

UNIVERSITY OF SÃO PAULO
SCHOOL OF ECONOMICS, BUSINESS AND ACCOUNTING
DEPARTMENT OF ECONOMICS
GRADUATE PROGRAM IN ECONOMIC THEORY

AGGLOMERATION ECONOMIES AND LABOUR MARKETS IN BRAZIL
ECONOMIAS DE AGLOMERAÇÃO E MERCADOS DE TRABALHO NO BRASIL

Ana Maria Bonomi Barufi

Supervisor: Prof. Dr. Eduardo Amaral Haddad

São Paulo

2015

Prof. Dr. Marco Antônio Zago
President of the University of São Paulo

Prof. Dr. Adalberto Américo Fischman
Dean of the School of Economics, Business and Accounting

Prof. Dr. Hélio Nogueira da Cruz
Head of the Department of Economics

Prof. Dr. Márcio Issao Nakane
Coordinator of the Graduate Program in Economics

ANA MARIA BONOMI BARUFI

**AGGLOMERATION ECONOMIES AND LABOUR MARKETS IN BRAZIL
ECONOMIAS DE AGLOMERAÇÃO E MERCADOS DE TRABALHO NO BRASIL**

Doctoral dissertation submitted to the Graduate Program in Economic Theory of the Department of Economics of the School of Economics, Business and Accounting at the University of São Paulo in partial fulfilment of the requirements for the degree of Doctor of Sciences.

Supervisor: Prof. Dr. Eduardo Amaral Haddad

Corrected version

(Original version available at the School of Economics, Business and Accounting)

São Paulo

2015

FICHA CATALOGRÁFICA

Elaborada pela Seção de Processamento Técnico do SBD/FEA/USP

Barufi, Ana Maria Bonomi

Agglomeration economies and labour markets in Brazil / Ana Maria Bonomi Barufi. – São Paulo, 2015.

185 p.

Tese (Doutorado) – Universidade de São Paulo, 2015.

Orientador: Eduardo Amaral Haddad.

1. Economia urbana 2. Aglomerações urbanas 3. Mercado de trabalho I. Universidade de São Paulo. Faculdade de Economia, Administração e Contabilidade. II. Título.

CDD – 301.36

To my family.

ACKNOWLEDGMENTS

Firstly, I would like to thank my supervisor, Professor Eduardo Amaral Haddad, for the unconditional support, dedication, teaching, trust and friendship. A special thank for his capacity of challenging me and giving me new opportunities. Working with him has been delightful and a continuous learning process in all dimensions of life.

To my co-supervisor, Professor Peter Nijkamp, from the VU Amsterdam, for the great times in Amsterdam. It was an honour to spend a semester working with him. He will always be an inspiration for my future academic endeavours. To the Spatial Economics Department of the VU, for the most valuable interactions over that semester.

To Professor Serge Rey, Professor Luc Anselin and Julia Koschinsky from Arizona State University, for the opportunity to spend a semester at the GeoDA Center, that was extremely valuable and enriched my work.

I would like to thank CNPq and CAPES for the financial support during the Sandwich Programs in Arizona, United States, and Amsterdam, Netherlands. To FIPE, Eduardo Zylberstajn and Professor Hélio Zylbestajn for providing the data of the identified RAIS. To Professor Raul Silveira Neto for the friendship and for granting me access to RAIS MIGRA.

To Professors Carlos Azzoni, Naercio Menezes Filho, and André Chagas for the valuable comments in my qualifying examination that enriched significantly the final work. To all the members of NEREUS, especially Professors Joaquim Guilhoto and Danilo Iglioni, for all the valuable advices. To the Economic Department of the University of São Paulo for giving me this opportunity to become a better economist. To Pinho, Leka, Valéria, Cida and Francisco for all the help. To my amazing academic network, especially the founding members of the Regional Science Academy, with whom I had the honour to share thoughts and learn so much.

To the Macroeconomic Research Department of Bradesco Bank for all the support on this journey. To Octavio de Barros, mentor and friend, for all the opportunities. To Fernando Freitas, for all these years working together and building new ideas from scratch. To Myriã for all the help and support. To Ellen, Ana Paula, Leandro, Felipe, Matheus, Thiago, Fabiana, Priscila, Thomas and all others, for the daily laughs and friendship.

To my friends from the Arizonian desert, Simón, Miguel, Wellington, Kevin, Scott and Joanna, for a fun and productive semester. To my friends of the fantastic Amsterdam, Zeynep, Jamie, Ting, Iannis, Karima, Elfie, Panayiotis, Ifa, Monica, and all others that had such an impact in my life.

To my friends of a lifetime, Paula, Keiti, Genta, Jesus, Acauã, Leandro, Sérgio, Attuy, Gervásio, Weslem, Daniela, Vanessa, João Eduardo, and all others, who make it worth to fight for a better world. To Andrea, Juliana, Dario, Mirela Denise, Eric, and Moisés for sharing the anxieties of this journey. To Luis, Antonio and Anna for sharing the love for Palmeiras.

