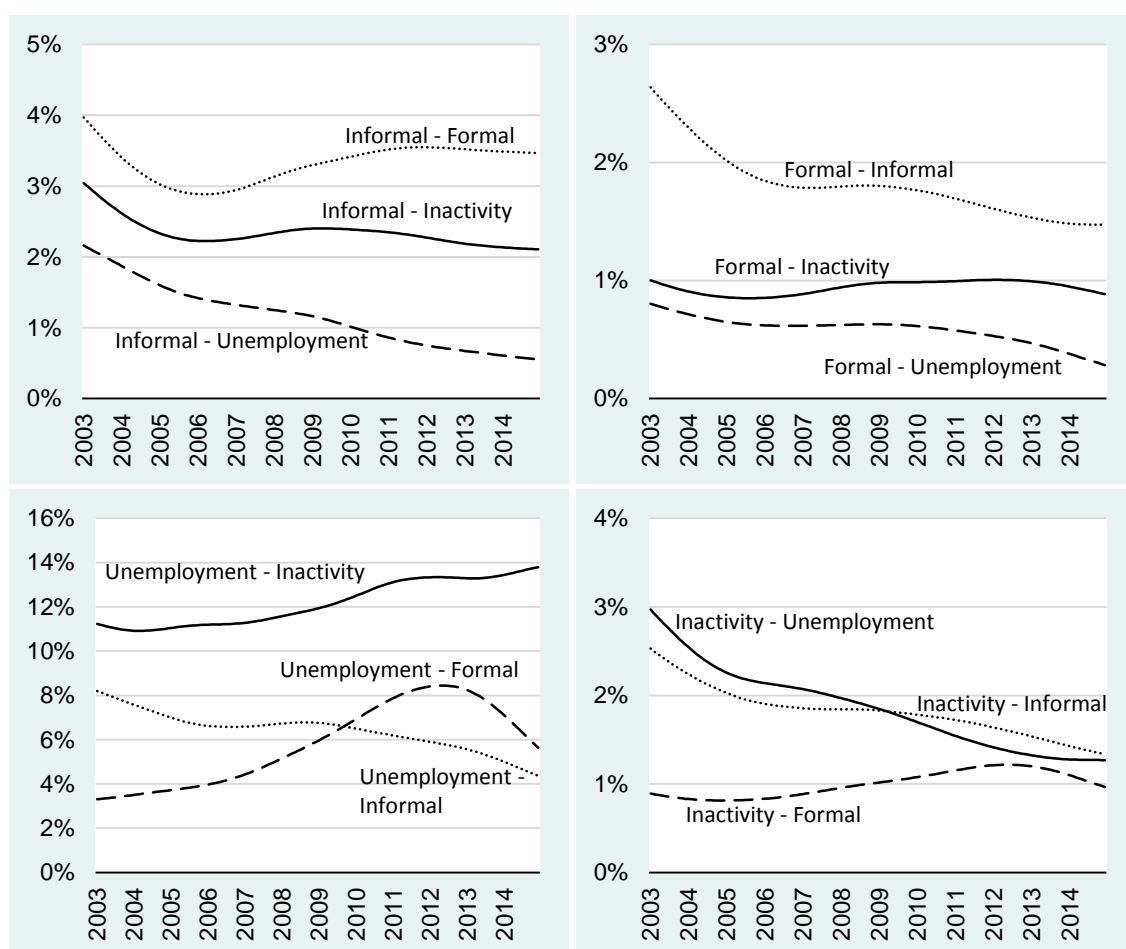
Note: The six monthly transition rates are derived from PME (IBGE) and smoothed using an HP filter with parameter 14.400. Response errors are corrected using the algorithm proposed in Chapter 4.

One important feature of developing economies is the presence of a large informal sector in the labor market. As Figure 3 shows, transitions into and out of formal employment have evolved in remarkably different paths when compared to transitions into and out of informal employment. If we followed the standard methodology of the literature and focused only on the traditional three states (considering employment as a whole), the relevant interactions between formality and informality would be missed.

Figure 3 breaks down the same transitions shown in Figure 2, but it considers formal and informal employment as two different states (we consider self-employment as informal and

employers as formal employment). For instance, the transition $U \rightarrow E$ has two distinct dynamics when the formal-informal dichotomy is considered (bottom left panel). The difference may be observed comparing top left panel of Figure 11. The informal job finding rate has been falling since 2006 and is still following this trend, whereas the formal job finding rate rose sharply since 2006 but since 2012 is rising by a smaller pace. Similar patterns are exhibited by other transitions: formal and informal employment have trailed opposing paths in the last decade. This is particularly relevant because, as we will show, the formalization of the labor market accounts for a large share of the decline in the unemployment rate.

Figure 3 – Monthly transition rates: unemployment, formal/informal employment and inactivity



Note: The six monthly transition rates are derived from PME (IBGE) and smoothed using an HP filter with parameter 14.400.

3. Using flows to assess unemployment dynamics

We begin describing a two-state model of employment status exactly as Petrongolo and Pissarides (2008) and Smith (2011). At any time t individuals can either be employed (E_t) or unemployed (U_t). The unemployment rate is, by definition, $u_t = U_t / (U_t + E_t)$ and evolves according to:

$$\dot{u}_t = s_t e_t - f_t u_t = s_t(1 - u_t) - f_t u_t \quad (1)$$

where e_t is the employment rate, f_t is the job finding rate and s_t is the separation rate at time t (or, respectively, the instantaneous time unemployment outflow and inflow rates). In steady state $\dot{u}_t = 0$, so the unemployment rate can be expressed as:

$$u_t = \frac{s_t}{s_t + f_t} \quad (2)$$

The advantage of using this expression to calculate the unemployment rate is that one can decompose changes in the unemployment rate according to changes in the inflow or outflow rates.

In order to account for inactivity, we continue to follow both Petrongolo and Pissarides (2008), Smith (2011) and also Shimer (2012). In a three-state world where individuals can also be out of the labor force (I), the dynamics of each state can be described by the following equations:

$$\dot{U}_t = \lambda_t^{EU} E_t + \lambda_t^{IU} I_t - (\lambda_t^{UE} + \lambda_t^{UI}) U_t \quad (3)$$

$$\dot{E}_t = \lambda_t^{UE} U_t + \lambda_t^{IE} I_t - (\lambda_t^{EU} + \lambda_t^{EI}) E_t \quad (4)$$

$$\dot{I}_t = \lambda_t^{UI} U_t + \lambda_t^{EI} E_t - (\lambda_t^{IU} + \lambda_t^{IE}) I_t \quad (5)$$

where λ_t^{AB} describes the transition rate at time t between states A and B .

In steady state, $\dot{U}_t = \dot{E}_t = 0$, so we can rearrange (3) and (4) and express U_t as a function of E_t and the transition probabilities:

$$(\lambda_t^{IE} \lambda_t^{UE} + \lambda_t^{IE} \lambda_t^{UI} + \lambda_t^{IU} \lambda_t^{UE}) U_t = (\lambda_t^{IU} \lambda_t^{EI} + \lambda_t^{IU} \lambda_t^{EI} + \lambda_t^{IE} \lambda_t^{EU}) E_t \quad (6)$$

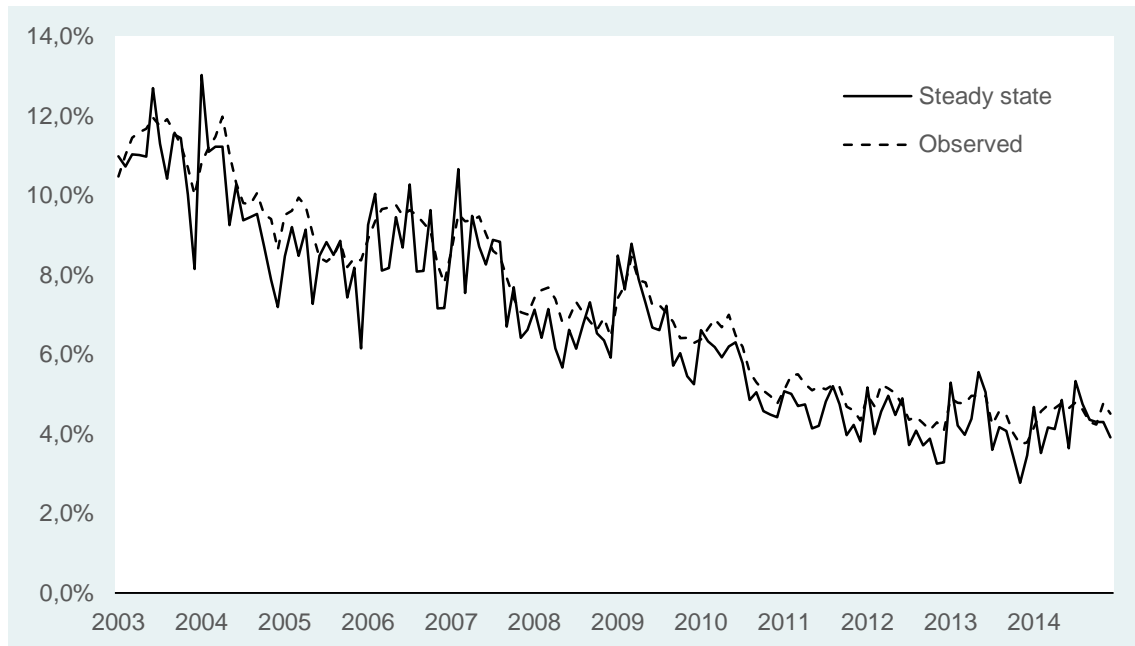
Recalling that $u_t = U_t / (U_t + E_t)$, we can express the steady state unemployment rate only in terms of the six transitions rates:

$$u_t = \frac{\lambda_t^{EU} + \frac{\lambda_t^{IU}}{\lambda_t^{IU} + \lambda_t^{IE}} \lambda_t^{EI}}{\lambda_t^{EU} + \frac{\lambda_t^{IU}}{\lambda_t^{IU} + \lambda_t^{IE}} \lambda_t^{EI} + \lambda_t^{UE} + \frac{\lambda_t^{IE}}{\lambda_t^{IU} + \lambda_t^{IE}} \lambda_t^{UI}} \quad (7)$$

The expression above is essentially similar to equation (2). The numerator accounts for the separation rates from employment to unemployment and to inactivity, but the latter needs to be ‘rescaled’ in order to have its impact in the unemployment rate appropriately measured. The same rationale applies to the finding rate, which now should be thought of as the share of unemployed that leaves unemployment (either to employment or to inactivity).

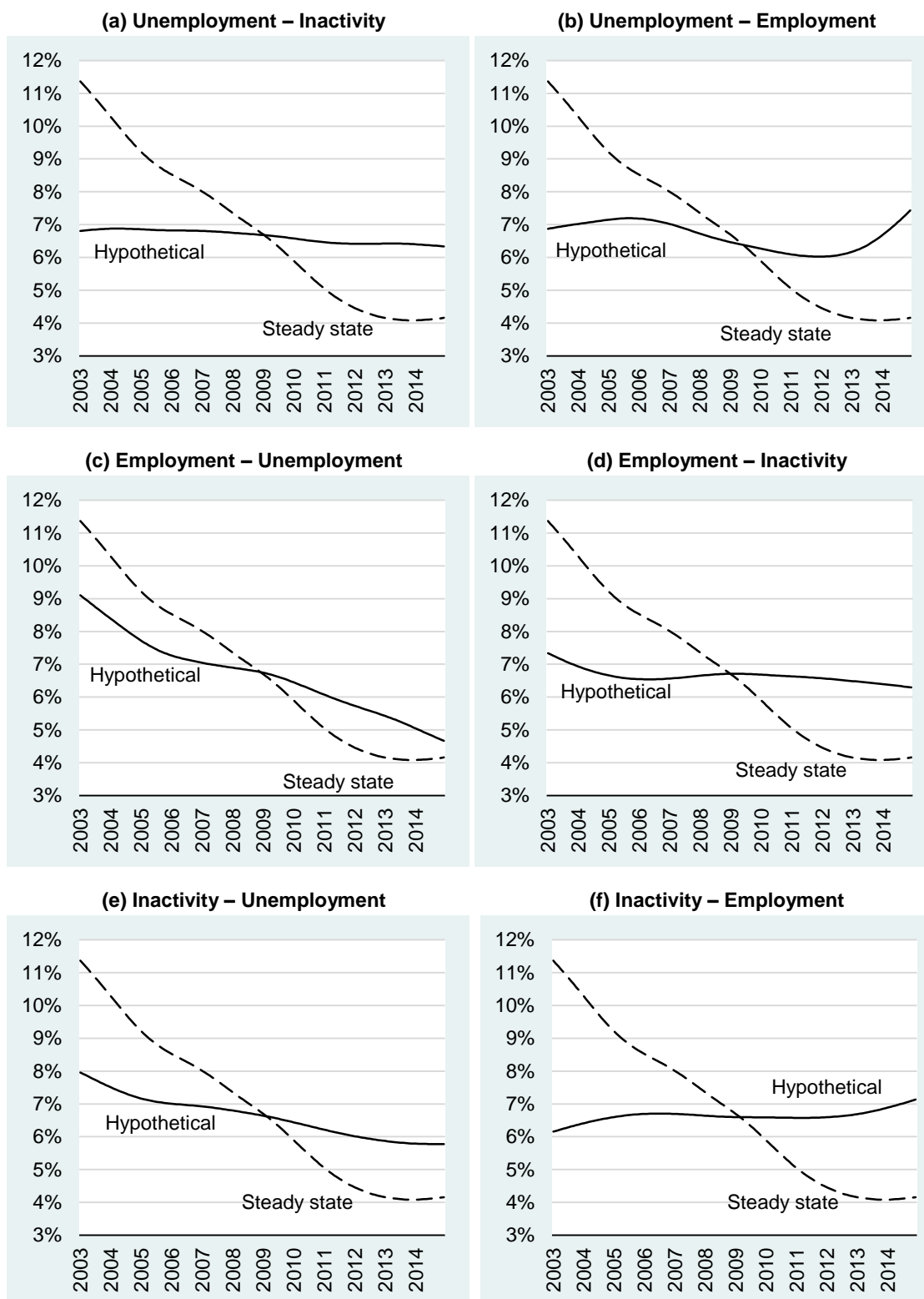
Using the measures of the transition rates described in the previous section, we are now able to construct a series for the steady state unemployment rate, as depicted in Figure 4. The adopted method replicates well the actual unemployment rate: the correlation between the actual and the steady state unemployment rates is 0.965.

Figure 4 – Unemployment rate



Note: Steady state unemployment is calculated according to Eq. (7) using the six monthly transition rates derived from PME (IBGE) data and corrected for response error as described in Chapter 4.

Figure 5 – Contribution of fluctuations in transition rates to the unemployment rate



Note: Each panel show the steady state unemployment rate (dashed line) and the hypothetical unemployment if there were only fluctuations in each subtitled transition rate (all others are held constant and equal to their 2003-2014 average). All calculations are based on PME (IBGE) data. Transitions rates are smoothed using an HP filter with parameter 14.400.

We follow Shimer (2012) to measure the contribution of the changes in each of the transition rates to the change in the unemployment rate. The method is simple: first, we construct the hypothetical unemployment rate using equation (7) holding every transition rate constant (and equal to their historical average), allowing only the transition of interest to change. For instance, the hypothetical unemployment rate had only the transition E→U changed is given by:

$$u_t^{EU} = \frac{\bar{\lambda}^{EU} + \frac{\bar{\lambda}^{IU}}{\bar{\lambda}^{IU} + \bar{\lambda}^{IE}} \bar{\lambda}^{EI}}{\lambda_t^{EU} + \frac{\bar{\lambda}^{IU}}{\bar{\lambda}^{IU} + \bar{\lambda}^{IE}} \bar{\lambda}^{EI} + \bar{\lambda}^{UE} + \frac{\bar{\lambda}^{IE}}{\bar{\lambda}^{IU} + \bar{\lambda}^{IE}} \bar{\lambda}^{UI}} \quad (8)$$

where $\bar{\lambda}^{AB}$ is the average transition rate A→B between January 2003 and December 2014. We do this for every transition rate between the three states u_t^{EU} , u_t^{EI} , u_t^{UE} , u_t^{UI} , u_t^{IE} , u_t^{IU} . This exercise yields the series shown in Figure 5.

Clearly, the main drivers of the decline in unemployment in the past decade were the decline in the E→U (panel c) and in the I→U (panel e) movements. On the other hand, it is worth noting that the job finding rate (U→E and I→E transitions, panels b and f) have both recently contributed to an increase in unemployment, even though this contribution has been offset by the movements in the other transition probabilities and we have not seen unemployment rise (at least not until the end of 2014).

Next, we calculate the contribution of each transition rate as $\beta_{AB} = \frac{cov(u_t^{AB}, u_t)}{var(u_t)}$, where u_t is defined in equation (7) and u_t^{AB} is defined in equation (8). Notice that β_{AB} is the coefficient of the regression of the hypothetical unemployment on the steady state unemployment rate². The results of this decomposition are shown in Table 3³.

The first column of the table below shows each β for the entire period (2003-2014), without any correction to the data. The second column shows the results after correcting for response bias and will be the subject of our analysis. The results confirm what we described based on the visual inspection we had done when analyzing Figure 5: transitions from employment into unemployment (inflows) contributed to roughly 48% of the decline in unemployment, whereas transitions from inactivity into unemployment accounted for approximately 27%. The third most relevant transition was U→E, which contributed to 13% to the change in the unemployment rate, but this contribution was partially offset by the I→E movements (-5%). Results do not change significantly if we don't correct for response bias, except for the contribution of I→U, which is higher in the original (uncorrected) data.

² This is not an exact decomposition; the sum of all betas for the entire period 2003:1 – 2014:12 is 0.970.

³ Other authors, such as Petrongolo and Pissarides (2008), adopt a slightly different approach. They decompose Δu_t into Δs_t and Δf_t and only then apply the beta approach to the Δ series. This method does not measure the contribution of each of the transitions to Δu_t ; rather, it yields the contribution of each of the transition rates variation to the unemployment rate **volatility**. Even though they clarify what they are measuring, other papers that followed (such as X and Y) seem to miss the difference.

We argue that this result is both expected and imprecise, because a share of these transitions are spurious.

Table 3 – Decomposition of the change in the unemployment rate

Transition	2003-2014 (Original)	2003-2014 (Corrected)
E → U	0,412	0.481
E → I	0,061	0.065
U → I	0,093	0.080
U → E	0,058	0.126
I → U	0,447	0.266
I → E	-0,076	-0.048

Note: Each row shows the covariance of u_t^{AB} and u_t divided by the variance of u_t . All calculations are based on PME (IBGE) data. Transitions rates are smoothed using an HP filter with parameter 14.400. The first column shows the results for the original dataset, while second shows the results after correcting for response errors following the algorithm described in Chapter 4.

In short, the results for the three-state model highlight two key factors that explain the decline in the unemployment rate between 2003 and 2014 in Brazil. First, contrary to what Shimer (2012) documented for the US, the ‘ins win’ in Brazil: inflows into unemployment were more relevant to explain the unemployment rate movements in the past decade than the outflows (given by job finding rates). Given that in this period there was a significant decline in unemployment, an appropriate expression for this phenomenon would be *employment hoarding*, a situation where workers are less likely to move into unemployment. It is worth mentioning that employment hoarding is different from job hoarding: an increase in the attachment to employment does not (necessarily) mean that the attachment to the same job increased (in fact, as Chapter 2 shows, job turnover increased dramatically, i.e. job-to-job transitions increased). It means only that workers are less prone to return to the condition of being unemployed.

Second, the *decrease in the participation rate* that happened through *less frequent movements into the workforce*, also played a key role in the unemployment rate movement. These two phenomena are perhaps two sides of the same coin: if there are less persons entering the workforce, those who want to work will be more likely to be employed, or perhaps if workers have increased their attachment to employment, there is less room for new entrants (we cannot establish any causality).

4. Accounting for informality

We now propose a model with four states. In this specific application, we consider state 0 (N_0) to be formal employment and state 1 (N_1) to be the informal sector, while unemployment (U) and inactivity (I) are the other possible positions. It is worth mentioning that in advanced economies the differentiation might also be useful in other labor market dimensions, such as part-time and contingent work.