To my family, for the unconditional support, especially to my mother, Cristina Bonomi, that followed the whole process over these four years, always giving me comfort and love. To my siblings Paolo, Chico, Tiago, João, Gabi, Clara and Zeca, for everything.

To Desnes Rosario, for the love, patience, and respect. Building our life together has been the most enjoyable adventure.

AGRADECIMENTOS

Em primeiro lugar, gostaria de agradecer ao meu orientador, Professor Eduardo Amaral Haddad, pelo apoio incondicional, dedicação, ensinamentos, confiança e amizade. Um agradecimento especial pela sua capacidade de me desafiar e de me oferecer novas oportunidades. Trabalhar com ele tem sido muito prazeroso e um contínuo processo de aprendizado em todas as dimensões da vida.

Ao meu co-orientador, Professor Peter Nijkamp, da VU Amsterdam, pelos excepcionais momentos em Amsterdam. Foi uma honra poder passar um semestre trabalhando com ele. Ele sempre será uma inspiração para os meus futuros esforços acadêmicos. Ao Departamento de Economia Espacial da VU, pelas excepcionais interações.

Ao Professor Serge Rey, Professor Luc Anselin e Julia Koshinky, da Arizona State University, pela oportunidade de ficar um semestre no GeoDA Center, que permitiu enriquecer o trabalho.

Ao CNPq e à CAPES pelo apoio financeiro durante os Programas de Doutorado Sanduíche no Arizona e em Amsterdam. À FIPE, Eduardo Zylberstajn e Professor Hélio Zylberstajn por me dar acesso aos dados da RAIS identificada. Ao Professor Raul Silveira Neto pela amizade e por me disponibilizar os dados da RAIS MIGRA.

Aos Professores Carlos Azzoni, Naercio Menezes-Filho e André Chagas pelos valiosos comentários no Exame de Qualificação, que enriqueceram de maneira significativa o trabalho final. Aos membros do NEREUS, em especial os Professores Joaquim Guilhoto e Danilo Iglori, pelos valiosos conselhos. Ao Departamento de Economia da Universidade de São Paulo por me dar a oportunidade de me tornar uma economista melhor. Ao Pinho, Leka, Valéria, Cida e Francisco por toda a ajuda. À minha fenomenal rede acadêmica, especialmente aos membros fundadores da Regional Science Academy, com os quais tive a honra de compartilhar pensamentos e aprender muito.

Ao Departamento de Pesquisas e Estudos Econômicos do Bradesco, por todo o apoio nesta jornada. Ao Octavio de Barros, mentor e amigo, por todas as oportunidades. Ao Fernando Freitas, por todos os anos trabalhando junto e criando novas ideias. À Myriã, por toda a ajuda e apoio. À Ellen, Ana Paula, Leandro, Felipe, Matheus, Fabiana, Priscila, Thomas e todos os outros, pelas risadas diárias e pela amizade.

Aos amigos do deserto do Arizona, Simón, Miguel, Wellington, Kevin, Scott e Joanna, por um semestre divertido e produtivo. Aos amigos de Amsterdam, Zeynep, Jamie, Ting, Iannis, Karima, Elfie, Panayiotis, Ifa, Monica, e todos os outros que impactaram tanto em minha vida.

Aos amigos de uma vida, Paula, Keiti, Genta, Jesus, Acauã, Leandro, Sérgio, Attuy, Gervásio, Weslem, Daniela, Vanessa, João Eduardo, e todos os que fazem valer a pena buscar um mundo melhor. À Andrea, Juliana, Dario, Mirela, Denise, Eric e Moisés por compartilhar as ansiedades desta jornada. Ao Luis, Antonio e Anna, por compartilhar a paixão pelo Palmeiras.

À minha família, pelo apoio incondicional, especialmente à minha mãe, Cristina Bonomi, que acompanhou todo o processo ao longo desses quatro anos, sempre me confortando com muito amor. Aos meus irmãos Paolo, Chico, Tiago, João, Gabi, Clara e Zeca, por tudo.

Ao Desnes Rosario, pelo amor, paciência e respeito. Construir a nossa vida juntos tem sido a mais adorável das aventuras.

Cities have the capability of providing something for everybody, only because, and only when, they are created by everybody.

Jane Jacobs

*The Economy of Cities: The death
and life of great American cities*

ABSTRACT

Agglomeration economies have a relevant impact on local labour markets. The interaction of workers and firms in dense urban areas may generate productivity advantages that result in higher wages. They may be accompanied by an increase in local costs, but the existence of cities that continue to grow is a sign by itself that these gains supersede higher costs.

Therefore, large urban areas have an expected positive impact on wages. However, not only the size of the city but also the sectoral composition is relevant to understand locational choices of firms of a specific sector. The industrial scope of agglomeration economies is investigated in the first chapter of this dissertation, and the main results indicate that there is not a unique optimal local industrial mix to foster productivity in different technological sectors. Furthermore, high-tech and low-tech manufacturing sectors benefit more from urban scale in Brazil, followed by services associated with higher knowledge intensity. These sectors are supposed to locate relatively more in large urban areas in order to profit from these advantages.