The procedure is the same as before. Unemployment, formal and informal employment and inactivity evolve according to:

$$\dot{U}_t = \lambda_t^{0U} E_t^0 + \lambda_t^{1U} E_t^1 + \lambda_t^{IU} I_t - (\lambda_t^{U0} + \lambda_t^{U1} + \lambda_t^{UI}) U_t \quad (9)$$

$$\dot{E}_t^0 = \lambda_t^{U0} U_t + \lambda_t^{I0} I_t + \lambda_t^{10} E_t^1 - (\lambda_t^{0U} + \lambda_t^{0I} + \lambda_t^{01}) E_t^0 \quad (10)$$

$$\dot{E}_t^1 = \lambda_t^{U1} U_t + \lambda_t^{I1} I_t + \lambda_t^{01} E_t^0 - (\lambda_t^{1U} + \lambda_t^{1I} + \lambda_t^{10}) E_t^1 \quad (11)$$

$$\dot{I}_t = \lambda_t^{UI} U_t + \lambda_t^{0I} E_t^0 + \lambda_t^{1I} E_t^1 - (\lambda_t^{IU} + \lambda_t^{I0} + \lambda_t^{I1}) I_t \quad (12)$$

In steady state, the flows in and out each type of employment are equal, as are the flows in and out of unemployment: $\dot{U}_t = \dot{E}_t^0 = \dot{E}_t^1 = 0$. Similarly to the case of three states, by manipulating (9), (10) and (11) we get:

$$(\lambda_t^{I0} (\lambda_t^{U0} + \lambda_t^{U1} + \lambda_t^{UI}) + \lambda_t^{IU} \lambda_t^{U0}) U_t = (\lambda_t^{IU} (\lambda_t^{01} + \lambda_t^{0U} + \lambda_t^{0I})) E_t^0 + (\lambda_t^{I0} \lambda_t^{1U} - \lambda_t^{IU} \lambda_t^{10}) E_t^1 \quad (13)$$

$$(\lambda_t^{I1} (\lambda_t^{U0} + \lambda_t^{U1} + \lambda_t^{UI}) + \lambda_t^{IU} \lambda_t^{U1}) U_t = (\lambda_t^{IU} (\lambda_t^{10} + \lambda_t^{1U} + \lambda_t^{1I})) E_t^1 + (\lambda_t^{I1} \lambda_t^{0U} - \lambda_t^{IU} \lambda_t^{01}) E_t^0 \quad (14)$$

The equations above are similar with each other and are also the four-state equivalent to equation (6). Because of the notation burden, we rewrite them simply as:

$$\tilde{f}U = \tilde{s}_0 E^0 + \tilde{s}_1 E^1 \quad (15)$$

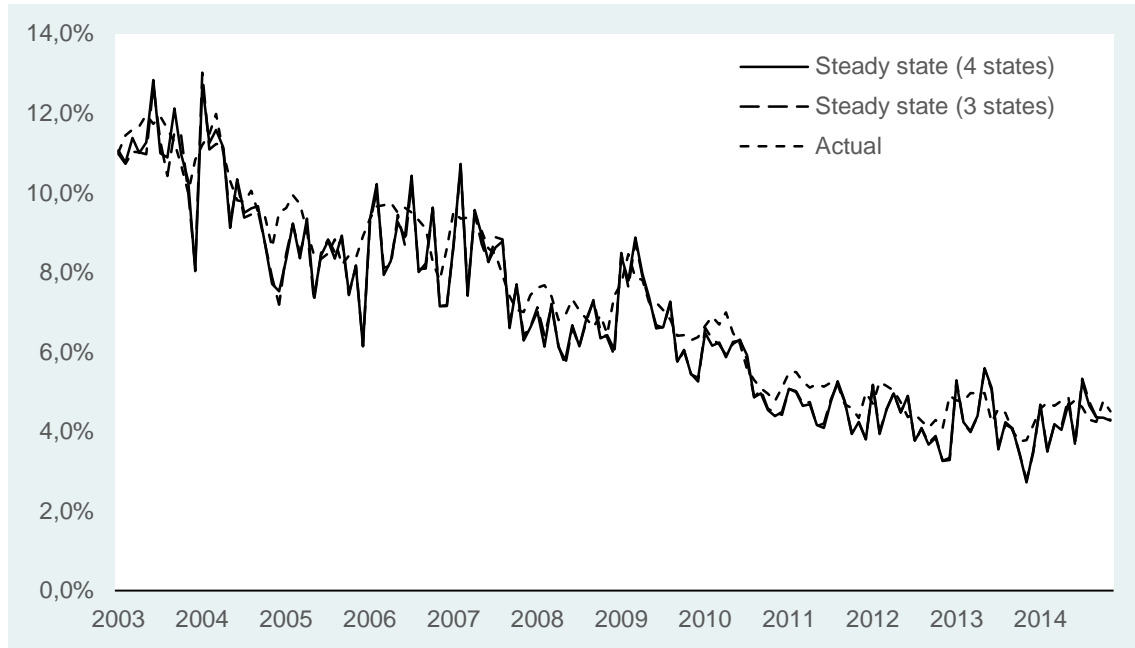
$$\hat{f}U = \hat{s}_1 E^1 + \hat{s}_0 E^0 \quad (16)$$

where the subscript t is omitted, f refers to job finding rates and s refers to separation rates. Manipulating them the equations again we obtain the unemployment rate only in terms of the twelve transition rates at time t :

$$u = \frac{\tilde{s}_0 \left(\frac{\tilde{f}\hat{s}_1 - \hat{f}\tilde{s}_1}{\tilde{f}\tilde{s}_0 - \hat{f}\hat{s}_0} \right) + \tilde{s}_1}{\tilde{s}_0 \left(\frac{\tilde{f}\hat{s}_1 - \hat{f}\tilde{s}_1}{\tilde{f}\tilde{s}_0 - \hat{f}\hat{s}_0} \right) + \tilde{s}_1 + \tilde{f} \left(1 + \left(\frac{\tilde{f}\hat{s}_1 - \hat{f}\tilde{s}_1}{\tilde{f}\tilde{s}_0 - \hat{f}\hat{s}_0} \right) \right)} \quad (17)$$

which is the four-state model equivalent to equations (2) and (7).

Figure 6 – Unemployment rate

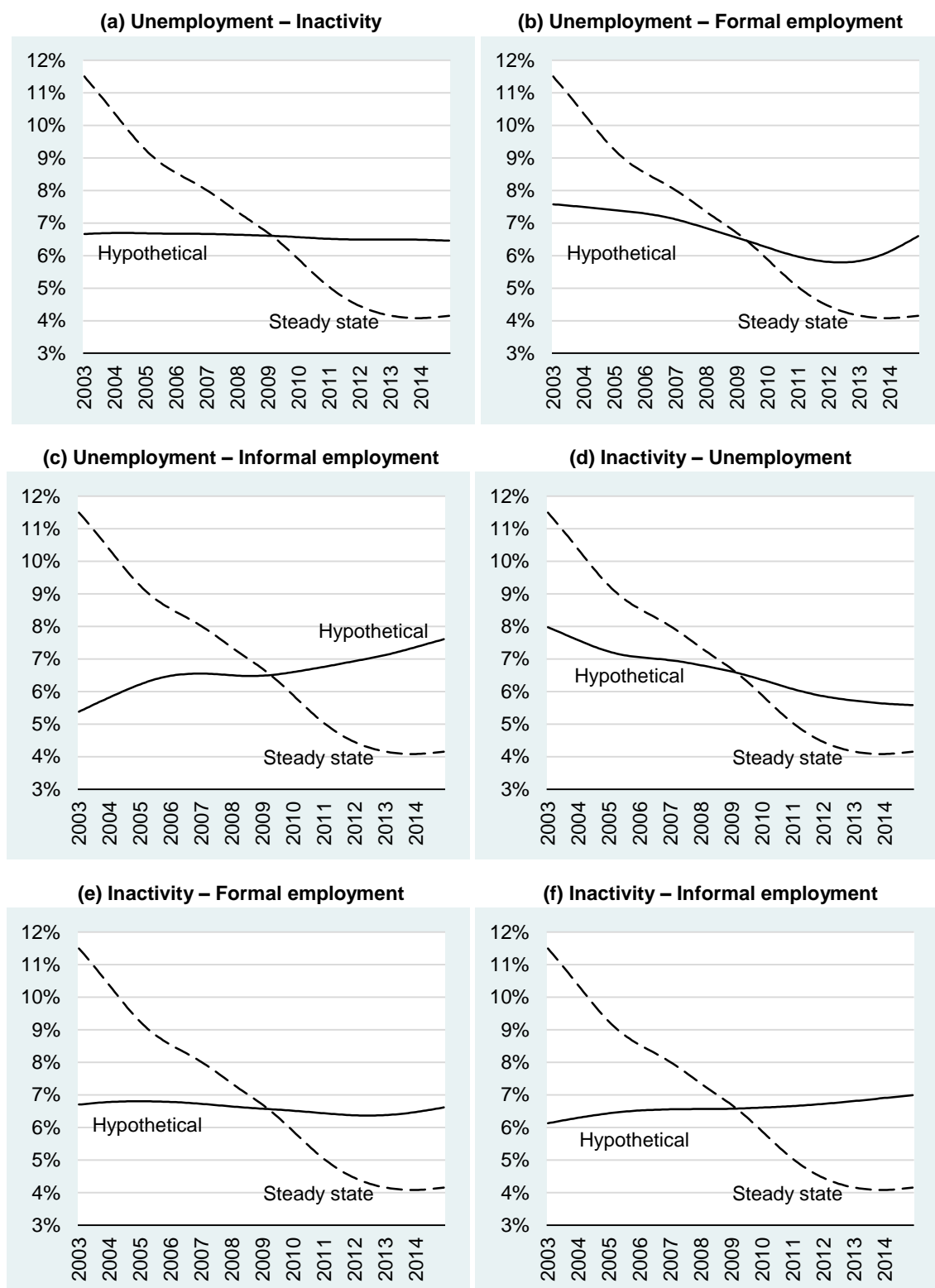


Note: Steady state unemployment is calculated according to Eq. (7) for the 3-state model and Eq. (17) for the 4 state-model, using the monthly transition rates derived from PME (IBGE) data after correcting for response bias. The correlation between both steady state series is 0.999.

Once again we follow Shimer (2012) and replicate the procedure from the previous section to measure the contribution of the changes in each of the transition rates to the change in the unemployment rate. We now have twelve hypothetical unemployment rates: u_t^{0U} , u_t^{0I} , u_t^{01} , u_t^{1U} , u_t^{1I} , u_t^{10} , u_t^{U0} , u_t^{U1} , u_t^{UI} , u_t^{I0} , u_t^{I1} and u_t^{IU} , where we remind that 0 refers to formal employment and 1 refers to informal employment. The series are shown in Figure 7.

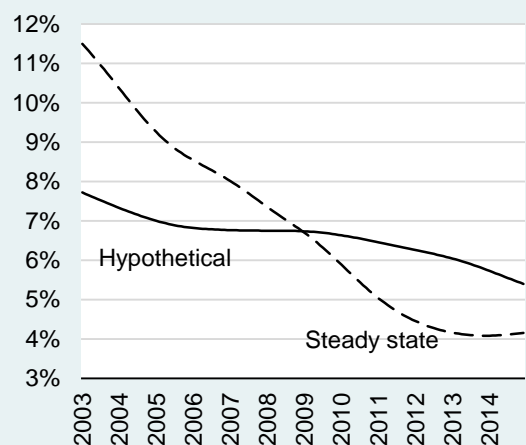
In Section 3, our results showed that the transitions from unemployment into employment contributed little to the change in the unemployment rate (roughly 13%), i.e. the job finding rate did not play a key role to explain the decline in unemployment. Moreover, the key conclusion was that the inflow into unemployment from employment was the most relevant factor for the decline in the unemployment rate.

Figure 7 – Accounting for informality to assess the contribution of fluctuations in transition rates to the unemployment rate

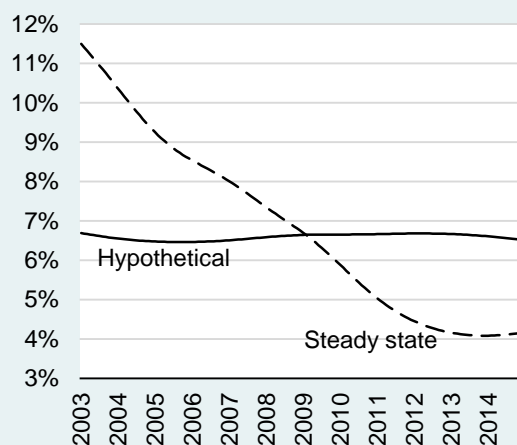


[continued on the next page]

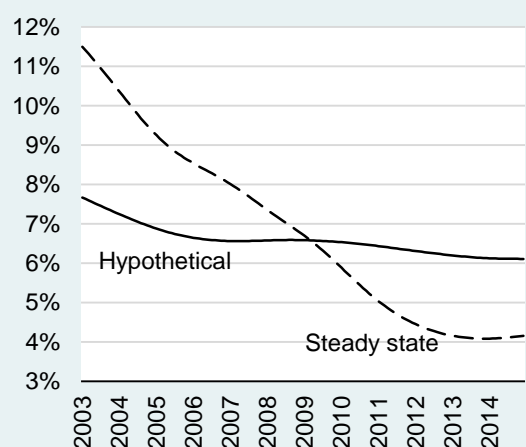
(g) Formal employment – Unemployment



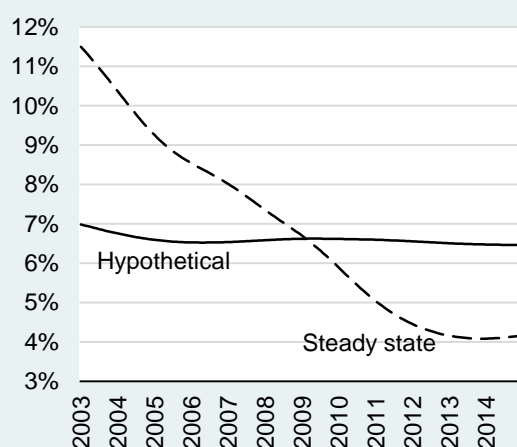
(h) Formal employment – Inactivity



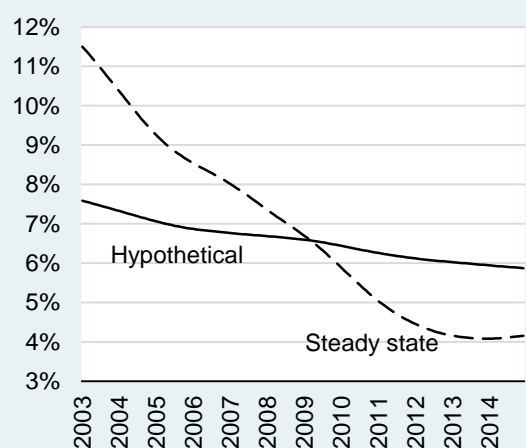
(i) Formal employment – Informal employment



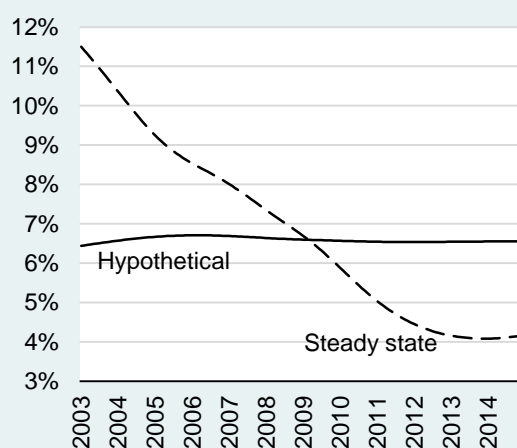
(j) Informal employment – Inactivity



(k) Informal employment – Unemployment



(l) Informal employment – Formal employment



Note: Each panel show the steady state unemployment rate (dashed line) and the hypothetical unemployment if there were only fluctuations in each subtitled transition rate (all others are held constant and equal to their 2003-2014 average). All calculations are based on PME (IBGE) data. Transitions rates are smoothed using an HP filter with parameter 14.400.

With the four-state model, however, we conclude differently as Figure 7 and Table 4 show. First, the transition that contributed the most to the change in the unemployment rate is $I \rightarrow U$, suggesting that the participation rate played a key role in the unemployment rate dynamics. The second most relevant contribution comes from $U \rightarrow \text{Formal}$, i.e. the formal job finding rate. Remember that in the previous section the (aggregate) job finding was not especially relevant, but this was actually the aggregation of three different and large effects: First, the $U \rightarrow \text{Formal}$ movements accounted for 26% of the change in the unemployment rate, and this was because the formal job finding rate increased between January/2003 (3.6%) and December/2014 (5.7%). Second, $U \rightarrow \text{Informal}$ contributed with -20% to the decline in unemployment, meaning that unemployment would have risen by a significant amount due to the decline in the informal job finding rate. Third, we have the 15% contribution to the decline in the unemployment rate from the transition $\text{Formal} \rightarrow \text{Informal}$.

The reduction in the $\text{Formal} \rightarrow \text{Informal}$ transition probabilities and its contribution to the unemployment reduction is a second-order effect, but with a high magnitude. The smaller frequency at which a worker moves from formal to informal employment itself plays a key role in the unemployment dynamics, possibly because once remaining formal, the likelihood of becoming unemployed is much smaller (see Figure 3).

Our methodology does not allow to assess any causality, but these results indicate that either unemployment fell because of the formalization that took place in the Brazilian labor market, or conversely, the decline in unemployment induced the increase in the formal share of the market. In any case, we can establish a strong correlation between formalization and the reduction in the unemployment rate, which by itself is relevant for policy-makers. Further research is necessary to clarify the mechanics behind this phenomenon.

It is also worth pointing out that while $\text{Formal} \rightarrow U$ and $U \rightarrow \text{Formal}$ both contributed positively to the decline in unemployment, the respective movements from and into informal jobs have opposite signs. That is, in the formal sector it became both more likely to find a job *and* to remain in employment, whereas in the informal sector it is now less likely to find a job, but those that have one are more likely to keep it⁴.

We can therefore summarize the forces that led to changes in the Brazilian unemployment rate between 2003 and 2014 in three main factors: (a) *Employment hoarding*, with 43% (the sum of $\text{Informal} \rightarrow U$, $\text{Informal} \rightarrow I$, $\text{Formal} \rightarrow U$ and $\text{Formal} \rightarrow I$); (b) *Participation effect* which involves all transitions from and into inactivity but is well represented by the sum of $I \rightarrow U$ and $U \rightarrow I$ contributions (both accounting to roughly 36%); and finally (c) *Formal job finding rate* was very relevant with a total contribution of 32% (the sum of $\text{Informal} \rightarrow \text{Formal}$, $U \rightarrow \text{Formal}$ and $I \rightarrow \text{Formal}$).

⁴ We loosely interpret here the informal job exit rate as separation rates. It is worth pointing out that we do not consider job-to-job flows, so using the term separation rate is imprecise. Still, the literature employs such terms and we follow the same standard.

Table 4 – Accounting for informality in the decomposition of the change in the unemployment rate

Transition	2003-2014 (Original)	2003-2014 (Corrected)
Informal → U	0,211	0.205
Informal → I	0,018	0.032
Informal → Formal	0,015	0.011
Formal → U	0,139	0.211
Formal → I	-0,035	-0.018
Formal → Informal	0,278	0.152
U → I	0,035	0.034
U → Formal	0,198	0.258
U → Informal	-0,317	-0.201
I → U	0,454	0.300
I → Formal	0,091	0.060
I → Informal	-0,103	-0.081

Note: Each row shows the covariance of u_t^{AB} and u_t divided by the variance of u_t . All calculations are based on PME (IBGE) data. Transitions rates are smoothed using an HP filter with parameter 14.400. Both Informal and formal refer to types of employment. The first column shows the results for the original dataset, while second shows the results after correcting for response errors as described in Chapter 4.

Notice that the magnitude of the contributions of the inflows from formal employment is similar to the ones from the outflows. Thus, we can conclude that both inflows and outflows are relevant to the changes in the unemployment rate. However, differently from what the three-state model analysis indicated, the single most important contribution does not come from the inflows from employment, but rather from inflows from inactivity.

5. Participation and formalization rates

Assuming steady state and based on equations (12), (13) and (14), we can write each state as a function of the transition rates and other states (we again suppress the subscript t for simplicity):

$$E^0 = \frac{\tilde{f}\hat{s}_1 - \hat{f}\tilde{s}_1}{\hat{f}\tilde{s}_0 - \tilde{f}\hat{s}_0} E^1 \quad (18)$$

$$U = \frac{\tilde{s}_0 E^0 + \tilde{s}_1 E^1}{\tilde{f}} \quad (19)$$

$$I = \frac{\lambda^{0I} E^0 + \lambda^{1I} E^1 + \lambda^{UI} U}{\lambda_t^{IU} + \lambda_t^{I0} + \lambda_t^{I1}} \quad (20)$$

We define the formality rate as the share of formal employment in total employment as:

$$\varphi_t = \frac{E_t^0}{(E_t^0 + E_t^1)} \quad (21)$$

We also define the participation rate (the share of the working age population that is either employed or unemployed) as:

$$p_t = \frac{(E_t^0 + E_t^1 + U_t)}{(E_t^0 + E_t^1 + U_t + I_t)} \quad (22)$$

We apply the five equations above to build the steady state series for the participation and the formality rates. By adopting the same procedure as we did when analyzing the changes in the unemployment rate, we obtain the betas for each of the twelve transition rates. Results are depicted in Tables 5 and 6.