Agglomeration economies may have static and dynamic effects for individuals. These effects are reinforced by a process of sorting of skilled workers into large urban areas. In fact, initial and return migration are mechanisms that select more skilled and more productive workers into large urban areas. Then, cities with a higher percentage of skilled workers attract more of these highly-qualified individuals. Second migration seems to reinforce these relations.

The estimation of static agglomeration economies indicate that the inclusion of individual fixed effects decreases density coefficient significantly. Then, dynamic agglomeration economies are estimated considering previous work experience in cities. In this case, static agglomeration advantages become insignificant and whenever years of previous experience are combined with the current place of work, individuals working in less dense cities who had previous experience in denser areas will benefit the most from these gains. Finally, controlling for worker heterogeneity previous experience has a relevant and positive impact on wage growth only in cities with at least the same density level of the current place of work.

Finally, city size has an important impact on the relative bargaining power of workers and firms in the labour market. When analysing the relationship of local wages and the business cycle, wage flexibility, measured by the wage curve, is higher in informal sectors in less dense areas. Therefore, large agglomerations are supposed to provide a higher bargaining power for workers, as they have further job opportunities.

All these results indicate that agglomeration economies in Brazil are likely to stimulate spatial concentration and increase regional inequalities. Workers and firms self-select themselves into agglomerated urban areas, in which they find a more diversified environment and a larger share of high-skilled individuals. Bigger centres also provide the conditions for workers to bargain for higher wages, even if they are in the informal sector.

RESUMO

Economias de aglomeração possuem um impacto importante sobre o mercado de trabalho. A interação entre trabalhadores e firmas em áreas de elevada densidade pode gerar ganhos de produtividade que resultam em salários mais elevados. Tais áreas também podem possuir custos de vida mais elevados, mas o crescimento recente das cidades parece indicar que os ganhos se sobrepõem aos custos.

Portanto, grandes áreas urbanas têm um impacto esperado positivo sobre os salários. No entanto, não só o tamanho da cidade, mas também a composição setorial é relevante para entender as escolhas de localização das empresas de um sector específico. O escopo industrial de economias de aglomeração é investigado no primeiro capítulo desta tese, e os principais resultados indicam que não há um único mix setorial local ótimo para fomentar a produtividade em diferentes setores tecnológicos. Além disso, setores de alta tecnologia e setores industriais de baixa tecnologia se beneficiam mais da escala urbana no Brasil, seguidos de setores de serviços associados a intensidade de conhecimento mais elevado.

As economias de aglomeração podem ter efeitos estáticos e dinâmicos. Eles são reforçados por um processo de seleção de trabalhadores qualificados para grandes áreas urbanas. As migrações inicial e de retorno constituem mecanismos essenciais para a auto-seleção de trabalhadores mais qualificados e mais produtivos para grandes áreas urbanas. Assim, cidades com maior percentual de trabalhadores mais habilitados deverão atrair mais indivíduos qualificados.

A estimação de economias de aglomeração estáticas indica que a inclusão do efeito fixo individual reduz o coeficiente da densidade de maneira significativa. Quando economias de aglomeração dinâmica são estimadas tendo por base a experiência prévia de trabalho em cidades, as vantagens estáticas se tornam não-significantes. Conforme esses anos de experiência são iterados com a densidade do local de trabalho atual, indivíduos trabalhando em cidades menos densas com experiência em cidades mais densas serão os maiores beneficiados. Por fim, a experiência prévia de trabalho tem um efeito positivo sobre o crescimento do salário somente no caso da experiência em cidades com ao menos a mesma densidade da cidade atual.

Finalmente, o tamanho da cidade tem um impacto importante sobre o poder de barganha relativo dos trabalhadores e das empresas no mercado de trabalho. Ao analisar a relação dos salários locais e do ciclo de negócios, a flexibilidade salarial, medida pela curva de salário, é maior em setores informais em áreas menos densas. Portanto, as grandes aglomerações supostamente oferecem maior poder de barganha dos trabalhadores, pois eles têm mais oportunidades de emprego.

Esses resultados indicam que as economias de aglomeração no Brasil parecem estimular a concentração espacial e ampliar as desigualdades regionais. Trabalhadores e firmas se auto-selecionam para grandes áreas urbanas, nas quais encontram um ambiente mais diversificado e outros trabalhadores altamente qualificados. Adicionalmente, grandes centros proporcionam maior poder de barganha aos trabalhadores em negociações salariais, mesmo que estejam no setor informal.