Table 5 – Decomposition of the change in the participation rate

Transition	2003-2014 (Original)	2003-2014 (Corrected)
Informal → U	-0,343	-0.307
Informal → I	-0,200	-0.574
Informal → Formal	-0,072	-0.107
Formal → U	-0,470	-1.058
Formal → I	0,536	-0.086
Formal → Informal	-0,838	-0.543
U → I	0,443	0.527
U → Formal	-0,404	-0.225
U → Informal	0,610	0.508
I → U	1,285	1.415
I → Formal	-0,858	-0.140
I → Informal	1,322	1.764

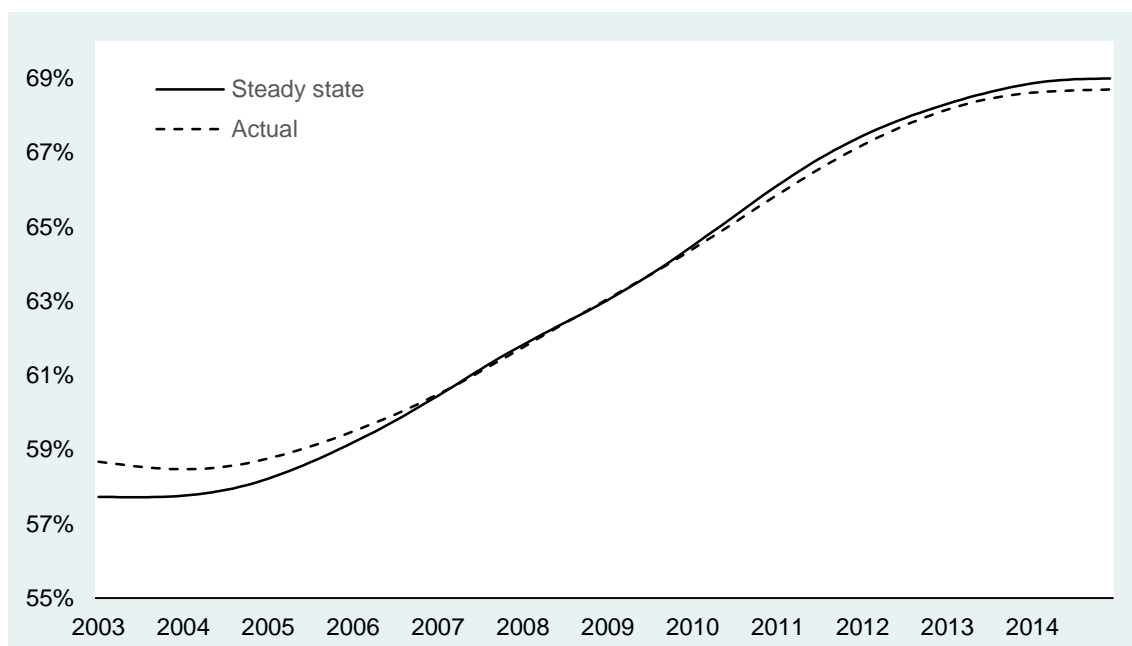
Note: Each row shows the covariance of p_t^{AB} and p_t divided by the variance of p_t . All calculations are based on PME (IBGE) data. Transitions rates are smoothed using an HP filter with parameter 14.400. Informal and formal both refer to types of employment. The first column shows the results for the original dataset, while second one shows the results after correcting for response errors as described in Chapter 4.

Table 5 shows that the coefficients in each row have larger magnitudes than those found in the unemployment rate exercise. Shimer (2012) finds a similar pattern. As for the results, we can classify each row of the table above into first and second order effects. First order effects are those involving transitions from and into inactivity. In this case, it is clear that inflows are dominant: the flow of workers engaging in the workforce has declined and thus the numerator of equation (22) decreases, leading to a smaller p . However, the outflows also play an important role: movements from informal employment into inactivity, for instance,

have become more likely; thus, this rate have contributed (*ceteris paribus*) to an increase in the participation rate.

Second order effects, on the other hand, are those involving transitions from states other than inactivity. They are also relevant to the analysis: U→Formal and Formal→U, for instance, contributed negatively (and significantly) to the decline in p . Because the formal job finding rate has increased and the formal job separation rate has decreased, workers are more likely to be in formal employment; once in formal employment, the likelihood of transiting into inactivity is much smaller.

Figure 8 – Formality rate



Note: Steady state formality rate is calculated according to Eqs. (18)-(21) for the 3-state model and Eq. (17)/ the actual rate is the share of formal employment in total employment. Rates are calculated using the PME (IBGE) panel data after correcting for response bias. Series are smoothed using HP filter with parameter 14,400.

Formalization of the labor market has increased dramatically in the Brazilian metropolitan areas in the past decade (see Figure 8), rising from 57% in 2004 to 70% of total employment in 2014. We once again decompose this change into first and second order effects. The most important contributions come from Formal→Informal and Formal→U transitions, confirming the argument of employment hoarding: Workers became less prone to move out of formality. However, the dynamics of the informal sector seems to have changed as well. Inflows into informality have fallen as a whole, which contributed to the decrease in the informal share. Conversely, the likelihood of transiting from informality to unemployment has decreased, which contributes negatively to the change in the formality rate. Also, second order effects in this case are small, such as the flow U→I. Overall, we can state that it was the *outflow* that were the main driver of the formality increase.

Table 6 – Decomposition of the change in the formality rate

Transition	2003-2014 (Original)	2003-2014 (Corrected)
Informal → U	-0,201	-0,386
Informal → I	-0,011	-0,053
Informal → Formal	0,145	0,165
Formal → U	0,143	0,283
Formal → I	-0,060	-0,043
Formal → Informal	0,394	0,297
U → I	0,053	0,074
U → Formal	0,137	0,225
U → Informal	0,230	0,263
I → U	-0,077	-0,120
I → Formal	0,120	0,129
I → Informal	0,119	0,151

Note: Each row shows the covariance of φ_t^{AB} and p_t divided by the variance of φ_t . All calculations are based on PME (IBGE) data. Transitions rates are smoothed using an HP filter with parameter 14.400. Informal and formal both refer to types of employment. The first column shows the results for the original dataset, while second one shows the results after correcting for response errors as described in Chapter 4.

6. Conclusion

This paper measures both formal and informal job finding rates, employment exit rates, and inactivity's inflow and outflow rates in Brazil from 2003 to 2014. Throughout this period, the decline in the entrance of inactive persons in the workforce and the formalization of the labor market were the main drivers of the decline in the unemployment rate. We found that in particular the reduction of inflows into unemployment were determinant for this trend.

The main contribution of this paper is the extension of the tradition three-state model used to study inflows and outflows contributions to changes in the unemployment rate. We included a fourth state in the model, allowing the separation of employment into two different forms. For instance, in developed countries, employment could be classified as full-time or part-time/contingent work. In the developing world, a useful classification is between formal and informal work.

In Brazil, we showed that formal and informal employment followed different paths and that not accounting for this fact leads to incomplete conclusions: when looking at employment as a whole, the most relevant flow that led the decline in unemployment was the employment exit rate; however, we show that the formal job finding rate increased substantially, while the corresponding informal rate contributed negatively to the drop in the unemployment rate. At the same time, both formal and informal employment exit rates decreased, contributing further to the reduction of the unemployment rate. In sum, while in the three-state approach, the job finding rate did not contribute much to explain the variation in the unemployment rate, we have shown that the dynamics are much more complex (as revealed with the four-state approach).

We must consider, however, that these findings do not allow to establish any causality, and we do not present any theoretical framework to help us do that. Also, we need to take into account that the data is available in Brazil for a short period of time (less than 15 years), and given that this period was particularly positive for the labor market, results should not be generalized without caution. Still, the evidence we presented helps to understand the key movements in the Brazilian labor market during the past decade.

References

- ABOWD, J.; ZELLNER, A. 1985. "Estimating gross labor force flows". *Journal of Business and Economic Statistics*, vol. 3(3), pp. 254-83.
- ATTUY, G. M. 2012. "Decomposição dos ciclos do Desemprego: Uma Aplicação Para o Brasil a Partir dos Fluxos do Trabalho". *Anais do Encontro Nacional de Economia da Anpec*.
- ELSBY, M. W. L.; MICHAELS, R.; SOLON, G. 2009. "The Ins and Outs of Cyclical Unemployment", *American Economic Journal: Macroeconomics*, American Economic Association, vol. 1(1), 84-110, January.
- FUJITA, S.; RAMEY, G. 2009. "The Cyclicalities of Job Loss and Hiring," *International Economic Review*, vol 50, 415-430.
- HALL, R. E. 2005. "Employment Efficiency and Sticky Wages: Evidence From Flows in the Labor Market". *Review of Economics and Statistics*. vol. 87 (3), 397-407.
- SMITH, J. C. 2011. "The Ins and Outs of UK Unemployment". *The Economic Journal*, 121: 402-444.
- MENEZES FILHO, N. A.; CABANAS, P. H. F.; KOMATSU, B. K. 2014. "Tendências Recentes do Mercado de Trabalho Brasileiro". IPEA: Boletim Mercado de Trabalho - Conjuntura e Análise, n. 56, fevereiro.
- PETRONGOLO, B.; PISSARIDES, C. A. 2008. "The Ins and Outs of European Unemployment". *American Economic Review: Papers and Proceedings*, vol. 98 (2), 256-262.
- RIBAS, R.P.; SOARES, S.D. 2008. "O atrito nas pesquisas longitudinais: o caso da pesquisa mensal de emprego do IBGE". IPEA, TD 1347
- SHIMER, R. 2007. "Reassessing the Ins and Outs of Unemployment," NBER Working Papers 13421, National Bureau of Economic Research, Inc.
- SHIMER, R. 2012. "Reassessing the Ins and Outs of Unemployment", *Review of Economic Dynamics*, vol. 15(2), pages 127-148, April.
- ZYLBERSTAJN, E. 2015. "Painéis de rotação e trajetórias inconsistentes: fontes de viés na Pesquisa Mensal de Emprego do IBGE".

Capítulo 2: Estimating the Returns to Education in Brazil: a parametric control function approach

Abstract

This paper investigates the causal effect of education on earnings in Brazil by employing a new method proposed by Klein and Vella (2010) that obtains identification on the presence of conditional heteroskedasticity. In contrast to traditionally used IV methods, this approach yields unbiased estimates in the absence of instruments and allows for interpretation of the coefficients that is not confined to local average treatment effects. Results indicate that the average return to education in Brazil was relatively stable from 1995 to 2003, declined afterwards until 2011 but has risen by one percentage point since then to 12.7% in 2013.

Keywords: return to education, wage equation, control function approach, second moment exclusion restriction

JEL Classification: C3, I21, J31

1. Introduction

Earnings inequalities have declined during the past decade in most of Latin American countries. In Brazil, the Gini coefficient declined from 0.567 in 2001 to 0.495 in 2013⁵. Still, Brazil remains a very unequal country, and recent developments suggest that this trend might be at risk: the Gini coefficient has been stable for the last three years.

Inequality levels in a given country are the result of the returns to individual characteristics and the distribution of these attributes in the population. In particular, the returns to education play a key role to determine earnings and are likely one of the most explored issues in the empirical economics literature. However, to these days several empirical challenges make it difficult to correctly estimate the returns to schooling. This is, in turn, a serious problem researchers and policy makers have to address in order to better understand income inequality dynamics and to foster investments in human capital.

This paper employs a novel technique developed by Klein and Vella (2010)⁶ to estimate the causal effect of schooling on wages in Brazil for the period 1995 to 2013. The difficulty in identifying the causal effects of education on earnings arise from the endogeneity of educational choices to wages. Since Mincer (1974) established a methodology to estimate wage equations, several authors have documented and tried to deal with the OLS bias, usually by employing instrumental variables (IV) strategies (see, for instance, Griliches, 1977; Angrist and Krueger, 1991; Duflo, 2001; Heckman et al. 2006; Card, 1999, in particular, presents a detailed survey on this subject).

Although large in quantity, the empirical literature lacks in robustness of results. Angrist and Krueger (1991) estimate returns to schooling in the US labor market between 6-10%, above the OLS estimates of 5-7%. Oreopoulos (2007) estimates that one extra year of education yields an average increase of 13% in wages, compared to a downward biased OLS estimate of 8%. Carneiro and Lee (2008) and Chen (2008) also make use of IV to estimate an average return to education between 13-15% for American men, but the validity of the instruments they use (proxies of the costs of school attendance) is questioned by authors such as Cameron and Taber (2004).

In Brazil, several studies have estimated the returns to education (see, for instance, Langoni, 1973; Senna, 1976; Tannen, 1991; Barros and Ramos, 1994), but the difficulty in dealing with endogeneity has been a constant issue. More recently, Teixeira and Menezes-Filho (2012) use a national survey with data from 1997 to 2007 and employ an IV strategy to find that returns were between 5.5-9.4%, lower than their OLS estimate of 11.6%.

The KV (2010) strategy uses a control function approach. Identification relies on heteroskedasticity of the error terms, which yields nonlinearity to the control term, allowing for non-biased estimates of the coefficient of interest. As Saniter (2012) points out, this approach makes use of second moment restrictions (variance) instead of first moment exclusion restrictions (inherent to any IV approach) and therefore has two key advantages. First, interpretation of the coefficient of interest is possible for the entire population of interest (i.e. inference is made on the average treatment effect – ATE), while any IV estimate

⁵ According to Instituto Brasileiro de Geografia e Estatística (IBGE).

⁶ Henceforth KV (2010).

only provides the causal effect for the subsample of individuals who are affected by variations in the instrument (compliers). Second, because there is no need for instruments, inference can be made at any moment in time. On the other hand, IV estimates are often also bounded by the time of the variation induced by the instrument. An example of such case is found in Ichino and Winter-Ebmer (2004), who use own and father's World War II involvement as instruments for schooling; thus, their results are estimates for a particular cohort of Germans and cannot be replicated for previous or subsequent cohorts.

However, the implementation of the KV (2010) estimator is not simple. On the contrary, it is difficult because of their semiparametric approach, which brings a heavy computational burden to the estimation procedure. We then turn to Farré, Klein and Vella (2013), who implement a parametric estimator for the KV identification strategy.

To provide a preview of the results found in this paper, the estimated average return to education in Brazil fell from 15.6% in 1995 to 12.7% in 2013. The OLS estimates are downward biased in the entire period by 2.8 percentage points on average.

These results are presented in greater detail in section 4. First, section 2 presents the econometric model and the implementation procedure of the estimator. Section 3 describes the data used in the estimation. Section 5 discusses in greater depth the OLV-IV bias and section 6 presents some evidence to support the robustness of our findings. Section 7 concludes.

2. Empirical exercise

2.1 Econometric model and identification

In this section, mainly drawn from Saniter (2012), we describe the KV (2010) approach. The ultimate goal of any linear model that estimates the causal effect of education on earnings is to obtain consistent estimates for the parameter δ in a wage equation of the following kind:

$$w_i = x_i\beta + \delta s_i + u_i \quad (23)$$

Where w_i represents hourly log wages, s_i represents years of schooling, x_i is a vector of exogenous regressors and u_i the error term for individual i . OLS estimation is infeasible because of the endogeneity bias caused by omitted variables. To better illustrate our estimation strategy, let's first write the education equation and then rewrite our model in a control function setting:

$$s_i = x_i\varphi + v_i \quad (24)$$

where x_i may be (but not necessarily is) identical for both equations. From this point on, we refer to (23) as the wage equation and to (24) as the education equation.

If the covariance between u_i and v_i is different from zero, then we can write $u_i = \lambda v_i + e_i$, where $cov(v_i, e_i) = 0$, and the endogeneity arises. Replacing this expression in equations (23) and (24) provides us the controlled function:

$$w_i = x_i\beta + \delta s_i + \lambda v_i + e_i \quad (25)$$

where v_i is the control term and λ is a measure of the degree of endogeneity. Because v_i is a perfect combination of s_i and x_i , the regressors are collinear and OLS is thus infeasible.

However, we can also write:

$$\lambda = \frac{cov(u, v)}{var(v)} = \frac{cov(u, v)}{\sigma_v \sigma_v} \frac{\sigma_u}{\sigma_u} = \rho \frac{\sigma_u}{\sigma_v} \quad (26)$$

where σ_j ($j=u, v$) denotes the standard deviations of the error terms u and v and $\rho = cov(u, v)/\sigma_u \sigma_v$ is the correlation coefficient between them. Notice that the subscripts are omitted for simplicity.

If the assumption that the errors are heteroskedastic holds, then $\sigma_j = H_j(x_j)$, where H_j is the heteroskedasticity function and $x_j \subseteq x$. Equation (25) can now be rewritten as:

$$w = x\beta + \delta s + \rho \frac{H_u(x_u)}{H_v(x_v)} v + e \quad (27)$$

which is the final estimation equation. Notice that x_u and x_v may be different or identical in both equations.

There are two important identification assumptions in order to properly estimate the equation above. First, $H_u(x_u)/H_v(x_v)$ needs to vary across x , so that regressors are not collinear. KV (2010) has named this assumption as the *variable impact property* (VIP), which simply requires that the heteroskedasticity is present in either one of both equations in such a way that the quotient of the two functions is not constant across x . A good candidate (but not the only one) for providing this is age. What is needed is that the heteroskedasticity due to age in the education equation is different than the one in the wage equation, which is economically plausible: on the one hand, schooling in Brazil has been expanding throughout cohorts and being universalized; on the other hand, we expect heteroskedasticity in the wage errors to increase as age increases due to heterogeneous experience and human capital accumulation (other than education). We show evidence in support of this in section 3.

The second assumption is that the errors correlation must be constant and independent of the regressors, i.e. $\rho = corr(uv|x) = corr(uv) = const$. KV (2010) refer to this as the *constant correlation condition* (CCC). Put differently, the CCC requires that the degree of

endogeneity in the model does not depend on x . As Saniter (2012) notes, “this assumption is inherent to any IV strategy” (Saniter, 2012, p. 9).

2.2 Implementation

The implementation of the KV (2010) estimator can be done either parametrically and non-parametrically. Notably, a non-parametric approach has the advantage of not imposing normality of errors for consistency nor a functional form for the heteroskedasticity functions. KV (2010) describe a semi-parametric method, for which they also provide proof of consistency; their method is implemented by Saniter (2012), Klein and Vella (2009), Schroeder (2010) and Farré, Klein and Vella (2009). Nonetheless, all these papers describe the computational burden of adopting such strategy. Saniter (2012), for instance, who estimated the returns to schooling in Germany, reports having used 500 computer cores at the same time.

Perhaps because of these implementation difficulties, the KV (2010) estimator has been used in only a few studies. Aside from the studies mentioned above, Schroeder (2010) estimates the impact of microcredit borrowing on per-capita household consumption in Bangladesh using conditional second moments. Wang (2010) adapts the KV (2010) estimator to a Chinese database and finds that the returns to education in urban China are in the range 2.3-4.6%, below the OLS estimates of 3.9%-7.3%. Figueirêdo et al. (2014) use the KV (2010) approach to assess the influence of family background in student achievement in ENEM, a national exam that is undertaken when students graduate from high school.

Farré, Klein and Vella (2013) follow a different path. They choose a functional form for the heteroskedasticity functions (that take an exponential form), thus largely simplifying the implementation of the estimator. They also simulate a Monte Carlo exercise in which the true heteroskedastic functions are exponential but the estimation is done parametrically assuming either heteroskedasticity in only one of the equations or a quadratic form for the heteroskedastic functions. In both cases, results show that the estimator performs quite well⁷.

The heavy computational demand in the nonparametric estimation arises for reasons that become clear below and we will highlight them when the method is described. We adopt a more flexible parametric approach than the one proposed by FKV (2013), treating $H_u(x_u)$ and $H_v(x_v)$ as quadratic functions of age plus a linear combination of dummies (gender, race etc.) and their interactions. More specifically, we define the following heteroskedastic function:

$$H_j^2(x_j) = z\theta_j + \varepsilon_j \quad (28)$$

⁷ Also using simulation techniques, Klein and Vella (2010) show that even with little heteroskedasticity the OLS bias is eliminated. These results are encouraging to the use of the estimation technique described in this section.

This approach makes the estimation feasible, with the risk of losing efficiency due to misspecification of the functional form. It should be noted that z may or may not be equal to x . In our benchmark specification, we use all individual characteristics in both equations. The only subtle change that we make is the use of state of residence dummies in the wage equation and state of birth dummies in the education equation. We provide some robustness checks in section 6.