TABLE OF CONTENTS

INTRODUCTION	9
1. AGGLOMERATION ECONOMIES	15
1.1 Micro-foundations of agglomeration economies	16
1.2 Industrial scope of agglomeration economies	18
1.3 Knowledge externalities, proximity and new technologies	20
1.4 Static and dynamic agglomeration externalities	22
2. INDUSTRIAL SCOPE OF AGGLOMERATION ECONOMIES IN BRAZIL.....	25
2.1 Introduction.....	25
2.2 Urban wage premium and increasing returns	26
2.2.1 Principles of agglomeration economies.....	27
2.2.2 Empirical studies on agglomeration externalities.....	29
2.3 Analytical Framework	32
2.4 Data.....	35
2.5 Results.....	39
2.6 Conclusion	51
3. SHOULD I STAY OR SHOULD I GO? SELECTION ON MIGRATION AND LEARNING IN CITIES IN BRAZIL	53
3.1 Introduction.....	53
3.2 Literature review	54
3.2.1 Static versus dynamic externalities.....	55
3.2.2 Migration and selection bias.....	56
3.2.3 Learning in cities	58
3.3 Theoretical frameworks and empirical strategies	61
3.3.1 Selection in initial and return migration	61
3.3.2 Dynamic advantages from agglomeration economies.....	63
3.4 Data.....	67
3.5 Descriptive statistics	69
3.5.1 Migration decision.....	69
3.5.2 Dynamic agglomeration economies	72
3.6 Results.....	73
3.6.1 Migration	74
3.6.2 Dynamic agglomeration economies	81

3.7	Conclusion.....	86
4.	A COMPREHENSIVE ANALYSIS OF THE WAGE CURVE IN BRAZIL: NON-LINEARITIES, URBAN SIZE, AND THE SPATIAL DIMENSION	89
4.1	Introduction	89
4.2	The wage curve	90
4.3	Empirical literature.....	92
4.4	Labour market dualism in Brazil and a theoretical framework.....	96
4.4.1	Theoretical framework: institutional costs, bargaining power and efficiency wages	100
4.5	Empirical strategy.....	103
4.5.1	Extensions	104
4.6	Empirical analysis	108
4.6.1	The wage curve in a more aggregated context: a comparison between the formal and the informal sectors.....	109
4.6.2	Rural and urban areas and the wage curve.....	116
4.6.3	City size, spatial dependence, and the wage curve	124
4.7	Conclusion.....	135
5.	CONCLUSION.....	137
5.1	Specific results	139
5.2	Future research	141
	REFERENCES.....	143
	APPENDIX A1 – ADDITIONAL TABLES FOR THE INTRODUCTION	155
	APPENDIX A2 – ADDITIONAL TABLES FOR CHAPTER 2.....	156
	APPENDIX A3 – ADDITIONAL TABLES FOR CHAPTER 3	163
	APPENDIX A4 – ADDITIONAL TABLES FOR CHAPTER 4.....	164

LIST OF TABLES

Table 2.1 - The effect of agglomeration externalities on employment and productivity	28
Table 2.2 - Descriptive statistics of the main variables of the first stage of the model, 2012 (for the database containing individuals present in the whole period, 2004-2008-2012)	38
Table 2.3 - Descriptive statistics of the main variables for the second stage of the model for 2012 (database 2004-2008-2012)	40
Table 2.4 - Second-stage OLS regressions for the spatial wage, with and without individual fixed effects in the first stage	46
Table 2.5 - Second-stage regressions for the spatial wage, with and without individual fixed effects on the first stage, and with instrumental variables for $\ln(\text{density})$	49
Table 3.1 – Descriptive statistics of individuals according to their migration status over the period 1995–2008.	70
Table 3.2 – Descriptive statistics of the employment density at the origin and destination.	72
Table 3.3 – Descriptive statistics of the main variables for the regressions of dynamic agglomeration externalities.	73
Table 3.4 – Logit models for the probability of first-time migration, 1995 to 2007.	74
Table 3.5 – Logit models for the probability of first-time migration towards a REGIC area with a density higher than 10 workers/km ² and with an increase in density, 1995 to 2007.	77
Table 3.6 – Multinomial logit models for the determinants of second-time migration (for individuals who have already migrated once), 1996 to 2007.	79
Table 3.7 – Partial correlation of the \ln of hourly wages in $t-1$ and migration decisions in t , 1996 to 2007.	80
Table 3.8 – Estimation of the dynamic and static city-size earnings premia (dependent variable is the logarithm of the hourly wage), 2000 to 2008.	82
Table 3.9 – Regression for the urban wage growth premium (dependent variable is the individual wage growth in one year), 2001 to 2008.	84
Table 3.10 – Regression for the urban wage growth premium with previous experience and current workplace (dependent variable is the individual wage growth in one year), 2001 to 2008.	85
Table 4.1 – Average unemployment rates and hourly wages (both sectors) by metropolitan region, 2003-2015.	112
Table 4.2 – Descriptive statistics of the PME database.	112
Table 4.3 – Elasticity of wages in relation to the unemployment rate for different estimation strategies and sectors.	114