Following FKV (2013), estimation is done following a three-step procedure:

- 1) Estimate \hat{v} through OLS in the education equation (Eq. (24)): $\hat{v} = s - x\hat{\phi}_{OLS}$.
- 2) Using OLS, estimate θ_v from Eq. (28) by regressing \hat{v}^2 on z . Then, compute the standard deviation of the reduced form error as $\hat{H}_v = \sqrt{z\hat{\theta}_v}$.
- 3) Using \hat{v} and \hat{H}_v and the assumed form for $H_u = z\theta_u + \varepsilon_u$, solve the following nonlinear least squares problem⁸:

$$\min_{\beta, \rho, \theta_u} \sum_i^n \left(e_i = w_i - x_i\beta - \delta s_i - \rho \frac{\sqrt{z_i\theta_u}}{\hat{H}_{vi}} \hat{v}_i \right)^2 \quad (29)$$

Solving equation (29) yields the coefficient estimates of interest, particularly $\hat{\delta}$. The optimization was programmed in SAS, using the NLP procedure (Non Linear Programming) and the default Conjugate Gradient Methods (CONGRA) optimization technique. Convergence in step 3 was typically achieved with 300 iterations or less. Notice that if the estimation was to be done non- (or semi-) parametrically, step 3 implies that instead of estimating $\hat{\theta}_u$, one needs to estimate the unknown functions H_j at each iteration, which is precisely the main computational burden in the non-parametric strategy. On the other hand, by imposing the functional form $H_{ij} = \sqrt{z_i\theta_j} + \varepsilon_j$ the optimization of equation (29) is done in approximately two minutes using a common server with standard CPU and RAM specifications.

Estimation was performed for several years of data. Standard errors are defined as square roots of the diagonal elements of the covariance matrix, which in turn is obtained numerically (see Nocedal, 1996).

3. Data

The analysis presented in the next sections are based on the *Pesquisa Nacional por Amostra de Domicílios* (National Household Sample Survey, henceforth PNAD) for the years of 1995 to 2013. PNAD is an annual household survey conducted by the *Instituto Brasileiro de Geografia e Estatística* (IBGE) and it is a nationally representative sample. In the most recent wave (2013), 362,555 individuals were surveyed.

⁸ Once again we remind that in our benchmark specification we have $z = x$ except for state of birth dummies.

We adopt a parsimonious specification, so we use few variables and compatibilization of different waves is straightforward. Nonetheless, we used the *DataZoom* Stata package from PUC-Rio⁹ in order to construct each year's dataset. Because our interest lies on returns to schooling, we restrained our sample to individuals between 25 and 55 years of age (to minimize censored schooling and selection into the labor market biases). Also, we have excluded individuals that did not report any labor income in the week of reference. Individuals with missing values in any of the variables used in the estimation were also dropped. We considered the log hourly wage from the main job as the labor income variable and dropped the top and bottom 1% observations to exclude outliers. Table 7 provides an overview of all the variables used in this study and Table 8 displays some summary statistics for the 2013 dataset.

Table 7 – Variables description

Variable	Description
LNWAGE	Log of hourly wage
YRSEDUC	Years of education
AGE	Years of age at interview
FEMALE	Dummy indicator with value 1 if female and 0 otherwise
WHITE	Dummy indicator for white individuals

Table 8 – Descriptive statistics

Variable	Mean	S.D.	Min	Max
LNWAGE	3.39	0.79	1.28	6.21
YRSEDUC	9.23	4.39	0	17
AGE	38.64	8.55	25	55
FEMALE	0.4310	-	0	1
WHITE	0.4432	-	0	1
N	109,687			

Source: PNAD 2013 (IBGE).

To summarize and facilitate the understanding of our empirical strategy, the final estimation models are outlined below, in the order in which they are estimated:

$$\begin{aligned}
s_i = & \varphi_0 + \varphi_1 Age_i + \varphi_2 AgeSq_i + \varphi_3 Female_i + \varphi_4 White_i + \varphi_5 Female_i \cdot Age_i \\
& + \varphi_6 Female_i \cdot AgeSq_i + \varphi_7 White_i \cdot Age_i + \varphi_8 White_i \cdot AgeSq_i \\
& + \varphi_9 White_i \cdot Female_i + \sum_{m=1}^{26} \tau_m dsb_{im} + v_i
\end{aligned} \tag{30}$$

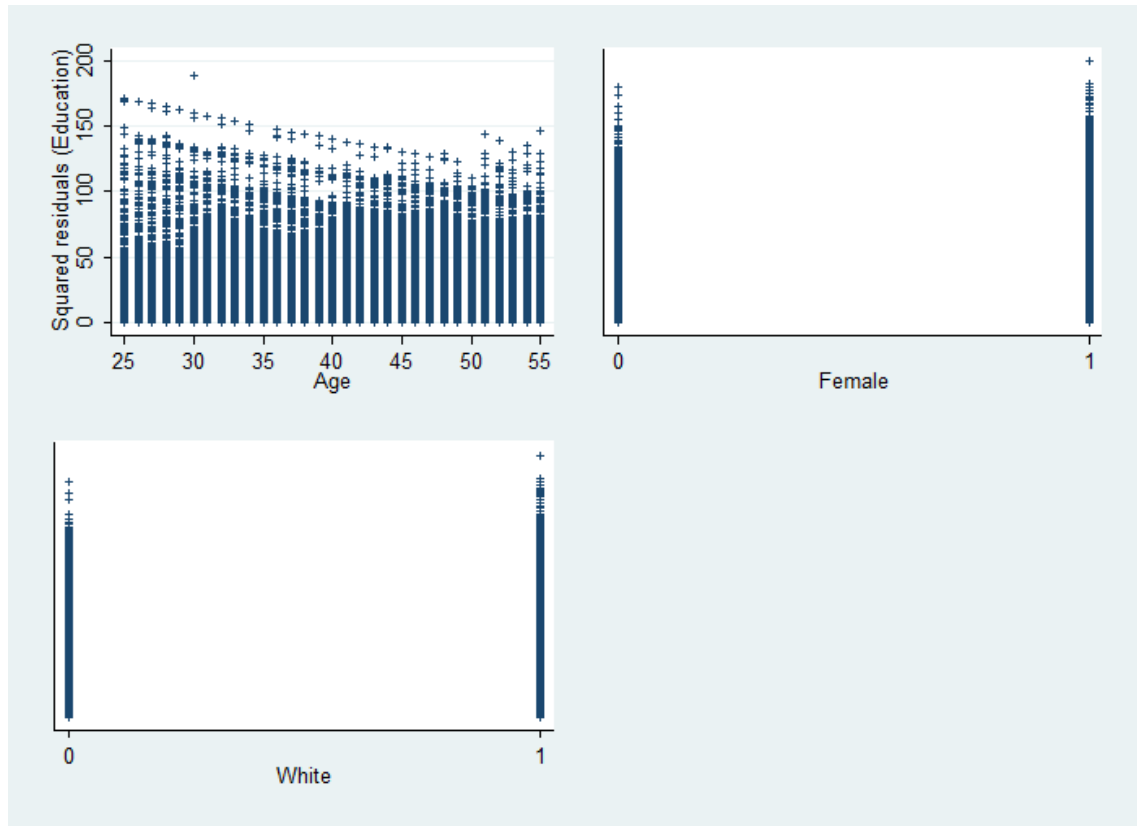
⁹ Available at <http://www.econ.puc-rio.br/datazoom>.

$$\begin{aligned}
H_v^2(x_i) = \hat{v}_i^2 = & \theta_{v0} + \theta_{v1}Age_i + \theta_{v2}AgeSq_i + \theta_{v3}Female_i + \theta_{v4}White_i \\
& + \theta_{v5}Female_i \cdot Age_i + \theta_{v6}Female_i \cdot AgeSq_i + \theta_{v7}White_i \cdot Age_i \\
& + \theta_{v8}White_i \cdot AgeSq_i + \theta_{v9}White_i \cdot Female_i + \sum_{m=1}^{26} \psi_m dsb_{im} + \varepsilon_{vi}
\end{aligned} \quad (31)$$

$$\begin{aligned}
LnWage_i = & \beta_0 + \delta s_i + \beta_1 Age_i + \beta_2 AgeSq_i + \beta_3 Female_i + \beta_4 White_i + \beta_5 Female_i \\
& \cdot Age_i + \beta_6 female_i \cdot AgeSq_i + \beta_7 White_i \cdot Age_i + \beta_8 White_i \cdot AgeSq_i \\
& + \beta_9 White_i \cdot Female_i + \sum_{m=1}^{26} \gamma_m dsr_{im} + \rho \frac{\sqrt{\exp(z_i \theta_u)}}{\hat{H}_{vi}} \hat{v}_i + e_i
\end{aligned} \quad (32)$$

where dss_{im} and dsr_{im} are state of birth and state of residence dummies and $z_i \theta_u = \theta_{u0} + \theta_{u1}Age_i + \theta_{u2}AgeSq_i + \theta_{u3}Female_i + \theta_{u4}White_i + \theta_{u5}Female_i \cdot Age_i + \theta_{u6}Female_i \cdot AgeSq_i + \theta_{u7}White_i \cdot Age_i + \theta_{u8}White_i \cdot AgeSq_i + \theta_{u9}White_i \cdot Female_i + \sum_{m=1}^{26} \zeta_m dsb_{im}$. Note that state of birth dummies were used in the education equation and in the $H_j^2(x_i)$ functions (we consider them exogenous), while state of residence were included as controls for local labor markets in the wage equation. In short, we cautiously adopted a parsimonious specification, with only a small number of strictly exogenous covariates. Also, the heteroskedasticity function is relatively flexible, with a set of dummies, a quadratic function of age and their interactions.

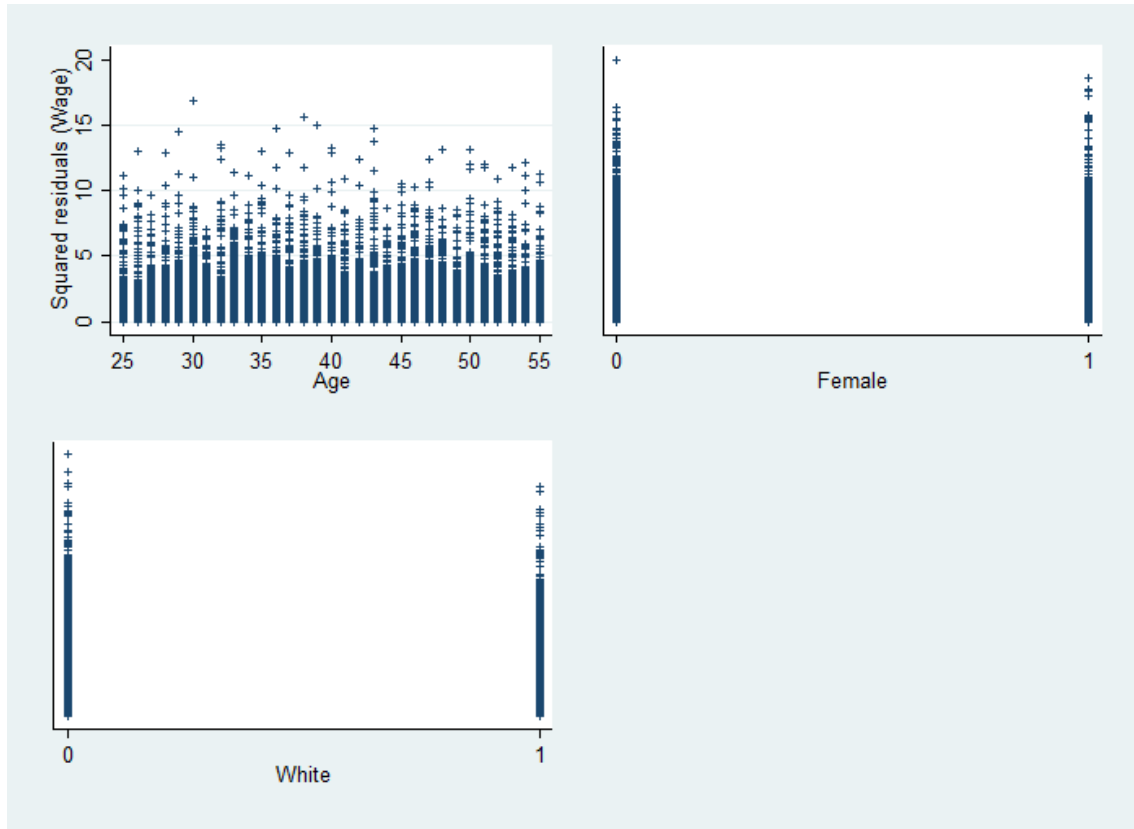
Figure 9 – Education equation: graphical analysis of heteroskedasticity



Source: Author's calculations based on PNAD 2013 (IBGE).

Note: residuals are obtained after the OLS estimation of equation (24).

Figure 10 – Wage equation: graphical analysis of heteroskedasticity



Source: Author's calculations based on PNAD 2013 (IBGE).

Note: residuals are obtained after the optimization of equation (29).

Figure 9 and Figure 10 above provide visual evidence for the presence of heteroskedasticity and, more importantly, also suggest that heteroskedasticity is not constant across x . See, for instance, that the slope of the relationship between age and the squared residuals \hat{v}^2 (Figure 9) and \hat{u}^2 (Figure 10) are different, as is the case with the female dummy. This is crucial because, as highlighted earlier, identification relies on these differences.

4. Results

The three-step procedure outlined in section 2 yields the results displayed in Table 9 and Table 10. We begin displaying in Table 9 the results for the estimation of the education equation, which yielded estimates for \hat{v} . In general, older and non-white individuals are less educated. Also of interest is the formal test for the presence of heteroskedastic errors in the education equation, displayed in the appendix. The test strongly rejects the hypothesis of homoscedastic errors, which is in favor of the identification strategy adopted in this paper. The estimates for $\hat{\theta}_v$ are also displayed in the appendix.

Next, we turn to the wage equation. Table 10 shows the differences between the OLS and the control function estimates for β , with standard errors displayed in parenthesis. As for the coefficients, we see that age affects wages in a linear fashion in the CF setting and that

men and white workers are expected to earn more. The coefficients are, in most cases, smaller in the CF setting.

Table 9 – OLS Estimates: Education equation (2013)

Variable	β_{OLS}
AGE	-0.1235*** (0.0224)
AGESQ	0.0170 (0.0283)
FEMALE	0.6339 (0.5528)
WHITE	1.5938*** (0.5506)
FEMALE*AGE	0.0435 (0.0289)
FEMALE*AGESQ	-0.0442 (0.0365)
WHITE*AGE	-0.0166 (0.0288)
WHITE*AGESQ	-0.0459 (0.0363)
WHITE*FEMALE	-0.3099*** (0.0501)
CONSTANT	14.5316*** (0.4370)
N	109,687
Adjusted R ²	0.1349

Source: Authors' calculations based on PNAD (IBGE). Standard errors in parentheses.

Note: OLS estimates for equation (30). State of birth dummies are also included in the regression.

* significant at 10%; ** significant at 5%; *** significant at 1%.

The main result of this paper is displayed in the first row, where we find an average return of 9.2% for each extra year of education in the OLS case and a 12.7% in the controlled setting, meaning that the OLS model is thus biased downward by 3.5 percentage points. The estimates for $\hat{\theta}_u$ are displayed in the appendix. The other coefficients have none or small changes in the controlled function setting, with the exception of the age coefficient, which is higher in the OLS estimate.

Using in our favor the key advantages of the method adopted in this paper, the same exercise is replicated for the years of 1995 to 2012. Remember that the KV (2010) approach allows to calculate the returns to schooling at any period (independent of any instrument) and for all individuals, not only compliers. The KV (2010) estimator, as we have argued before, measures the ATE, while any IV strategy yields a LATE.

For simplicity, results are displayed graphically in Figure 11. Each year represents the equivalent to the first row of Table 10, which depicts the estimates for 2013. A clear pattern of declining returns to education emerges, either in the OLS or the CF setting. Moreover, when comparing 1995 to 2013 one finds similar declines in return rates: 3.0 and 2.9 respectively. However, recent patterns diverge: the CF estimation unveils an important and

more pronounced increase in returns since 2011 when compared to the OLS estimates: 1.0 p.p. (controlled) versus 0.3 p.p. (OLS).

Table 10 – OLS and Control function estimates: Wage equation (2013)

	β_{OLS}	β_{CF}
YrsEDUC	0.0919*** (0.0005)	0.1269*** (0.0022)
AGE	0.0487*** (0.0035)	0.0100*** (0.0023)
AGESQ	-0.0004*** (0.0000)	0.0000 (0.0000)
FEMALE	0.0894 (0.0874)	-0.3333*** (0.0188)
WHITE	-0.0511 (0.0870)	-0.0131 (0.0188)
FEMALE*AGE	-0.0151*** (0.0046)	0.0004 (0.0005)
FEMALE*AGESQ	0.0002*** (0.0001)	0.0000 (0.0001)
WHITE * AGE	-0.0002*** (0.0000)	0.0020*** (0.0005)
WHITE * AGESQ	0.0000 (0.0000)	0.0000 (0.0001)
FEMALE * WHITE	-0.0069 (0.0079)	0.0097 (0.0080)
Constant	1.7556*** (0.0687)	1.8508*** (0.0545)
ρ		-0.2378*** (0.0097)
N	109,687	109,687
Adjusted R ²	0.4160	0.3352

Source: Authors' calculations based on PNAD (IBGE). Standard errors in parentheses.

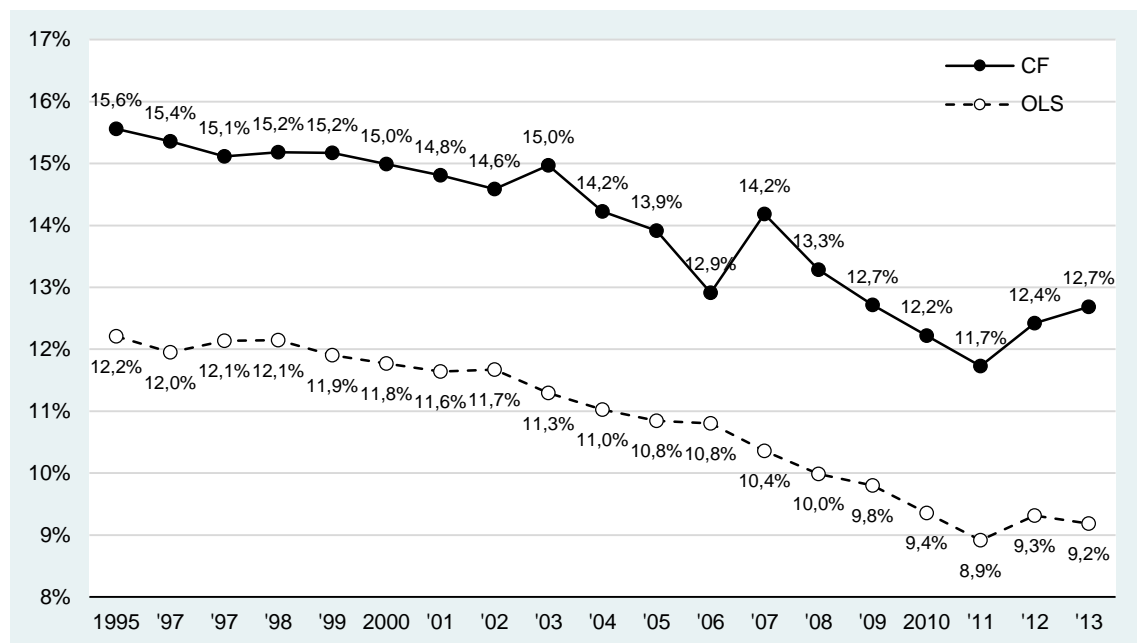
Note: OLS estimates for equation (33) and CF estimates for equation (32). State of birth dummies are also included in the regression.

* significant at 10%; ** significant at 5%; *** significant at 1%.

Menezes-Filho, Fernandes and Picchetti (2006) and Tavares and Menezes-Filho (2011) have already documented the decline in returns to schooling in the past decade and its importance to the decline in earnings inequality, even though they are unable to properly account for endogeneity. Fernandes and Menezes-Filho (2012) present evidence that this phenomenon might be related to the sharp increase in the relative supply of medium and high-skilled workers. It seems a reasonable explanation, but arguably further research is needed, especially to identify the reasons for the recent increase in the returns.

The results presented thus far are not directly comparable to other studies that attempt to measure the causal impact of education on wages in Brazil. Menezes-Filho, Fernandes and Pichetti (2006) report returns of 14% in 1997 and Tavares and Menezes-Filho (2011) report declining returns between 1995 and 2009, when they reached slightly less than 12%. Both studies rely on repeated cross-sections to estimate average returns to schooling. Teixeira and Menezes-Filho (2012), on the other hand, use an IV approach to estimate a much lower return of 5.5% per year of education, leading to a conclusion of an upward bias in OLS estimates. The instruments they use are the number of schools in the state and year when the individual was born and an educational law passed in 1971, which unified primary education in Brazil. While their results seem to contradict the ones presented in this paper, they can be reconciled taking into account their LATE interpretation. As argued by Imbens and Angrist (1994), the coefficients estimated by means of IV represent the causal effect only for the subsample of compliers (the individuals who were *actually* affected by variations in the instrument). It seems sound to assume that for Teixeira and Menezes-Filho (2012) the compliers are individuals with lower levels of education, so their estimates are not directly compared to the ones presented in this section. In order to make results comparable, we would need to be able to identify the compliers and then estimate the returns to education only for that subsample.

Figure 11 – Returns to education in Brazil: 1995-2013



Source: Authors' calculations based on PNAD (IBGE).

Note: Each line depicts the evolution of δ in equations (33) for OLS and (29) for the control function estimates. Estimations are done separately for each year. All coefficients are significant at the 1% level in all years. In the last years of each decade, IBGE does not carry PNAD: values for 2000 and 2010 are simple linear interpolations.

The results presented in this paper also seem reasonable when compared internationally. Psacharopoulo and Patrinos (1994) argue that, overall, the international average of the Mincerian return to schooling is 10% and that it is higher in middle and low income countries due to diminishing marginal returns. Interestingly, three recent studies apply

similar approaches to ours. Klein and Vella (2009) find a return of 10% for Australian workers and Saniter (2012) estimated an average return of 8.5% in Germany (both studies estimate H_j semiparametrically). Farré, Klein and Vella (2013) employ the same parametric methodology we used in this paper and find an average return of 11.2% for the US using the NLSY79 database. All three studies find that the OLS estimates are biased downwards, as our results also suggest. As extensively documented in the literature, it seems reasonable that in Brazil, a middle-income country, returns to schooling are higher than those in Australia, Germany and the US. Less obvious, but also of interest, is the fact that the OLS estimates are biased downwards in these studies and in our results as well.

5. Discussion: the OLS-IV gap

In this section we shed some light on the direction of the OLS bias that was found and reported earlier. The negative estimate for ρ implies that the OLS estimate will be smaller than the one obtained in the controlled function setting. One important test is to measure the actual correlation between \hat{u} and \hat{v} , since that is, in principle, what ρ yields. We found that, indeed, the observed correlation is -0.2126 in 2013, within the 95% confidence interval of the estimated correlation of -0.2378 (Table 10). This is not obtained by construction; on the contrary, we tested different specifications for H_j (such as the exponential form as FKV, 2013) and found that, in general, the average return schooling is consistent throughout the alternative forms, but $\hat{\rho}$ estimates vary significantly. Figure 12 displays the plot between both residuals and clearly highlight the slightly negative relationship between them.

The OLS-IV gap has been extensively discussed in the literature. Some early papers suggest that the omitted variable bias arises because of unaccounted ability, which in turn would produce an upward bias in OLS estimates. Other authors have claimed that measurement errors in the education variable would produce downward biased OLS estimates (see Angrist and Krueger 1991; Card 1995, 1999; Cameron and Taber 2004).

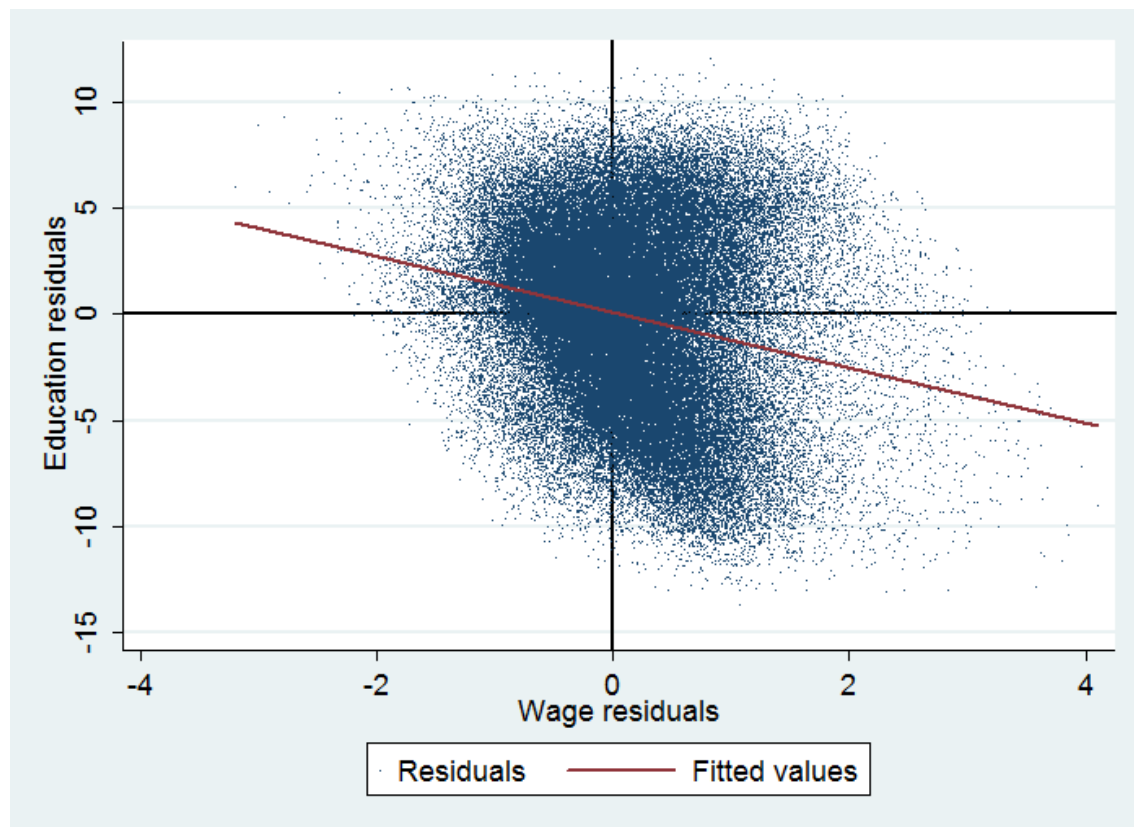
A third trend in the literature, however, claims that aside from ability, the error component of the education equation captures other factors, such as motivation, that would lead individuals to obtain what Vella and Gregory (1996) call ‘over-education’. This over achievement would, in turn, yield lower returns to schooling because returns to over education are lower than the returns to education (Dolton and Vignoles 2000; Groot and van den Brink 2000; Rubb 2002; Farré, Klein and Vella 2013), yielding what has been called the ‘over education penalty’. Looking once again at Figure 12, we see that there is

Mathematically, this penalty is only an interpretation for the negative value of ρ . If the correlation between u and v is negative, then if one has a higher education than what is predicted by the education model, she will likely have a lower than expected wage. In a developing country context, however, it might be useful to look at the other half of the coin.

If the claim that there is an over-education penalty is valid, then for the same reasons one might expect an ‘under-education premium’. This means that if the education of one individual is lower than expected, she probably earns more than what is predicted by the wage model. This could happen in a context where factor such as credit constrains or school

availability would lead individuals to get less education. In this context, skilled individuals would be able to somehow surpass this education constraint and earn more than what is predicted given their level of schooling.

Figure 12 – Residuals analysis: education and wage equations (2013)



Source: Authors' calculations based on PNAD (IBGE). Education residuals are \hat{v} wage residuals are given by \hat{u} , which is calculated as the difference between observed and predicted log wages (predicted with the unbiased estimate of δ).

To see this more clearly, we cross-tabulated the residuals dividing the sample between those with positive or negative residuals. Results depicted in

Table 11 show that among the less educated the 'under education premium' prevails, whereas among those with more schooling the over education penalty is more frequent. Comparing both groups, the under-education premium is slightly more frequent than the under education penalty.

Table 11 – Cross tabulation of positive and negative residuals (2013)

Entire sample		
	$v \geq 0$	$v < 0$
$u \geq 0$	24,1%	28,5%
$u < 0$	30,5%	16,9%
9 years or less of schooling		
	$v \geq 0$	$v < 0$
$u \geq 0$	4,1%	61,1%
$u < 0$	5,3%	29,5%
More than 9 years of schooling		
	$v \geq 0$	$v < 0$
$u \geq 0$	39,1%	4,1%
$u < 0$	49,5%	7,4%

Source: Authors' calculations based on PNAD (IBGE). Note: A negative (positive) residual means that the expected education or wage is higher (lower) than what is observed. Over- (under-) education means that $v > 0$ ($v < 0$), and a wage premium (penalty) means that $u > 0$ ($u < 0$).

6. Robustness check

The first robustness check has been presented in the previous section, when it was verified that the observed correlation between residuals is not statistically different from $\hat{\rho}$. In this section we replicate the estimation of the returns to schooling using different specifications for the functional forms of the heteroskedasticity functions to examine the robustness of the findings presented in previous sections. Two different tests are performed and results are displayed in Table 12. The estimate presented in the second columns uses a less rigid form for H_v and H_u : age enters the regression as 5-year interval dummies and we don't use any state (of birth or residence) variables (we do use age and gender and age and race interactions). The model depicted in the third column follows FKV (2013) with an exponential specification for H_j with age, age squared, gender, race and state variables (no interactions).

These tests are relevant because the identification strategy employed earlier in the estimation requires that the variables used in each specification of the equations generate enough and consistent heteroskedasticity. Thus, changing the functional form and the set of variables used in each heteroskedasticity function are key tests for the robustness of our findings.

The estimates for these alternate specifications are 12.3% and 12.6%, slightly below the 12.7% estimate obtained under the benchmark specification. Note, however, that the difference between all these estimates are within the 95% confidence interval of each other. Thus, we can state that the results presented earlier are unaffected by the use of different functional forms for heteroskedasticity. It should also be noted, however, that the estimates for $\hat{\rho}$ change substantially when the exponential form is used. Not only it changes, it is also biased, since the observed correlation between residuals under this specification is -0.2072.

Table 12 - Control function estimates with alternative specifications: Wage equation (2013)

	β_{CF} (bechmark especification)	β_{CF} (dummies specification)	β_{CF} (exponential heteroskedasticity)
YRSEDUC	0.1239*** (0.0015)	0.1226*** (0.0015)	0.1259*** (0.0037)
ρ	-0.2590*** (0.0104)	-0.1984*** (0.0112)	-0.0892** (0.0363)
N	109,687	109,687	109,687
Adjusted R ²	0.3352	0.3070	0.3443

Source: Authors' calculations based on PNAD (IBGE). Standard errors in parentheses.

Note: Column 1 reports estimates for equation (32). Results reported in column 2 uses a more flexible specification for H_j with only dummy variables, while in column 3 H_j is modeled as an exponential function of age, age squared, gender and race. Standard errors in parentheses.

* significant at 10%; ** significant at 5%; *** significant at 1%.

7. Conclusion

This paper estimates the causal returns to education for the Brazilian population during the period 1995-2013. The naïve OLS regression of earnings on years of schooling yields estimates with the well-known endogeneity bias. Klein and Vella (2010) developed a control function setting in which heteroskedasticity provides identification without the need for exclusion restrictions. The key advantage of this approach is the fact that estimation can be done for the entire population at any point in time, allowing for a more general interpretation than the IV's LATE.

One possible drawback of the KV (2010) method is the computational demands that arise due to their semiparametric estimators. Farré, Klein and Vella (2013) propose a fully parametric approach that allows for the implementation of the KV (2010) estimator in practice. We apply this parametric approach and find that the average return to education have declined in Brazil from 15.6% in 1995 to 11.7% in 2011 and since then has bounced back to 12.7% in 2013. These estimates are higher than the OLS estimated coefficients suggest, pointing to a downward bias in the OLS estimation. We interpret this bias as a sign of over-education penalties and under-education premiums. Also, the decline in returns to schooling seem to be associated with the well documented increase in the supply of more educated workers observed in the past two decades in Brazil.

References

- Angrist J. and Krueger, A. B. (1991) "Does compulsory school attendance affect schooling and earnings". *Quarterly Journal of Economics* 106:979–1014
- Cameron, S. and Taber, C. (2004) "Estimation of educational borrowing constraints using returns to schooling". *Journal of Political Economy* 112:132–182.
- Card, D. (1995) "Earnings, schooling and ability revisited". In: Polachek, S. W. (ed) *Research in labor economics*. JAI Press, Greenwich, Connecticut, pp 23–48
- Card, D. (1999) "The causal effect of education on earnings". In: Ashenfelter O, Card D (eds) *Handbook of labor economics*, vol 3A, Chap. 30. Elsevier Science/North-Holland, Amsterdam.
- Carneiro P, Lee S (2008) "Changes in college enrolment and wage inequality: distinguishing price and composition effects", UCL Working Paper.
- Chen, S. H. (2008) "Estimating the variance of wages in the presence of selection and unobserved heterogeneity". *Review of Economics and Statistics*, 90(2):275–289.
- Dolton P, Vignoles A (2000) "The incidence and effects of overeducation in the UK graduate labour market". *Econ Educ Rev* 19(2):179–198
- Duflo E (2001) "Schooling and labor market consequences of school construction in Indonesia: evidence from an unusual policy experiment". *American Economic Review*, 91(4):795–813.
- Figueirêdo, Erik, Nogueiray, Lauro and Santanaz, Fernanda Leite. (2014) "Igualdade de Oportunidades: Analisando o Papel das Circunstâncias no Desempenho do ENEM". *Rev. Bras. Econ.*, Rio de Janeiro, v. 68, n. 3, p. 373-392.
- Griliches, Zvi (1977). "Estimating the Returns to Schooling: Some Econometric Problems". In: *Econometrica* 45.1, pp. 1-22.
- Groot W, van den Brink HM (2000) "Overeducation in the labor market: a meta-analysis". *Econ Educ Rev* 19(2):149–158
- Heckman J, Urzua S, Vytlacil E (2006) "Understanding instrumental variables in models with essential heterogeneity". *Rev Econ Stat* 88(3):389–432.
- Ichino, Andrea and Rudolf Winter-Ebmer (2004). "The Long-Run Educational Cost of World War II". In: *Journal of Labor Economics* 22.1.
- Imbens, Guido W. and Joshua D. Angrist (1994). "Identification and Estimation of Local Average Treatment Effects". In: *Econometrica*, pp. 467-475.
- Klein, Roger and Vella, Francis (2010). "Estimating a class of triangular simultaneous equations models without exclusion restrictions," *Journal of Econometrics*, Elsevier, vol. 154(2), pages 154-164, February.
- Menezes-Filho, Naercio Aquino, Fernandes, Reynaldo, and Picchetti, Paulo. (2006). "Rising human capital but constant inequality: the education composition effect in Brazil". *Revista Brasileira de Economia*, 60(4), 407-424.

- Mincer, Jacob (1974). "Schooling, Experience, and Earnings". New York: Columbia University Press.
- Nocedal, J. (1996). Conjugate gradient methods and nonlinear optimization. Linear and Nonlinear Conjugate Gradient-Related Methods, 9-23.
- Oreopoulos, Philip (2007). "Do Dropouts Drop Out Too Soon? Wealth, Health and Happiness from Compulsory Schooling". In: Journal of Public Economics 91.11-12, pp. 2213-2229.
- Psacharopoulos, George, and Harry Anthony Patrinos*. "Returns to investment in education: a further update." Education economics 12.2 (2004): 111-134.
- Saniter, Nils, 2012. "Estimating Heterogeneous Returns to Education in Germany via Conditional Heteroskedasticity," IZA Discussion Papers 6813, Institute for the Study of Labor (IZA).
- Rubb S (2002) "Overeducation in the labor market: a comment and re-analysis of a meta-analysis". Econ Educ Rev 22(6):621-629
- Schroeder, Elizabeth (2010). "The Impact of Microcredit Borrowing on Household Consumption in Bangladesh". Department of Economics, Georgetown University.
- Tavares, P. A. and Menezes-Filho, N. A. (2011). Human Capital and the Recent Decline of Earnings Inequality in Brazil. Brazilian Review of Econometrics, 31(2), 27.
- Teixeira, Wladimir Machado and Menezes-Filho, Naécio Aquino. (2010). "Estimando o retorno à educação do Brasil considerando a legislação educacional brasileira como um instrumento". Revista de Economia Política (Impresso), v. 32, p. 479-496, 2012

Appendix

Table 13 – Heteroskedasticity test: Education equation (2013)

Breusch-Pagan/Cook-Weisberg test for heteroskedasticity		
H₀: Constant variance		
Variable	Chi²	p
AGE	2179.47	0.0000
AGESQ	2062.45	0.0000
FEMALE	103.42	0.0000
WHITE	20.96	0.0000
FEMALE*AGE	0.31	0.5786
FEMALE*AGESQ	83.09	0.0000
WHITE*AGE	26.55	0.0000
WHITE*AGESQ	155.54	0.0000
WHITE*FEMALE	103.88	0.0000
Simultaneous	2509.72	0.0000

Source: Authors' calculations based on PNAD 2013 (IBGE). State of birth dummies were also included in the regression.

Table 14 – Heteroskedasticity test: Wage equation (2013)

Breusch-Pagan/Cook-Weisberg test for heteroskedasticity		
H₀: Constant variance		
Variable	Chi²	p
YRSEDUC	739.12	0.0000
AGE	699.67	0.0000
AGESQ	670.53	0.0000
FEMALE	0.30	0.5861
WHITE	239.84	0.0000
FEMALE*AGE	12.62	0.5784
FEMALE*AGESQ	43.31	0.0000
WHITE*AGE	55.52	0.0000
WHITE*AGESQ	429.04	0.0000
WHITE*FEMALE	552.64	0.0000
Simultaneous	1929.10	0.0000

Source: Authors' calculations based on PNAD 2013 (IBGE). State dummies were also included in the regression.

Table 15 –Heteroskedasticity functions: education and wage equations (2013)

	θ_v	θ_u
AGE	1.4706*** (0.1105)	-0,0002*** (0,0000)
AGESQ	-1.4725*** (0.1411)	0,0000 (0,0000)
FEMALE	2.1922 (2.7327)	-0,0856 (0,3107)
WHITE	5.6344** (2.7259)	-0,4219*** (0,0235)
FEMALE*AGE	-.3104** (0.1448)	0,0515 (1,9715)
FEMALE*AGESQ	-0.5696*** (0.2311)	0,0547 (2,3876)
WHITE*AGE	-0.3063** (0.1443)	-0,0969*** (0,0088)
WHITE*AGESQ	-0.7979** (0.1315)	-0,0120** (0,0880)
WHITE*FEMALE	-0.9715*** (0.2573)	-0,0605 (0,0396)
CONSTANT	-11.4832* (1.3299)	0,7054*** (0,0737)
N	109,687	109,687
Adjusted R ²	0.0366	0.3352

Source: Authors' calculations based on PNAD 2013 (IBGE). Standard errors in parentheses.

Note: Linear Least Squares estimates for θ_v (equation (31)) are depicted in column (i) and Nonlinear Least Squares for θ_u (equation (32)) are depicted in column (II). State of birth dummies were included in the regression.

* significant at 10%; ** significant at 5%; *** significant at 1%.

Capítulo 3: Diferencial de salários da mão de obra terceirizada no Brasil¹⁰

Resumo

Esse artigo analisa empiricamente o diferencial de salários da mão de obra terceirizada no Brasil. Uma simples comparação entre as remunerações médias indica que os salários dos terceirizados são 17% inferiores, mas quando o diferencial é controlado pelo efeito fixo do trabalhador, a diferença cai para -2%. Além disso, as evidências apontam para uma grande heterogeneidade no diferencial salarial. Trabalhadores de ocupações de baixa qualificação têm uma remuneração até 16% inferior quando estão terceirizados. Por outro lado, as ocupações de alta qualificação oferecem salários iguais ou maiores, em média, para os terceirizados. Além disso, as evidências indicam que o diferencial desfavorável ao terceirizado se reduziu entre 2007 e 2012. Uma possível explicação para esse resultado pode estar na especialização e na necessidade de acumulação de capital humano específico. O trabalho também discute alternativas de políticas para a dicotomia entre atividade-fim e atividade-meio.

Palavras-chave: mercado de trabalho, terceirização, diferencial de salários

Códigos JEL: J310, J010, J400

¹⁰ Esse capítulo foi escrito com a coautoria de Guilherme Stein.

1. Introdução

Por que e sob quais circunstâncias o processo produtivo é organizado por firmas? Coase (1937) respondeu a essa pergunta chamando a atenção para os custos de transação, mas desde então setenta e oito anos se passaram. Nesses três quartos de século houve mudanças tecnológicas e sociais profundas (internet, transporte, logística etc.) que estão reorganizando o trabalho no mundo (Eichhorst, 2015 e Zimmerman, 2015). A teoria de Coase continua válida, mas certamente mudaram os custos de transação que fazem ou faziam com que a coordenação da produção fosse mais eficiente sob a coordenação de um empreendedor do que sob a coordenação do sistema de preços. Nesse sentido, observamos novas formas de organização do trabalho e da produção e em muitos casos a verticalização das firmas dá lugar a redes horizontais de produção

Essas mudanças nas cadeias produtivas levam a um fenômeno conhecido internacionalmente como *outsourcing*. No Brasil, esse termo ficou conhecido como terceirização, que nada mais é do que a transferência de parte das atividades de uma firma para outra. O *outsourcing* só ocorrerá se a houver alguma vantagem de custo para a firma que terceiriza sua mão de obra. Essa vantagem pode ser obtida caso (i) a mão de obra terceirizada seja mais eficiente (eficiência obtida por especialização, gestão etc.) ou (ii) a remuneração dos trabalhadores terceirizados seja menor do que a que seria recebida caso o mesmo trabalhador fosse contratado diretamente. Essa segunda possibilidade é, inclusive, frequentemente utilizada como argumento por entidades sindicais e alguns setores da sociedade brasileira (CUT, 2014; CNBB, 2015) para justificar posições contrárias à regulamentação da terceirização.