Table 4.4 – Elasticity of wages in relation to their time lags for different estimation strategies and sectors.	115
Table 4.5 – Descriptive statistics of the PNAD database.	119
Table 4.6 – Elasticity of individual wages in relation to the state-level unemployment rate, for the whole country, urban and rural areas, and for the whole labour force, formal and informal sectors, 1996 to 2013.	121
Table 4.7 – Elasticity of individual wages in relation to the state-level unemployment rate, for the whole country, urban and rural areas, and for the whole labour force, formal and informal sectors, 2002 to 2009.	122
Table 4.8 – Second-stage results of the IV estimation of the elasticity of individual wages in relation to the state-level unemployment rate with a Heckman selection model in the first stage.	123
Table 4.9 – Descriptive statistics of the variables in the first stage (individual level).	127
Table 4.10. Descriptive statistics of the second stage – spatial wage.	128
Table 4.11 – Regression of the spatial wage against local characteristics, OLS and IV, 2000 and 2010.	129
Table 4.12 – Regressions for the spatial wage with MCAs fixed effects.	130
Table 4.13 – Regression of the spatial wage against local characteristics, OLS, FE and SAR, 2000 and 2010.	133
Table 4.14 – Regression of the spatial wage against local characteristics and the iteration of density groups and the local unemployment rate, OLS, FE and SAR, 2000 and 2010.	134

LIST OF FIGURES

Figure 2.1 - Regional distribution of the mean logarithm of the hourly wage and the spatial wage (for OLS and FE in the first stage), by sector in 2012 (database 2004-2008-2012)	42
Figure 4.1 – Average hourly wage in the formal and the informal sectors and unemployment rate, 2003-2015.	111
Figure 4.2 – Coefficients and confidence intervals for OLS regressions, with the sample comprising individuals in 12-month periods ending in each month.	115
Figure 4.3 – Logarithm of the average hourly wage for the sample and logarithm of the unemployment rate at the state level, formal and informal sectors, 1996 and 2013.	120

LIST OF ACRONYMS

2SLS – Two-Stages Least Squares

CAPES – *Coordenação de Aperfeiçoamento Pessoal de Nível Superior* – Coordination for the Improvement of Higher Education Personnel

CBO – *Classificação Brasileira de Ocupações* – Brazilian Classification of Occupations

CLT – *Consolidação das Leis do Trabalho* – Labour Law Consolidation

CNAE – *Classificação Nacional de Atividades Econômicas* – National Classification of Economic Activities

CNPq – *Conselho Nacional de Desenvolvimento Científico e Tecnológico* – National Counsel of Technological and Scientific Development

F2F – Face-to-face

FE – Fixed Effects

FE-2SLS – Fixed Effects – Two-Stages Least Squares

FIPE – *Fundação Instituto de Pesquisas Econômicas* – The Institute of Economic Research Foundation

GDP – Gross Domestic Product

IBGE – *Instituto Brasileiro de Geografia e Estatística* – Brazilian Institute of Geography and Statistics

ILO – International Labour Organisation

INPC – *Índice Nacional de Preços ao Consumidor* – National Consumer Price Index

IPEADATA – *Instituto de Pesquisas Econômica Aplicada DATA* – Institute for Applied Economic Research DATA

IV – Instrumental Variables

LM – Lagrange Multiplier

MAR – Marshall-Arrow-Romer

MCA – Minimum Comparable Area

MTE – *Ministério do Trabalho e Emprego* – Ministry of Labour

NEG – New Economic Geography

NEREUS – *Núcleo de Economia Regional e Urbana da USP* – The University of São Paulo Regional and Urban Economics Lab

NUTS – Nomenclature of Territorial Units for Statistics

OLS – Ordinary Least Squares

PME – *Pesquisa Mensal de Emprego* – Monthly Employment Survey

PNAD – *Pesquisa Nacional por Amostra de Domicílios* – National Household Sample Survey

RAIS – *Relatório Anual de Informações Sociais* – Annual Report of Social Information

RAIS-MIGRA – *Relatório Anual de Informações Sociais – Migração* – Annual Report of Social Information - Migration

REGIC – *Regiões de Influência de Cidades* – Regions of Influence of Cities

SAR – Spatial Auto-Regressive

INTRODUCTION

The general pattern of growing cities around the world is a sign of their success as ways to organize economic production in space. Apart from a few exceptions (‘shrinking cities’, ‘cities in decline’), the global map shows an increasing urbanization, especially in the developing world. Clearly, cities may turn into diffuse urban agglomerations (including suburban urban areas, poly-nuclear satellites), but the trend is undeniably towards a rise in urbanized settlement patterns. It is noticeable that not only there is a trend towards mega-cities, but there are also many movements directed to smaller agglomerations (with less than 500,000 inhabitants), which may also act as relevant drivers of urban growth (UNFPA, 2007).

The process of accelerated urbanisation is characterized by migratory inflows and population growth, with relevant impacts on the competition for scarce local resources. Consequently, there may be agglomeration disadvantages, which are better expressed by the concept of rising congestion costs. They appear through the cost of land use, of housing, of transportation and of the labour force (VAN OORT; LAMBOOY, 2014). Pollution, crime and congestion are non-market public bads that can reduce the quality of life in big cities (KAHN, 2010).