Teoricamente, só haveria motivo para encontrarmos diferenças salariais caso o mercado de trabalho terceirizado fosse de alguma forma segmentado. Do contrário, se houvesse mobilidade perfeita dos trabalhadores, os salários deveriam ser equivalentes por conta da possibilidade de arbitragem. Ou seja, é difícil explicar porque um terceirizado aceitaria ganhar menos nessa condição, a menos que haja diferença em aspectos institucionais relevantes para a formação dos salários, como por exemplo a representação sindical ou a insegurança jurídica que faria com que a contratante impusesse um ‘prêmio’ para contratar terceirizados.

De fato, a legislação trabalhista no Brasil não tem regulamentada de forma clara a possibilidade de se terceirizar mão de obra. Para suprir essa lacuna, o Tribunal Superior do Trabalho estabeleceu jurisprudência sobre o tema através da súmula 331, que define que as empresas podem terceirizar “serviços de vigilância e de conservação e limpeza, ou outros serviços especializados ligados à atividade-meio do tomador, desde que inexistente a pessoalidade e a subordinação direta”. Essa preocupação em proibir a terceirização mais ampla parece estar ligada à ideia de que de alguma forma a terceirização implica na contratação dos trabalhadores em condições piores do que eles teriam caso fossem contratados diretamente.

Apesar disso, não existem no Brasil estudos empíricos que avaliam o diferencial de salários e de condições de trabalho dos trabalhadores terceirizados e próprios¹¹. Idealmente, para

¹¹ Neste trabalho nos referimos a trabalhadores contratados diretamente como ‘trabalhadores próprios’, sendo que o termo próprio não é empregado no sentido de propriedade, mas sim, de

responder se há diferenças salariais entre as duas formas de contratação, seria necessário observar num experimento o mesmo trabalhador sendo contratado pela mesma empresa na condição de próprio e na condição de terceiro, ao mesmo tempo. Como isso não é possível no mundo real, estudos empíricos adotam estratégias que tentam tornar comparáveis os salários observados por diferentes tipos de trabalhadores. Uma primeira abordagem seria a comparação simples entre os salários médios dos terceirizados e dos próprios, como proposto em CUT (2014). Entretanto, essa abordagem desconsidera as diferenças nas características tanto dos trabalhadores quanto das firmas que os contratam. Por exemplo, é possível que no início de carreira seja mais comum que um vigia seja contratado por uma empresa que presta serviços de vigilância, e que depois que certa experiência seja adquirida, esse vigia seja contratado diretamente. Também é possível que as empresas que terceirizam sejam intrinsicamente diferentes das empresas que contratam diretamente, necessitando de trabalhadores mais ou menos qualificados. Por exemplo, um prédio que contrata diretamente uma faxineira pode demandar características (observáveis ou não observáveis) diferentes das demandadas por prédios que terceirizam esse trabalho.

Para tentar expurgar essas diferenças, uma segunda abordagem seria a de comparar os salários médios dos terceirizados e dos não terceirizados controlando pelas características observáveis das firmas e dos trabalhadores. Fatores como escolaridade, idade, ramo de atividade e tamanho das empresas explicam, como veremos, parte da diferença de remuneração que observamos entre os dois tipos de contratação. Porém, mesmo controlando por essas características, ainda notamos que no Brasil os trabalhadores terceirizados ganharam, em média, -XX% entre 2007 e 2012.

O controle das características observáveis pode ser um exercício incompleto se, além delas, aspectos não observáveis também forem relevantes para a determinação do salário. Habilidades não cognitivas são relevantes para a determinação da remuneração e incluem motivação, dedicação, capacidade de comunicação e maturidade emocional são alguns exemplos de características que podem ser valorizadas pelos empregadores, mas não são facilmente observadas pelos pesquisadores. Nesse sentido, é possível que algumas dessas características sejam mais comumente encontradas nos trabalhadores próprios. Dessa forma, a comparação dos salários médios controlada pelas características observáveis também não seria suficiente para a correta medida do diferencial de salários.

Diferenciais salariais já foram amplamente discutidos na literatura nacional (ver, por exemplo, Botelho e Ponzek, 2011 e Menezes-Filho et al, 2004), mas mesmo com nossos melhores esforços não encontramos estudos empíricos que discutam a questão do impacto da terceirização nos salários no Brasil. Nos EUA, Dube e Kaplan (2010) é um raro exemplo de estudo empírico nos moldes deste artigo, com uma modelagem de efeito fixo. Os autores encontram um diferencial de -4% para serviços de limpeza e de -9% para vigias. Novamente, essas evidências apontam para impactos heterogêneos da terceirização.

Este estudo busca chegar o mais próximo possível do exercício hipotético de observar o mesmo trabalhador sob duas formas de contratação, mas no mesmo momento e na mesma firma. Utilizando dados em painel, observamos o mesmo trabalhador ao longo do tempo e conseguimos comparar as mudanças em seu próprio salário, de acordo com a forma de

acordo com as definições usuais encontradas nos dicionários, no sentido de que o contrato de trabalho é inerente e característico da firma tomadora dos serviços.

contratação. Não observamos dois salários ao mesmo tempo e nem na mesma firma, mas conseguimos controlar as comparações através dos chamados efeitos fixos dos trabalhadores. Com essa metodologia, mostramos para um conjunto de seis ocupações tipicamente terceirizáveis no Brasil, o diferencial de salários dos trabalhadores terceirizados é de -3%. Mais ainda, mostramos que esse diferencial é bastante heterogêneo: enquanto serviços como Limpeza e Telemarketing têm um diferencial bastante negativo para os terceirizados, a terceirização de serviços de Tecnologia da Informação (TI) ou Segurança/Vigilância implica em uma remuneração média maior.

Essas evidências são importantes para embasar o debate acerca da regulamentação da terceirização no Brasil. Em 2015, a Câmara dos Deputados aprovou o Projeto de Lei 4.330, que autoriza a terceirização para todas as atividades da empresa e estabelece as regras para que esse procedimento possa ser adotado. Diante do debate que se estabelece na sociedade brasileira, buscamos contribuir mostrando evidências sobre os efeitos da terceirização nos salários dos trabalhadores.

O restante do trabalho está organizado da seguinte forma. A seção 2 descreve a base de dados utilizada. A estratégia empírica é descrita e explicada na seção 3, enquanto a seção 4 apresenta os resultados. A seção 5 exibe alguns testes de robustez e a seção 6 conclui.

2. Dados

Este estudo utiliza a base de dados da Relação Anual de Informações Sociais (RAIS) para os anos de 2007 a 2012. A RAIS é compilada pelo Ministério do trabalho e Emprego (MTE) anualmente desde 1975 e contém informações detalhadas sobre todos os vínculos de emprego registrados no Brasil, já que todas as pessoas jurídicas e outros tipos de empregadores devem enviar as informações referentes ao contrato de trabalho (salário, data de início do vínculo, afastamentos, desligamentos etc.) e também informar características sociodemográficas de cada trabalhador (idade, escolaridade, raça/cor, gênero etc.).

O MTE disponibiliza publicamente os microdados da RAIS, porém sem a identificação dos indivíduos. A base de dados que utilizamos (também fornecida pelo MTE), porém, é ligeiramente diferente porque contém variáveis que permitem a identificação do mesmo indivíduo ao longo do tempo. A seção seguinte ilustra a importância dessa possibilidade de acompanhamento dos trabalhadores ao longo dos anos. Para construir a base de dados para este estudo, mantivemos apenas os trabalhadores que tinham um único vínculo ativo em 31 de dezembro de cada ano. Além disso, utilizamos apenas as informações dos vínculos ativos no final do ano, ou seja, caso tenha havido mais de uma mudança de emprego ao longo do ano, essa informação não foi utilizada. Também excluímos os registros nos quais o salário por hora registrado era inferior ao salário mínimo por hora vigente.

A identificação dos trabalhadores terceirizados não é possível diretamente. Esse talvez seja, inclusive, um dos principais motivos que justificam a ausência de trabalhos empíricos sobre o tema. De forma semelhante a Dube e Kaplan (2010), adotamos um procedimento que permite identificar indiretamente quem são os trabalhadores terceirizados em determinadas ocupações.

O método é simples e pode ser dividido em três etapas. Primeiro, definimos¹² um conjunto de ocupações que são tipicamente terceirizáveis, tais como: *Porteiros, vigias e afins; Operadores de telemarketing; Trabalhadores nos serviços de manutenção de edificações* (que inclui, entre outros, faxineiros) etc. Segundo, definimos atividades econômicas que são tipicamente de empresas prestadoras de serviços de terceirização de mão de obra, como por exemplo, *Atividades de vigilância, segurança privada e transporte de valores; Atividades de limpeza* etc. Finalmente, cruzamos os dois conjuntos de classificações anteriores e identificamos os trabalhadores em ocupações tipicamente terceirizáveis que trabalhavam em empresas cuja atividade é a prestação de serviços de terceirização. Esses trabalhadores foram então classificados como terceirizados em nossa base de dados. Em anexo ao final do artigo listamos todas as ocupações e setores considerados. Além disso, também consideramos o código de atividade 829 (*Outras atividades de serviços prestados principalmente às empresas*) como atividade típica de empresas prestadoras de serviços terceirizados (desde que a ocupação do trabalhador coincida com as listadas anteriormente).

Um exemplo pode ajudar na compreensão do método. Suponha que identificamos na base da RAIS em um dado ano um indivíduo que trabalha na ocupação de vigia. Se esse trabalhador for contratado por uma empresa cuja atividade econômica é, por exemplo, *Fabricação de calçados* ou *Comércio varejista*, assumimos esse trabalhador não é terceirizado. Por outro lado, se o ramo de atividade da empresa em que esse vigia trabalha for *Atividades de vigilância, segurança privada e transporte de valores*, então assumimos que esse indivíduo é terceirizado.

Há dois tipos de erro que podem ser cometidos com o procedimento descrito acima. Primeiro, podemos classificar um trabalhador que é terceirizado por uma empresa cuja atividade principal não é a de prestação de serviços terceirizados. Segundo, podemos classificar um trabalhador próprio de uma empresa de terceirização de mão de obra como terceirizado (como um vigia que cuida da segurança da sede da empresa). Acreditamos que os dois casos devem representar uma parcela muito pouco significativa da base de dados. Tanto o primeiro quanto o segundo tipo de erro tendem a subestimar o efeito da terceirização.

A Tabela 1 lista os seis tipos de trabalhos terceirizados que consideramos para este estudo e mostra a quantidade de trabalhadores em cada um deles. Para identificar os terceirizados, utilizamos a Classificação Brasileira de Ocupações (CBO) no nível da *família ocupacional* (4 dígitos) e a Classificação Nacional de Atividades Econômicas (CNAE) no nível de grupos de atividade (3 dígitos). Reconhecemos que a classificação que este estudo adota não inclui a totalidade de ocupações terceirizáveis (ocupações como recepcionistas e serviços como jardinagem, por exemplo, não foram considerados). Cabe lembrar que o Brasil carece de uma definição precisa sobre quais ocupações são terceirizáveis, o que aumenta a complexidade do problema de classificação do que é ou não é 'terceirizável'. Nesse sentido, a generalização dos resultados que serão apresentados nas próximas seções deve ser feita com cautela. Além disso, também foram excluídos das bases os indivíduos que transitaram entre as ocupações analisadas (por exemplo, um trabalhador que apareceu em um dado momento em atividades de limpeza e em outro ano em atividades de vigilância). Ao final do processo de montagem da base de dados havia 8.245.683 indivíduos diferentes ao longo dos seis anos disponíveis. Note-se que o painel não é balanceado, já que apenas 4,7% dos

¹² Os tipos de trabalho terceirizados foram adaptados de CNI (2014). A compatibilização CBO-CNAE foi feita pelos autores deste estudo.

indivíduos estavam empregados formalmente nas ocupações consideradas ao final de cada um dos anos entre 2007 e 2012.

De maneira geral, chama a atenção o fato de que a participação da mão de obra terceirizada no conjunto das ocupações analisadas diminui 0,2 ponto percentual entre 2007 e 2012, atingindo 24,5% do total da mão de obra nesse conjunto de ocupações. Esse percentual permite supor que a estimativa de que 26,8% de terceirizados do total do emprego formal em 2013 CUT (2014) é superestimado, uma vez que mesmo entre as ocupações tipicamente terceirizáveis o percentual é menor.

Além disso, é interessante notar que a disseminação da terceirização é heterogênea: enquanto ocupações relacionadas a *Montagem e manutenção de equipamentos* tinham 5,4% de terceirizados em 2012 (6,8% em 2007), *Telemarketing* tinha 59,5% dos trabalhadores sendo contratados como terceirizados no mesmo ano (e 65,2% em 2007). Em 2012, a base de dados tinha 3,7 milhões de indivíduos, em torno de 8% de todo o emprego formal do Brasil naquele ano segundo o MTE.

Tabela 1 – Serviços consideradas para a análise da terceirização e suas participações no emprego formal do Brasil (2007/2012)

Serviços	2007			2012		
	Terceiros	Próprios	% Ter.	Terceiros	Próprios	% Ter.
MONTAGEM E MANUTENÇÃO DE EQUIPAMENTOS	25.901	352.567	6,8%	36.455	638.727	5,4%
SEGURANÇA/VIGILÂNCIA	188.949	456.033	29,3%	283.251	695.519	28,9%
TECNOLOGIA DA INFORMAÇÃO	30.916	93.244	24,9%	83.919	146.270	36,5%
LIMPEZA E CONSERVAÇÃO	188.899	685.018	21,6%	359.936	1.211.968	22,9%
PESQUISA E DESENVOLVIMENTO	1.218	13.371	8,3%	2.492	30.595	7,5%
TELEMARKETING	95.967	51.123	65,2%	148.896	101.236	59,5%
TOTAL	531.850	1.651.356	24,4%	914.949	2.824.315	24,5%

Fonte: RAIS (MTE). Elaboração própria.

A Tabela 2 exibe algumas estatísticas descritivas da base de dados. Inicialmente, notamos que os trabalhadores terceirizados ganham, em média, menos do que os trabalhadores próprios. A comparação das demais estatísticas mostra, porém, que a comparação dos salários médios de terceirizados e não terceirizados deve ser feita com cautela, uma vez que há diferenças nos valores médios de variáveis importantes para a determinação da remuneração dos trabalhadores. Por exemplo, os terceirizados são mais de três anos mais novos, em média, e uma menor proporção deles tem ensino superior completo. Além disso, a duração média dos vínculos é consideravelmente maior no caso dos trabalhadores diretamente contratados. Com essas diferenças registradas, é natural que a remuneração média observada dos terceirizados seja menor. A questão relevante é se essas diferenças das características observáveis são suficientes para explicar toda a diferença nos salários nessas duas formas de contratação.

Tabela 2 – Estatísticas descritivas

Variável	Terceirizados				Próprios			
	Média	DP	Mín.	Máx.	Média	DP	Mín.	Máx.
REMUNERAÇÃO MÉDIA *	993,06	1382,85	21,69	85523,76	1280,12	1876,18	2,92	91447,76
IDADE	34,74	10,39	10,47	100,32	37,94	11,73	10,54	110,00
HORAS CONTRATADAS	41,52	5,07	1	44	41,51	4,990856	1	44
MULHERES	0,4156	-	0	1	0,3437	-	0	1
BRANCO	0,5125	-	0	1	0,5851	-	0	1
ENSINO MÉDIO	0,5123	-	0	1	0,4739	-	0	1
ENSINO SUPERIOR	0,0598	-	0	1	0,0627	-	0	1
TEMPO NO EMPREGO (MESES)	31,64	41,03	0	599,3	65,92	82,73	0	599,9
NÚMERO DE DIAS AFASTADO	3,88	24,87	0	392	3,65	24,40	0	730
ANO = 2007	0,1201	-	0	1	0,1190	-	0	1
ANO = 2008	0,1862	-	0	1	0,1775	-	0	1
ANO = 2009	0,1780	-	0	1	0,1774	-	0	1
ANO = 2010	0,1929	-	0	1	0,1904	-	0	1
ANO = 2011	0,1160	-	0	1	0,1322	-	0	1
ANO = 2012	0,2067	-	0	1	0,2035	-	0	1
TOTAL DE OBSERVAÇÕES	4.426.553				13.877.680			

Fonte: RAIS (MTE).Elaboração própria.

(*) Salário mensal em 31/dezembro de cada ano corrigido pelo IPCA para dezembro de 2012.

3. Metodologia e modelo econométrico

O objetivo deste trabalho é avaliar o diferencial de salários entre trabalhadores terceirizados e próprios. A estratégia empírica utiliza modelos com efeito fixo dos indivíduos. Abordagem similar pode ser encontrada, por exemplo, em Botelho e Ponczek (2011), que estimam modelos com efeito fixo para avaliar o diferencial de salários entre trabalhadores formais e informais.

Neste trabalho, nossa equação de referência é a seguinte:

$$w_i = \alpha_0 + \delta o_{it} + \beta x_{it} + \mu_i + e_{it} \quad (33)$$

onde w_i é o logaritmo natural do salário (real) do indivíduo i , o_{it} é uma variável binária que assume valor um caso o trabalhador i seja terceirizado no instante t , x_{it} é um vetor de covariadas observáveis relacionadas ao indivíduo i no instante t ; em nosso caso, x_{it} inclui *dummies* de ano (para controlar por eventuais tendências nos salários comuns a todos os indivíduos), *dummies* relacionadas ao tamanho do estabelecimento, *dummies* do setor de atividade, o número de horas contratadas, *dummies* para o grau de instrução (para controlar

mudanças na escolaridade dos trabalhadores ao longo do tempo¹³), a duração do vínculo de trabalho e finalmente seis *dummies* que controlam para os seis grupos ocupacionais de interesse. Por fim, μ_i é o efeito fixo do indivíduo i (potencialmente correlacionado com O_{it}) que captura características não observáveis e que não variam ao longo do tempo (habilidade, preferências etc.), enquanto e_{it} é erro.

A inclusão do efeito fixo μ_i é fundamental para a estimação de δ livre de viés. A ausência do efeito fixo faria com que diferenças nas características não observáveis dos indivíduos correlacionadas com a alocação entre terceirizados e próprios tornassem nosso estimador viesado e inconsistente. Como discutido anteriormente, é possível que as empresas que empregam mão de obra terceirizada sejam menos intensas em habilidade ou que sejam intrinsicamente diferentes e, portanto, contratariam trabalhadores com salários naturalmente diferentes. O efeito fixo, porém, controla por essas características dos trabalhadores não observáveis e fixas ao longo do tempo, eliminando a correlação espúria.

Quando expusemos os argumentos teóricos que levariam à segmentação do mercado de trabalho entre trabalhadores terceirizados e próprios, notamos que aspectos institucionais do mercado de trabalho poderiam explicar esse fenômeno. O que faria um trabalhador aceitar um salário mais baixo na condição de terceirizado? Uma explicação bastante razoável seria uma diferença no poder de barganha do trabalhador. Por exemplo, trabalhadores próprios podem ser representados por sindicatos de categoria diferente à dos terceirizados, e então diferentes condições de negociação de acordos e convenções coletivas surgiriam. Imaginemos um vigia que resida na região do ABC Paulista, que tradicionalmente abriga muitas indústrias do setor de autopeças. Caso seja terceirizado, esse vigia deixará de ser representado pelo Sindicato dos Metalúrgicos do ABC, uma instituição tradicionalmente forte. Nesse caso, as condições de negociação seriam afetadas e o impacto disso nos salários poderia ser negativo.

Assim, é razoável supor que diferentes ocupações sintam impactos diferentes da terceirização. Para avaliar essa possibilidade, estimamos o seguinte modelo:

$$w_i = \alpha_0 + \delta o_{it} + \phi_j o_{it} V_{ij} + \beta x_{it} + \mu_i + e_{it} \quad (34)$$

onde V_{ij} é um vetor de *dummies* das ocupações de interesse j e ϕ_j é um vetor de coeficientes que mede o impacto da terceirização em cada ocupação j separadamente¹⁴. Para que a equação acima seja estimada, é necessário definir uma ocupação como referência, para que as demais $j-1$ ocupações tenham suas *dummies* incluídas na regressão, interagindo também com o_{it} . Nos resultados apresentados nas próximas seções, a categoria de referência foi *Montagem e manutenção de equipamentos*.

Algumas ressalvas devem ser feitas quanto à estratégia adotada. Primeiro, ao incluir o efeito fixo, não podemos incluir na regressão variáveis de interesse que não mudam ao longo do tempo, como por exemplo raça ou gênero. Segundo, a identificação só é possível porque há indivíduos que transitam entre terceirização e contratação própria ao longo do tempo.

¹³ Cerca de 10% dos trabalhadores registram mudança nessa variável ao longo dos 4 anos de painel.

¹⁴ O efeito da terceirização em cada ocupação j será dado pela soma $\delta + \phi_j$.

Portanto, o efeito da terceirização nos salários será medido apenas para esses trabalhadores que transitaram. No entanto, é possível que haja algum mecanismo de seleção que faça com que indivíduos com características específicas transitem mais frequentemente entre os tipos de emprego, o que traria algum viés para nossas estimativas. A Tabela 3 mostra a matriz de transição entre a contratação direta e a terceirização. Nota-se que cerca de 8% dos trabalhadores que eram terceirizados em um dado ano assumem uma condição de contratação direta no ano seguinte. Por outro lado, apenas pouco mais de 2,4% dos contratados diretamente tornam-se terceirizados no ano seguinte. Na seção 5 apresentamos alguns testes que indicam que os resultados não se alteram de acordo com o status inicial dos trabalhadores, mas esse fato não garante por si que não haja algum mecanismo que faça com que a contratação de terceirizados seja endógena. Caso a probabilidade de se tornar terceirizado seja afetada por características não observáveis, então nossas estimativas podem estar potencialmente afetadas se o efeito dessas características não observáveis na probabilidade de se tornar terceiro não for fixo no tempo. Nossa hipótese é de que isso não ocorre e que, ao longo do tempo, o efeito fixo consegue captar as diferenças na probabilidade de terceirização.

Tabela 3 – Transições de e para a terceirização

Tipo de contratação no ano inicial	Tipo de contratação no ano seguinte		
	TERCEIRIZADO	NÃO TERCEIRIZADO	NÚMERO DE OBSERVAÇÕES
TERCEIRIZADO	91,6%	8,4%	7.738.912
NÃO TERCEIRIZADO	2,4%	97,6%	2.319.312
NÚMERO DE OBSERVAÇÕES	7.748.313	2.309.911	10.058.224

Fonte: RAIS (MTE).Elaboração própria.

Nota: A remuneração em 31/dezembro de cada ano foi corrigida pelo IPCA para dezembro de 2012.

4. Resultados

A Tabela 4 exibe os resultados para a estimativa do diferencial de salários entre terceirizados e diretamente contratados. A primeira coluna mostra o coeficiente encontrado através da regressão de w em uma constante e na variável *dummy* de terceirização. Como esperado, o diferencial (não condicional) é bastante negativo (-17%). Já a segunda coluna da mesma tabela mostra que o diferencial de salários cai para -12% quando controlamos por diversas características observáveis.

Por outro lado, a terceira coluna da Tabela 4 nos leva a uma conclusão diferente a respeito do diferencial de salários entre os terceirizados e os trabalhadores próprios. A diferença média na remuneração de um mesmo indivíduo que muda de emprego e transita de uma forma de contratação para a outra é de -3%, uma diferença significativamente menor do que o encontrado nos dois exercícios anteriores.

Tabela 4 – Efeitos da terceirização nos salários

	MQO (sem controles)	MQO (com controles)	Efeito fixo
δ	-0.1653*** (0.0004)	-0.1155*** (0.0004)	-0.0298*** (0.0007)
Observações	18.304.233	14.515.798	18.301.656
Indivíduos	-	-	8.245.683
R ² (ajustado/total)	0,0092	0,6049	0,1983

Fonte: RAIS 2007-2012 (MTE). Elaboração própria.

Nota: Erros padrões entre parênteses. A regressão estimada via MQO sem controles tem o logaritmo do salário real como variável dependente e apenas uma constante e a dummy 'terceirizado' como regressores; a regressão com controles utiliza também dummies de UF, setor, família ocupacional, gênero, cor, nível de escolaridade, tamanho do estabelecimento e ano, além de idade, horas contratadas e número de dias de afastamento.

* significativo a 10%; ** significativo a 5%; *** significativo a 1%.

Em seguida, buscamos responder se o diferencial é heterogêneo entre as ocupações/serviços. A Tabela 5 abaixo mostra que de fato há bastante diferença quando desagregamos os resultados. Para encontrar o diferencial total devemos somar o coeficiente associado à *dummy* terceirização ao coeficiente da interação entre terceirização e ocupação (com exceção do caso da categoria de referência, *Montagem e manutenção de equipamentos*).

O maior diferencial nos salários foi encontrado para os trabalhadores de *Telemarketing* que transitam entre as duas formas de contratação com -12% no salário mensal. Trabalhadores de *Limpeza e conservação*, *Montagem e manutenção de equipamentos* e TI também apresentam um diferencial negativo, todos ao redor de -5%. Por outro lado, os trabalhadores das atividades de Segurança/vigilância recebem, em média, 5% a mais quando são terceirizados. Já os trabalhadores envolvidos em atividades de P&D têm diferenciais que não são significativos estatisticamente. A próxima seção mostra como esses prêmios ou penalizações evoluíram ao longo do tempo.

Esses resultados mostram que os efeitos da terceirização são de fato heterogêneos. Atividades que requerem baixa qualificação parecem resultar em remunerações inferiores para os terceirizados. Já atividades de alta qualificação vão na direção contrária e remuneram de maneira igual ou melhor os trabalhadores terceirizados. Uma hipótese para encontrarmos esses resultados pode ser a de que a arbitragem só funciona no caso de ocupações de média e alta qualificações, enquanto haveria maior poder de barganha para os empregadores no caso das ocupações de menor qualificação. Outra possibilidade (não excludente) pode ser a necessidade de determinadas ocupações por especialização e/ou em capital humano específico. Quando for esse o caso, então os trabalhadores podem ter vantagem em se especializar e acumular conhecimento e experiência atendendo diferentes empresas ao longo do tempo. Esse pode ser o caso, também, de ocupações cuja terceirização é aceita tacitamente pela Justiça Trabalhista, como por exemplo advogados, publicitários ou contadores.

Tabela 5 – Efeitos da terceirização nos salários de acordo com as ocupações terceirizadas

Ocupação	Coefficiente	MQO (sem controles)	MQO (com controles)	Efeito fixo
MONTAGEM E MANUTENÇÃO DE EQUIPAMENTOS	δ	-0,1191*** (0,0017)	-0,0421*** (0,0023)	-0,0557*** (0,0025)
SEGURANÇA/VIGILÂNCIA	$\delta + \phi_{SV}$	-0,0153*** (0,0001)	-0,1365*** (0,0010)	0,0508*** (0,0012)
TECNOLOGIA DA INFORMAÇÃO	$\delta + \phi_{TI}$	0,0140*** (0,0018)	-0,0279*** (0,0026)	-0,0510*** (0,0033)
LIMPEZA E CONSERVAÇÃO	$\delta + \phi_L$	-0,1255*** (0,0004)	-0,1317*** (0,0005)	-0,0555*** (0,0007)
PESQUISA E DESENVOLVIMENTO	$\delta + \phi_{P\&D}$	0,1661*** (0,0054)	-0,1984*** (0,0079)	-0,0183 (0,0118)
TELEMARKETING	$\delta + \phi_{TM}$	-0,3110*** (0,0009)	-0,0246*** (0,0025)	-0,1171*** (0,0022)
Observações		19.672.902	14.515.798	18.301.656
Indivíduos		-	-	8.245.683
R ² (ajustado/total)		0,3421	0,4572	0,2055

Fonte: RAIS 2007-2012 (MTE). Elaboração própria. Nota: Erros padrões entre parênteses. A regressão estimada via MQO sem controles tem o logaritmo do salário real como variável dependente e apenas uma constante e a dummy 'terceirizado' como regressores; a regressão com controles utiliza também dummies de UF, setor, família ocupacional, gênero, cor, nível de escolaridade, tamanho do estabelecimento e ano, além de idade, horas contratadas e número de dias de afastamento.

* significativo a 10%; ** significativo a 5%; *** significativo a 1%.

5. Testes de robustez

Esta seção avalia a robustez dos resultados apresentados anteriormente de duas formas distintas. Primeiro, avaliamos se os resultados se alteram quando utilizamos períodos menores dentro de nossa amostra. Ao invés de utilizar os seis anos para montarmos um único painel, dividiremos a base de dados em três painéis distintos: um com as observações de 2007 e 2008, outro para 2009 e 2010 e um último com os dados de 2011 e 2012. Uma das motivações para essa separação da base é que, quando dividimos a amostra em intervalos de tempo menores, temos menos atrito nos dados e isso nos permite realizar o exercício econométrico com painéis balanceados. Além disso, esse exercício é particularmente interessante porque avaliaremos se o período 2007-2008, pré e durante a crise financeira internacional, apresentou resultados distintos dos períodos subsequentes. Nesse período inicial, pode-se considerar que um choque exógeno (a crise) afetou o mercado de trabalho e, portanto, fez com que parte importante dos desligamentos ocorresse exogenamente. Essa variação pode ter causado alguma seleção entre os trabalhadores que permaneceram no mercado de trabalho em 2009, ou mesmo naqueles que regressaram para as mesmas ocupações depois disso. Como vemos na Tabela 6, os apontam para uma redução do diferencial negativo dos terceirizados, particularmente para a subamostra 2011/2012. Esse período foi marcado por uma redução significativa da taxa

de desemprego e pela expansão do setor de serviços como um todo, o que pode ter elevado a demanda por trabalhadores dos serviços analisados. De qualquer forma, o padrão observado anteriormente se mantém: ocupações de baixa qualificação registram remuneração menor para os terceirizados, enquanto as ocupações de maior qualificação registram salários iguais ou superiores para a forma de contratação terceirizada.

Tabela 6 – Efeitos da terceirização nos salários em subamostras (2007-2008, 2009-2010 e 2011-2012)

Ocupação	Coeficiente	2007-2012	2007-2008	2009-2010	2011-2012
MONTAGEM E MANUTENÇÃO DE EQUIPAMENTOS	δ	-0,0557*** (0,0025)	-0,0350*** (0,0066)	-0,0077 (0,0063)	-0,0006 (0,0109)
SEGURANÇA/VIGILÂNCIA	$\delta + \phi_{SV}$	0,0508*** (0,0012)	0,0647*** (0,0039)	0,0467*** (0,0037)	0,0562*** (0,0051)
TECNOLOGIA DA INFORMAÇÃO	$\delta + \phi_{TI}$	-0,0510*** (0,0033)	-0,0633*** (0,0088)	-0,0201*** (0,0060)	0,1952*** (0,0262)
LIMPEZA E CONSERVAÇÃO	$\delta + \phi_L$	-0,0555*** (0,0007)	-0,0419*** (0,0027)	-0,0538*** (0,0021)	0,0002 (0,0019)
PESQUISA E DESENVOLVIMENTO	$\delta + \phi_{P\&D}$	-0,0183 (0,0118)	-0,0246 (0,0434)	-0,0065 (0,0298)	-0,0969*** (0,0370)
TELEMARKETING	$\delta + \phi_{TM}$	-0,1171*** (0,0022)	-0,1310*** (0,0077)	-0,1357*** (0,0092)	-0,0046 (0,0174)
Observações		18.301.656	5.468.338	6.746.435	19.670.099
Indivíduos		8.245.683	4.121.621	4.641.320	8.687.580
R ² (ajustado/total)		0,2055	0,1580	0,1067	0,5295

Fonte: RAIS 2007-2012 (MTE). Elaboração própria.

Nota: Erros padrões entre parênteses.

* significativo a 10%; ** significativo a 5%; *** significativo a 1%.

Outro teste importante busca esclarecer se há tendências heterogêneas ao longo do tempo para os trabalhadores terceirizados em relação aos não terceirizados. Estimaremos a seguinte equação:

$$w_{it} = \alpha_0 + \gamma_j V_{ij} + \delta o_{it} + \phi_j o_{it} V_{ij} + \beta x_{it} + \sum_t \tau_t w_i^1 d_t + \sum_t \varphi_t o_i^1 d_t + \mu_i + e_{it} \quad (35)$$

onde o_i^1 é o valor de o_{it} na primeira observação do indivíduo i e w_i^1 é o logaritmo da remuneração do trabalhador na mesma primeira observação.

Portanto, a equação acima repete o modelo estimado anteriormente, acrescentando dois novos conjuntos de variáveis. Primeiro, entram interações entre as *dummies* de ano e o primeiro salário observado para o indivíduo. Além disso, acrescentamos também interações entre *dummies* de ano e a forma de contratação inicial, isto é, a *dummy* terceirizado ou próprio no primeiro ano em que o trabalhador aparece na base de dados. Com isso, flexibilizamos o modelo para permitir que aqueles que inicialmente são terceirizados ou próprios tenham tendências diferentes ao longo do tempo. Assim, o efeito da terceirização

medido por δ é aquele para além da tendência de cada indivíduo. Na estimação da equação (35), $\hat{\delta} = -0.0336$ com intervalo de confiança (a 95%) entre -0.03488 e -0.0323. O tamanho grande da base de dados ajuda na obtenção de erros padrões pequenos: esse intervalo de confiança não contém o valor encontrado anteriormente (-0.0298). Mesmo assim, é seguro dizer que as conclusões não mudam e permanecem válidas mesmo sob a hipótese de que terceirizados têm trajetórias de salários diferentes das apresentadas pelos não terceirizados.

6. Conclusões e considerações finais

Este trabalho estimou o diferencial de salários entre a mão de obra terceirizada e os trabalhadores contratados diretamente pelos tomadores dos serviços. As ocupações ou serviços considerados para a análise foram *Montagem e manutenção de equipamentos, Segurança/Vigilância, TI, Limpeza e conservação, P&D e Telemarketing*.

Os resultados indicam que, numa comparação não condicional, os trabalhadores terceirizados recebem em média um salário 17% menor do que no caso em que a contratação é direta. Esse diferencial se mantém quando controlado pelas características observáveis dos trabalhadores, mas é substancialmente reduzido para cerca de -3% quando controlado pelo efeito fixo dos indivíduos. Ou seja, as características não observáveis exercem um papel relevante na seleção e na determinação da remuneração dos terceirizados.

Outro fato importante é a heterogeneidade dos diferenciais. De maneira geral, ocupações de baixa qualificação (como *Telemarketing* e *Limpeza*) oferecem remunerações menores aos terceirizados. Porém, ocupações de alta qualificação e que necessitam de acúmulo de capital humano específico, como P&D e TI, pagam salários maiores aos terceirizados.

Esses resultados se baseiam no salário em carteira informado através da RAIS (MTE) e, portanto, não levam em conta outros benefícios (pecuniários ou não). Também não são medidas diferenças entre as condições de trabalho nas duas formas de contratação (acesso a vestiário, banheiros, alimentação em restaurante etc.). Porém, os resultados são robustos e permitem avaliarmos algumas opções de políticas. Vale comentar também que os resultados são medidos nos salários e não há nenhum teste no número de empregos (i.e. efeito quantidade). É possível que serviços nos quais os terceirizados têm salários mais baixos, o nível do emprego seja maior exatamente por conta de o 'preço' (salário) ser menor.

Visto que o diferencial de salários parece estar associado à qualificação da mão de obra e não propriamente ao serviço realizado, parece razoável que a atuação do Estado não se dê na definição dos limites das firmas, mas sim nas condições de negociações entre as partes envolvidas no contrato de trabalho. Ao invés de basear a terceirização na definição muitas vezes subjetiva do que é atividade-meio, o Estado poderia criar mecanismos institucionais para que o poder de barganha dos trabalhadores de baixa qualificação seja mais equilibrado ao dos empregadores. Uma alternativa seria oferecer ao trabalhador a liberdade e o direito de escolher qual sindicato lhe representa. Com isso, os trabalhadores menos qualificados poderiam buscar equilibrar as condições de negociação com a empresa através da associação a sindicatos mais fortes.

Referências

Botelho, F.; Ponczek, V. (2011) "Segmentation in the Brazilian Labor Market," *Economic Development and Cultural Change*, University of Chicago Press, vol. 59(2), pages 437 - 463.

Central Única dos Trabalhadores – CUT. (2014) “Terceirização e desenvolvimento: uma conta que não fecha: dossiê acerca do impacto da terceirização sobre os trabalhadores e propostas para garantir a igualdade de direitos”, Secretaria Nacional de Relações de Trabalho e Departamento Intersindical de Estatística e Estudos Socioeconômicos. - São Paulo.

Coase, Ronald H. (1937) “The nature of the firm.” *economica* 4.16 (1937): 386-405.

Confederação Nacional dos Bispos do Brasil – CNBB. (2015). “Nota da CNBB sobre o momento nacional”, 53^a Assembleia Geral da CNBB. Disponível em <http://www.cnbb.org.br/eventos-1/assembleia-geral-1/16376-cnbb-divulga-nota-sobre-o-momento-nacional>. Acesso em 11/6/2015.

Confederação Nacional da Indústria – CNI (2014). “Sondagem Especial: Terceirização”, *Sondagem Especial*, vol. 4(2), Julho.

Dube, A.; Kaplan, E. (2010). "Does Outsourcing Reduce Wages in the Low-Wage Service Occupations? Evidence from Janitors and Guards," *Industrial and Labor Relations Review*, ILR Review, Cornell University, ILR School, vol. 63(2), pages 287-306, January.

Eichhorst, W. (2015). "Do We Have to Be Afraid of the Future World of Work?," *IZA Policy Papers* 102, Institute for the Study of Labor (IZA).

Menezes Filho, Naércio Aquino; Mendes, Marcos; Almeida, Eduardo Simões de. (2004). “O diferencial de salários formal-informal no Brasil: segmentação ou viés de seleção?”. *Revista Brasileira de Economia*, 58(2), 235-248.

Zimmermann, Klaus. F. (2015). "The Big Trade-Off in the World of Labor," *IZA Policy Papers* 100, Institute for the Study of Labor (IZA).

Anexo

Códigos CBO considerados para a definição de terceirização

Montagem e manutenção de equipamentos

CBO	Descrição
1238	Diretores de manutenção
9511	Eletricistas de manutenção eletroeletrônica
1427	Gerentes de manutenção e afins
9513	Instaladores e mantenedores de sistemas eletroeletrônicos de segurança
7321	Instaladores e reparadores de linhas e cabos elétricos, telefônicos e de comunicação de
7313	Instaladores-reparadores de linhas e equipamentos de telecomunicações
9141	Mecânicos de manutenção aeronáutica
9193	Mecânicos de manutenção de bicicletas e equipamentos esportivos e de ginástica
9111	Mecânicos de manutenção de bombas, motores, compressores e equipamentos de transmissão
9113	Mecânicos de manutenção de máquinas industriais
9131	Mecânicos de manutenção de máquinas pesadas e equipamentos agrícolas
9112	Mecânicos de manutenção e instalação de aparelhos de climatização e refrigeração
7312	Montadores de aparelhos de telecomunicações
7311	Montadores de equipamentos eletroeletrônicos
7257	Instaladores de equipamentos de refrigeração e ventilação
7252	Montadores de máquinas industriais
7253	Montadores de máquinas pesadas e equipamentos agrícolas
7251	Montadores de máquinas, aparelhos e acessórios em linhas de montagem
7254	Mecânicos montadores de motores e turboalimentadores
7202	Supervisores da fabricação e montagem metalmecânica
9501	Supervisores de manutenção eletroeletrônica industrial, comercial e predial
9503	Supervisores de manutenção eletromecânica
9101	Supervisores em serviços de reparação e manutenção de máquinas e equipamentos industriais, famí
9153	Técnicos em manutenção e reparação de equipamentos biomédicos
9151	Técnicos em manutenção e reparação de instrumentos de medição e precisão
3141	Técnicos mecânicos na fabricação e montagem de máquinas, sistemas e instrumentos
3144	Técnicos mecânicos na manutenção de máquinas, sistemas e instrumentos
9192	Trabalhadores de manutenção de roçadeiras, motosserras e similares

Segurança/Vigilância

CBO	Descrição
5172	Policiais, guardas-civis municipais e agentes de trânsito
5174	Porteiros, vigias e afins
2526	Profissionais da administração dos serviços de segurança
5103	Supervisores dos serviços de proteção, segurança e outros
5173	Vigilantes e guardas de segurança

TI

CBO	Descrição
2124	Analistas de tecnologia da informação
1236	Diretores de serviços de informática
2122	Engenheiros em computação
2123	Administradores de tecnologia da informação
1425	Gerentes de tecnologia da informação
3171	Técnicos de desenvolvimento de sistemas e aplicações

Limpeza e conservação

CBO	Descrição
5133	Camareiros, roupeiros e afins
5143	Trabalhadores nos serviços de manutenção de edificações
5142	Trabalhadores nos serviços de coleta de resíduos, de limpeza e conservação de áreas públicas

P&D

CBO	Descrição
1237	Diretores de pesquisa e desenvolvimento
1426	Gerentes de pesquisa e desenvolvimento e afins
2030	Pesquisadores das ciências biológicas
2034	Pesquisadores das ciências da agricultura
2033	Pesquisadores das ciências da saúde
2031	Pesquisadores das ciências naturais e exatas
2035	Pesquisadores das ciências sociais e humanas
2032	Pesquisadores de engenharia e tecnologia
3951	Técnicos de apoio em pesquisa e desenvolvimento

Telemarketing

CBO	Descrição
4223	Operadores de telemarketing

Códigos CNAE considerados para a definição de terceirização

Montagem e manutenção de equipamentos

CNAE	Descrição
331	Manutenção e reparação de máquinas e equipamentos
332	Instalação de máquinas e equipamentos
951	Reparação e manutenção de equipamentos de informática e comunicação
952	Reparação e manutenção de objetos e equipamentos pessoais e domésticos

Segurança/Vigilância

CNAE	Descrição
801	Atividades de vigilância, segurança privada e transporte de valores

TI

CNAE	Descrição
620	Atividades dos serviços de tecnologia da informação

Limpeza e conservação

CNAE	Descrição
811	Serviços combinados para apoio a edifícios
812	Atividades de limpeza
970	Serviços domésticos

P&D

CNAE	Descrição
721	Pesquisa e desenvolvimento experimental em ciências físicas e naturais
722	Pesquisa e desenvolvimento experimental em ciências sociais e humanas

Telemarketing

CNAE	Descrição
822	Atividades de teleatendimento

Capítulo 4: Painéis de rotação e trajetórias inconsistentes: Fontes de viés na Pesquisa Mensal de Emprego do IBGE

Resumo

Essa nota documenta dois problemas comuns em painéis amostrais rotativos e que são existentes na Pesquisa Mensal de Emprego (PME/IBGE): o viés de grupo rotacional e trajetórias de emprego inconsistentes. No primeiro caso, a taxa de desemprego é inversamente correlacionada ao número da entrevista, sistematicamente ao longo de todo o período para o qual os dados são disponibilizados. Já no segundo caso, erros na informação reportada pelo entrevistado fazem com que sejam observadas transições espúrias no mercado de trabalho.