Despite all these difficulties, the city has triumphed (GLAESER, 2011). In fact, the impracticability of big cities is debatable, as size is a relative concept to the available technology in a certain moment of time (JACOBS, 1969, p. 103–104):

A city that is large for its time is always an impractical settlement because size greatly intensifies whatever serious practical problems exist in an economy at a given time. [...] Cities magnify an economy’s practical problems, but they can also solve them by means of new technology.

In addition, the upward trend in urbanisation is associated with the fact that urban areas may be characterized as powerhouses with unprecedented agglomeration benefits of all kinds (see, for example, COMBES et al., 2012). Cities increase the proximity between economic agents and allow the exchange of information, goods and ideas at a lower cost. Therefore, returns to urban concentration measure the reduction of transportation costs (GLAESER; GOTTLIEB, 2009). The conceptualization and measurement of agglomeration (dis)advantages has prompted a wealth of urban-economic research (see, for example, DURANTON; PUGA, 2014; GLAESER, 1998; HENDERSON, 2010).

Urban success is usually inferred in the urban economics literature through the analysis of population growth, real estate prices and local wages (GLAESER, 2010; ROSENTHAL; STRANGE, 2004). The latter is most commonly used due to data availability and its linkage with local productivity. Therefore, agglomeration externalities are supposed to affect local labour markets, creating distinct conditions for knowledge creation and for the reduction of production costs.

In the developing world, the pace of change towards the concentration of people in cities has been even faster, with a more recent displacement of large population groups from rural to urban areas and the exponential growth in world population (KOURTIT; NIJKAMP, 2013). In Latin America, the rate of urbanisation is already reaching similar figures to the developed world, close to 80% in 2015 (UNITED NATIONS, 2015). Brazil is one of the main drivers of this process in the region, with a steady growth of its urbanization rate from 77.6% to 85.7% between 1995 and 2015.

The urbanisation process in Brazil generated a highly concentrated spatial distribution of the population. Consequently, the sixteen largest cities – with one million inhabitants or more – held 21.5% of the population in 2012, according to estimates from IBGE. The GDP per capita in these cities was more than two times higher than the GDP per capita observed in municipalities with population up to 50,000. This pronounced disparity is directly related to wage and productivity differentials according to city size.

Moreover, rural-urban migration and flows from lagging to leading regions are largely associated to the spatial concentration of the population observed in the past few decades. In this context, the population of the eleven biggest cities in 1970 grew at rates above the average until 1980, with a significant slowdown afterwards. On the other hand, the period from 1980 to 2010 was characterized by a raise in the population share of cities sized 100,000 to 1,000,000 from 23.5% to 33.7%. While this is an indication of a positive trade-off related to life quality in cities and of a tendency towards medium-sized urban centres, the main factors driving migration decisions may vary significantly according to the schooling or skill level of the individuals (LALL; TIMMINS; YU, 2009). The aforementioned slowdown in population growth for the biggest cities can be related to increasing life costs and congestion caused by a higher demand for public services, land, and transportation, among other factors.

Consequently, the combination of push factors with heterogeneous effects over different groups and high congestion levels in large cities seems to be driving migration flows in the recent period. Saturated cities in terms of the access to public services are attractive for individuals

who can profit the most from their advantages. At the same time, medium-sized cities may become the destination for less-skilled individuals, looking for a lower life-cost.

In this context, agglomeration economies may provide an explanation for the attracting power large cities still exert over a handful of individuals, with a clear positive selection of high-skilled workers, despite congestion. Workers and firms interact in the city, generating productivity gains not fully internalized by a few economic agents. These positive externalities may benefit workers and firms heterogeneously, according to their skill level and sector, among other characteristics. Consequently, city size will influence labour market relations, productivity levels, and wage-setting processes.

Thus, this dissertation intends to measure agglomeration economies in Brazil through the analysis of local labour markets. A special focus is given to the relationship between city size and workers' productivity in a certain moment of time. Moreover, over time, a learning process can lead to a boost in productivity, generating dynamic agglomeration externalities. Finally, the fact that firms and workers are located in a specific place with a relevant relocation cost implies that city size can affect their relative bargaining power over wage determination. Therefore, smaller places with a lower level of accessibility to job opportunities may result in a higher wage flexibility in economic downturns, as firms can easily adjust wages in real terms in case there is a rise in the local unemployment rate.

Based on these topics, the following questions are addressed:

A. Industrial scope of agglomeration economies in Brazil:

- a) What is the size of urban agglomeration economies in Brazil?
- b) To what extent does the sorting of workers affect the measurement of static agglomeration externalities?
- c) What is the relative importance of urbanisation and localisation economies and local competition to foster productivity in different sectors of the Brazilian economy?

B. Selection of workers and dynamic agglomeration economies:

- a) Labour supply
 - i. How does the sorting process of highly skilled workers into large cities happen in Brazil?
 - ii. Which characteristics indicate that a worker will be more prone to migrate and decide to stay in the destination?

- iii. Which are the main implications of this sorting process to the spatial distribution of different types of workers?
- iv. What is the relevance of the sorting of workers to the urban wage premium?
- b) Labour demand
 - i. What is the effect of the inclusion of dynamic gains over the static agglomerations estimator?
 - ii. Who benefits the most from dynamic agglomeration gains?
- C. Wage flexibility in local labour markets – the wage curve:
 - a) What is the difference between the formal and the informal sectors in relation to wage flexibility?
 - b) How does market size affect the elasticity of local unemployment in relation to wages?
 - c) How relevant are neighbourhood effects in the context of the wage curve in Brazil?

The answers to these questions not only contribute to the debate of the importance of agglomeration economies in Brazil, but also involve considering the best theoretical frameworks for each case and the main strategies to avoid data limitations, which are common in developing economies. The conclusions obtained are expected to provide a better understanding of the functioning of local labour markets in Brazil, with a special focus on how city size can affect labour productivity.

An additional contribution of this work involves the investigation of how local wage determination may be affected by city size. As the probability of obtaining a matching in the labour market is higher in large urban areas, the wage bargaining power of workers and firms can be influenced by local conditions. In this context, the estimation of the wage curve provides an insight on the way agglomeration economies may affect local labour markets.

In addition to multiple methods of analysis, the study of agglomeration economies in Brazil developed here is based on many databases, briefly summarised in Table A.1 of the Appendix. Tables, figures and equations are identified with two digits, the first referring to the chapter and the second to the ordination along the chapter. In the Appendix, a letter A precedes the numbering of tables.

The remaining of this dissertation is organised as follows. Chapter 1 reviews the literature on agglomeration economies, identifying their main micro-foundations, and distinguishing between static and dynamic agglomeration externalities. Even though the empirical literature has been extensively developed in the last decades, there are some limitations in the main

theoretical approaches underlying agglomeration economies. There is not yet a comprehensive framework to generate empirical predictions to be tested (COMBES; GOBILLON, 2015).

Then, Chapter 2 investigates the industrial scope of agglomeration economies in Brazil, with a special focus on the combination of diversity, specialisation and competition in the generation of local advantages for different sectors. In a sense, this framework provides an indirect evaluation of the way agglomeration economies may affect location decision of firms (demand of workers).

Based on the worker's perspective, Chapter 3 explores the main factors that induce the sorting of highly skilled individuals into larger cities through selection in initial and return migration. After that, static and dynamic agglomeration advantages are estimated for different types of workers, aiming to understand who benefits the most from large cities. Therefore, the main factors underlying the location decision of workers are explored in this empirical exercise (supply of workers), among which there are agglomeration economies.

Chapter 4 explores the relationship between the wage curve and urban size in Brazil. A special focus is given to the differences in wage determination in the formal and informal sectors. To achieve this, an analytical framework that embodies different levels of bargaining power for workers in each sector is proposed to generate the wage curve in the context of a developing economy. City size is supposed to enhance the likelihood of a matching in the labour market, raising the bargaining power of workers and decreasing wage flexibility. In this context, agglomeration economies are supposed to affect the equilibrium in local labour markets that results from the combination of the supply and demand of workers.

Finally, Chapter 5 concludes, summarizing the main findings of the previous chapters and providing some guidance for future research on agglomeration economies, with a special focus on developing countries. In fact, apart from estimating the urban wage premium, the main consequences of these agglomeration advantages over economic relations established in urban areas should be understood. Furthermore, this research area still lacks a well-defined theoretical framework, covering all relevant dimensions of the supply and demand sides of local labour markets.

1. AGGLOMERATION ECONOMIES¹

The determination of wages at the regional level is a combination of two mechanisms that may have distinct effects. On the one hand, there is the concept of spatial equilibrium, in which, under free mobility of individuals, wages are supposed to equalize the utility obtained in all regions. Therefore, people should be indifferent between distinct regions, and high wages would be accompanied by low amenities and/or high prices.

The other mechanism is related to the concept of labour market equilibrium, through which local labour demand and supply produce an equilibrium wage in a certain region. This wage may reflect some positive local characteristic, which could be the reason why firms would prefer to remain in that place and accept a higher labour cost. If wages reflect the marginal productivity of labour (for instance, under a standard Cobb-Douglas production function), they will be a signal of the local productivity level (GLAESER, 2010). Consequently, local labour markets outcomes should be understood as a result of the inclusion of the spatial dimension in the labour market.

The spatial distribution of economic activities determines the existence of cities and trade flows between them. However, the importance of the spatial dimension is not identified by the competitive paradigm of the neoclassical economic theory (exemplified by ARROW; DEBREU, 1954). In fact, these type of models could not properly include space and increasing returns as additional dimensions of economic modelling (FUJITA; THISSE, 2012).