1. Introdução

Esta nota documenta dois aspectos importantes da Pesquisa Mensal de Emprego (PME/IBGE) para a análise do mercado de trabalho: o chamado *rotating group bias*, ou viés do painel rotacional; e a presença de trajetórias inconsistentes no mercado de trabalho que, caso não sejam tratadas, podem levar ao cômputo de transições espúrias.

2. Viés do painel rotacional

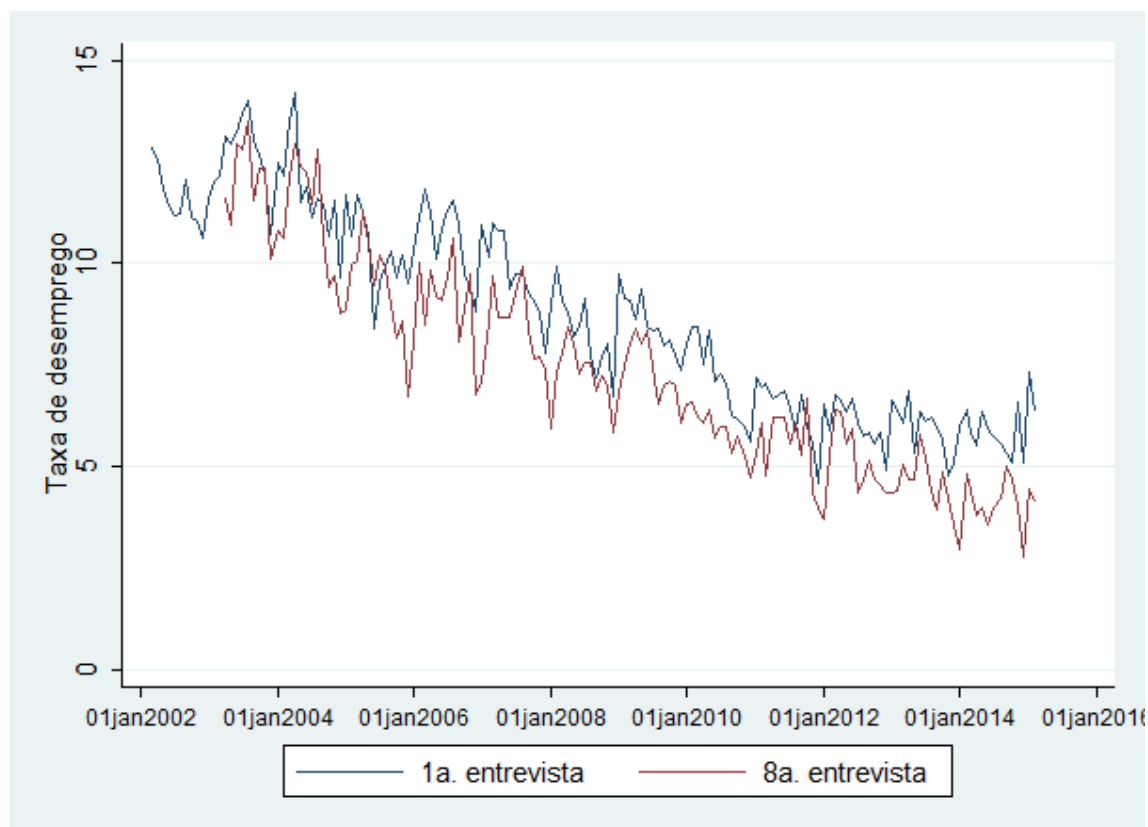
Muitos países têm à sua disposição painéis amostrais de indivíduos para aprimorar o acompanhamento e entendimento do mercado de trabalho local. Frequentemente, esses painéis são rotativos, nos quais os domicílios são entrevistados durante alguns meses, ficam fora da amostra por outro período e finalmente retornam para uma sequência final de novas entrevistas. No caso da PME, o esquema de rotação é 4-8-4, ou seja, os domicílios são entrevistados por quatro meses consecutivos, ficam fora da amostra por 8 meses e voltam para mais quatro entrevistas mensais.

Bailar (1975) foi um dos primeiros a documentar, para o caso da *Current Population Survey* (CPS) americana, a presença do que chamou de *rotating group bias*, ou simplesmente viés do painel rotacional. Ela notou que para um dado período, a taxa de desemprego estimada era sistematicamente diferente de acordo com o número da entrevista dos domicílios. Recentemente, Krueger et al (2014) mostraram que o problema se acentuou após a reformulação do questionário da CPS.

No Brasil, ainda não há registros sobre a existência desse possível viés. Ribas e Soares (2010) mostram que a PME tem uma presença grande de atrito e muitos domicílios deixam de responder as últimas entrevistas. Os autores mostram que o atrito não é aleatório e depende das características dos domicílios e medem o efeito desse atrito em dois modelos: um que mede probabilidades de transição no mercado de trabalho e outro de progressão escolar.

Outro impacto relevante é do atrito na própria taxa de desemprego. A Figura 1 mostra a evolução da taxa de desemprego média para os indivíduos que estavam na primeira e na oitava (última) entrevista. Já a Figura 2 mostra a taxa de desemprego média anual para cada um dos grupos rotacionais presentes na amostra no momento da pesquisa. Vê-se, claramente, um padrão onde a taxa de desemprego é inversamente correlacionada com o número da entrevista. Para testar o tamanho desse viés, regredimos a taxa de desemprego média de cada grupo rotacional contra o número da entrevista. Os resultados estão na Tabela 7.

Figura 1 – Taxa de desemprego mensal da PME para os indivíduos na primeira e oitava entrevistas



Fonte: PME (IBGE). Período considerado: mar/2002 a fev/2015. Elaboração própria.

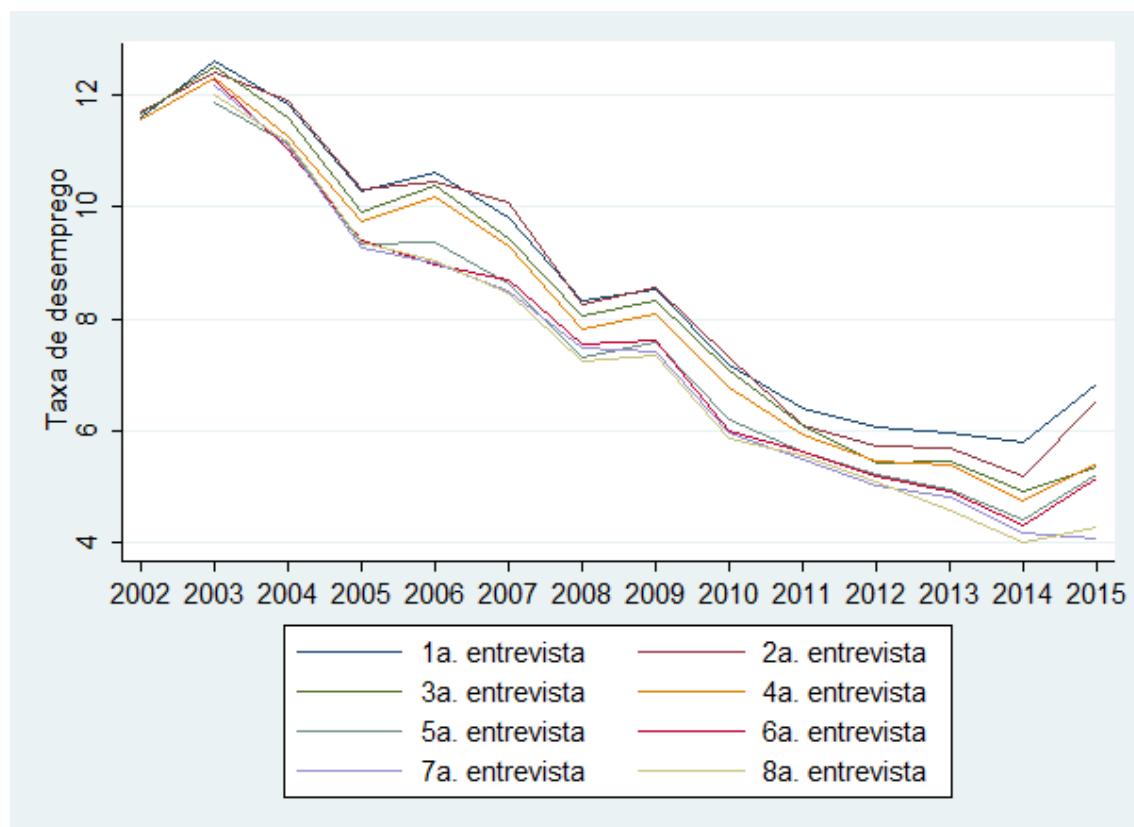
Tabela 7 – Viés de grupo rotacional na taxa de desemprego

Painel rotacional	Coefficiente	Erro padrão	P-valor
Entrevista = 1	(referência)	-	-
Entrevista = 2	-0,1031	0,2882	0,721
Entrevista = 3	-0,3656	0,2887	0,206
Entrevista = 4	-0,5638	0,2892	0,051
Entrevista = 5	-1,2079	0,2936	0,000
Entrevista = 6	-1,2490	0,2942	0,000
Entrevista = 7	-1,3955	0,2947	0,000
Entrevista = 8	-1,4720	0,2952	0,000
Constante	8,8124	0,2035	0,000
N	1.204		
R ² Ajustado	0,0467		

Fonte: PME (IBGE). Período considerado: mar/2002 a fev/2015.

Nota: os coeficientes são obtidos através de uma regressão entre a taxa de desemprego média para cada grupo de pessoas classificadas segundo o número da entrevista e dummies para cada um desses grupos.

Figura 2 – Taxa de desemprego (média anual) da PME para segundo o painel rotacional



Fonte: PME (IBGE). Período considerado: mar/2002 a fev/2015. Elaboração própria.

Esses resultados levantam dúvida quanto à real taxa de desemprego no Brasil. Também chama a atenção o fato de que a taxa de desemprego vem subindo particularmente entre aqueles que estão nas primeiras entrevistas. Não obstante o fato de que a PME tem cobertura geográfica limitada, fica claro que o valor normalmente aceito como a medida de taxa de desemprego pode estar viesado. Não é possível determinar, a priori, qual o sentido desse viés. Porém, duas hipóteses indicariam que a taxa de desemprego é maior do que o número normalmente divulgado pelo IBGE.

Primeiro, o próprio fato de o atrito não ser aleatório. Se os domicílios de maior probabilidade de desemprego forem os mais propensos a não completar as oito entrevistas, então ao menos parte do viés seria explicado pelo atrito. A Tabela 8 mostra que o total de pessoas que completa as oito entrevistas é, de fato, pequeno (menos da metade do total de indivíduos).

Além disso, Krueger et al (2014) propõem também uma segunda explicação, relacionada à uma situação embaraçosa, na qual desocupados tenderiam a reportar uma condição de ocupação não verdadeira. Independentemente do motivo, conclui-se que é importante que se leve em consideração o viés do painel rotacional no momento de se calcular a taxa de desemprego.

Tabela 8 – Distribuição do total de entrevistas completadas na PME

Quantidade de entrevistas registradas	Quantidade de pessoas	%
1	243,59	2,1%
2	357,137	3,1%
3	584,351	5,0%
4	2,607,393	22,4%
5	430,309	3,7%
6	562,512	4,8%
7	1,053,324	9,1%
8	5,784,749	49,8%
Total	11,623,365	100,0%

Fonte: PME (IBGE). Período considerado: 2003 a 2014. Elaboração própria.

3. Trajetórias inconsistentes no mercado de trabalho

Um segundo aspecto relevante da PME que essa nota busca documentar é o da existência de trajetórias inconsistentes. Entende-se que uma trajetória no mercado de trabalho é inconsistente quando o indivíduo reporta uma determinada posição na ocupação em um dado mês e no mês subsequente sua posição na ocupação e a duração de seu vínculo são incompatíveis.

Inicialmente, vamos analisar a matriz de transição observada com os dados da PME (**Erro! fonte de referência não encontrada.**). Note que a matriz é semelhante ao que normalmente é reportado na literatura nacional (ver, por exemplo, De Moura e Filho, 2014). Porém, chama a atenção nessa tabela a baixa aderência dos indivíduos em suas posições na ocupação. Por exemplo, dos trabalhadores que se declararam empregados formais em m dado mês, cerca de 10% deixaram essa condição no mês seguinte em média entre 2003 e 2014, um número aparentemente elevado mesmo para o padrão de rotatividade que encontramos no Brasil. Também os entre os desempregados, 31% deixam a PEA entre dois meses consecutivos, outro número potencialmente elevado.

Tabela 9 – Matriz de transição mensal sem correção (PME/IBGE)

Pos. Ocup em t	Fora da Força de Trabalho	Desempregado	Conta própria	Empregador	Empregado informal	Empregado formal
Fora da Força de Trabalho	85,8%	5,1%	3,0%	0,2%	3,2%	2,6%
Desempregado	30,9%	48,5%	4,3%	0,2%	8,4%	7,7%
Conta própria	8,8%	1,7%	74,3%	3,9%	7,2%	4,0%
Empregador	3,4%	0,4%	16,8%	70,7%	4,1%	4,5%
Empregado informal	9,1%	3,2%	7,2%	1,0%	64,8%	14,7%
Empregado formal	3,1%	1,1%	1,3%	0,3%	4,1%	90,2%

Fonte: PME (IBGE). Período considerado: 2003 a 2014. Elaboração própria.

Uma possível explicação para essas elevadas probabilidades de transição pode ser a existência de transições espúrias. O exemplo abaixo ilustra um caso real encontrado no painel da PME de um indivíduo da RM de Recife que em novembro de 2004 foi registrado como trabalhador por conta própria há dez anos. No mês seguinte, o mesmo indivíduo aparece como trabalhador informal há 11 anos e, na sequência, a posição na ocupação muda para empregado formal, sendo que o vínculo durava 10 anos.

id	month	year	age	sex	education	HH_status	Occupation Status	Region	Race	duration	interview
14	11	2004	29	Male	Primary	Chief of HH	Self employed	Recife	Pardo	120	2
14	12	2004	29	Male	Secondary	Chief of HH	Informal employee	Recife	Pardo	132	3
14	1	2005	30	Male	Secondary	Chief of HH	Formal employee	Recife	Pardo	120	4

Para corrigir esse problema, propõe-se nessa nota um algoritmo simples:

1. Na segunda ou terceira entrevista, compara-se a posição na ocupação ($POSOCUP_t$) com as posições na ocupação do mês imediatamente anterior ($POSOCUP_{t-1}$) e imediatamente posterior ($POSOCUP_{t+1}$). Caso ($POSOCUP_{t-1} = POSOCUP_{t+1}$) e ($POSOCUP_t \neq POSOCUP_{t+1}$), analisamos a duração do vínculo em $t+1$. Caso seja maior do que dois meses, então $POSOCUP_t$ é entendida como inconsistente e substituída por $POSOCUP_{t+1}$.
2. Na primeira entrevista, a comparação é feita apenas entre $POSOCUP_t$ e $POSOCUP_{t+1}$.
3. Na quarta entrevista, a comparação é feita somente entre $POSOCUP_t$ e $POSOCUP_{t-1}$.

Esse procedimento resulta em alterações nas posições na ocupação para 5,3% das observações e altera a matriz de transição. Após a correção, a matriz de transição apresenta maior aderência para as posições na ocupação em geral, como mostra a **Erro! Fonte de eferência não encontrada..**

Tabela 10 – Matriz de transição mensal após a correção das trajetórias (PME/IBGE)

Pos. Ocup em t	Fora da Força de Trabalho	Desempregado	Conta própria	Empregador	Empregado informal	Empregado formal
Fora da Força de Trabalho	91,3%	3,0%	1,7%	0,1%	2,1%	1,7%
Desempregado	19,5%	63,7%	3,0%	0,2%	6,7%	6,9%
Conta própria	4,9%	1,1%	85,5%	2,1%	4,0%	2,3%
Empregador	1,9%	0,3%	8,7%	84,2%	2,4%	2,5%
Empregado informal	5,7%	2,5%	4,1%	0,6%	78,8%	8,4%
Empregado formal	2,0%	0,9%	0,8%	0,2%	2,3%	93,9%

Fonte: PME (IBGE). Período considerado: 2003 a 2014. Elaboração própria.

Após as correções, as transições tornam-se menos frequentes. Como vimos no Capítulo 1, não atentar para as transições espúrias pode fazer trazer imprecisão a análises que utilizam a estrutura em painel da PME.

4. Considerações finais

Esta nota teve como objetivo o registro de duas características relevantes da Pesquisa Mensal de Emprego (IBGE) que ainda não estavam devidamente documentadas na literatura nacional.

O primeiro aspecto levantado diz respeito ao viés de grupo rotacional. Esse viés se manifesta ao observarmos que a taxa de desemprego é maior entre os indivíduos que estão nas primeiras entrevistas do painel rotativo. Não apenas o nível da taxa de desemprego parece estar viesado, como também a trajetória da taxa pode ser afetada (particularmente nos últimos meses).

Além disso, também se registrou a presença de trajetórias inconsistentes dos indivíduos no mercado de trabalho. A inconsistência surge quando se compara mudança nas posições na ocupação e a duração reportada dos vínculos. Não corrigir essas transições espúrias pode levar a inconsistência e viés em estudos que utilizam a estrutura de painel da PME. Para fazer a correção, essa nota trouxe um algoritmo que compara as posições na ocupação em meses sequenciais com a duração dos vínculos. O algoritmo está em uma versão preliminar e inicial e deve ainda ser refinado e sofisticado. De qualquer maneira, a correção parece levar a uma estrutura de transições mais consistente com os demais dados de mercado de trabalho observados para o Brasil.

Referências

Bailar, Barbara A. (1975). "The effects of rotation group bias on estimates from panel surveys." *Journal of the American Statistical Association* 70.349 (1975): 23-30.

de Moura, Rodrigo Leandro e Filho, Fernando Holanda Barbosa, (2014), EVOLUÇÃO RECENTE DA INFORMALIDADE NO BRASIL: UMA ANÁLISE SEGUNDO CARACTERÍSTICAS DA OFERTA E DEMANDA DE TRABALHO, Anais do XLI Encontro Nacional de Economia, ANPEC - Associação Nacional dos Centros de Pós-graduação em Economia.

Krueger, Alan, Alexandre Mas, and Xiaotong Niu. (2014). "The Evolution of Rotation Group Bias: Will the Real Unemployment Rate Please Stand Up?". No. w20396. National Bureau of Economic Research.

Ribas, Rafael Perez and Soares, Sergei Suarez Dillon. (2010). "O atrito nas pesquisas longitudinais: o caso da pesquisa mensal de emprego (PME/IBGE)". *Estud. Econ.* [online]. 2010, vol.40, n.1, pp. 213-244.