In more detail, even if spatial interdependencies are considered in the neoclassical model as additional characteristics of the choice of place to work and consume, space is still taken as homogeneous (it does not affect the utility of consumers and the production set of firms, and consumers do not have any a priori preference for a specific location). As shown by Starrett (1978), under a homogeneous space with a finite number of locations, transportation frictions and locally non-satiated preferences, there is no competitive equilibrium with goods transported between locations (the ‘spatial impossibility theorem’). In these models, trade is substituted for the mobility of production factors and all local consumption is satisfied by local production.

¹ Part of this chapter will be published in: BARUFI, A.M.B.; KOURTIT, K. Agglomeration economies and smart cities. In: KOURTIT, K.; NIJKAMP, P.; STOUGH, R. (Eds.) *The Rise of the City*, Edward Elgar, Cheltenham, pp. 33-54, 2015 (forthcoming).

Relaxing the assumptions of the neoclassical model, one of the possibilities to help understand the existence of cities and their trade flows is to allow for spatial heterogeneity. Nevertheless, this factor has less variation within countries, making it more appropriate for modern urban economic theory to consider increasing returns to scale as an explanation for the existence of cities (BRAKMAN; GARRETSEN; VAN MARREWIJK, 2009). In addition, such agglomeration forces are generated by non-market interactions between economic agents. Another possibility is that under imperfect competition, firms are price-makers (by product differentiation or oligopolistic strategies), meaning that their spatial distribution, along with consumers, will determine regional prices and may generate spatial agglomerations (COMBES; MAYER; THISSE, 2008b).

The focus here will be on the latter mechanism (based on urban-economic theories), as interest lies in the urban level of analysis. In this context, agglomeration economies are defined as external economies of scale generated by the spatial concentration of economic agents. Therefore, individuals and firms may benefit from being concentrated in cities by obtaining higher wages and higher productivity levels, respectively. In addition, space becomes an essential dimension in the generation of equilibrium outcomes in local labour markets.

The remaining of this chapter aims to discuss relevant topics for the empirical exercises provided in Chapters 2, 3 and 4. The next section will discuss the micro-foundations of agglomeration economies. Then, Section 1.2 will present concurrent theoretical perspectives for the industrial scope of agglomeration economies, while Section 1.3 will explore the concept of knowledge externalities and how new technologies can change their formation. Finally, Section 1.4 will provide an overview of the main differences between static and dynamic agglomeration externalities.

1.1 Micro-foundations of agglomeration economies

The concept of agglomeration economies is strongly tied to the idea of proximity, as they result from the “interaction (often involuntary) among economic agents made possible by the lesser amount of spatial friction that occurs in concentrated locations” (CAPELLO, 2014, p. 163). They are of particular interest due to their positive effect on the productivity of firms and the utility of individuals. Among other advantages, they present a hedge against uncertainty in economic relations (GORDON; MCCANN, 2000). Models of economic growth have been focusing mostly on the gains generated by knowledge externalities and innovations, which are

classified as dynamic externalities, as they increase the innovative capacity of firms and aim to explain not only how cities form, but also why they grow (GLAESER et al., 1992; PESSOA, 2014).

It is noteworthy that static externalities are associated with an increase in the efficiency of technologies that are already in use, explaining the existing distribution of economic activity, productivity and amenities, but not their growth. Among these, there are urbanization and localization externalities, which refer to the co-localization with firms from different sectors or with firms of the same sector, respectively. In the case of urbanization economies, an increase in urban scale or population density may allow firms to increase their production without improving productivity or efficiency (DE GROOT; POOT; SMIT, 2009).

On the other hand, spatial equilibrium models applied to an urban context are usually concerned with the effect dynamic externalities may have over the economic activity distribution and the interactions between economic agents. Glaeser and Gottlieb (2009) develop a basic framework to understand the effects of agglomerations on productivity and income, which should encompass the production side, individual utilities and the housing market (the Rosen-Roback tradition). Departing from a basic production function with an explicit term related to the productivity level, the labour demand function derived from the optimization process of the firm will present a positive relationship between wages and productivity. In a dynamic setting, the model includes urban changes and an adjustment of the local economy caused by migration and changes in the housing market. In this setting, agglomeration economies are multipliers of the exogenous factors that affect productivity, population, wages and prices, amplifying the relationship between amenities, housing supply and economic agents.

The measurement of agglomeration economies requires a better understanding of the mechanisms that generate them. In this context, there has been a large discussion in the literature on the micro-foundations of agglomeration externalities. Marshall (1890) identifies three main channels through which proximity can lead to extraordinary gains: it makes possible the exchange of information and knowledge; there are potential benefits from labour market pooling; and firms may share inputs, specialized or not. These interactions are stronger between firms belonging to the same sector and they happen outside traditional market mechanisms, without prices.

Duranton and Puga (2004) and Puga (2010) review these mechanisms and present some more specific concepts driving the productive interaction between economic agents: (i) sharing

(facilities, gains from individual specialization and variety, labour pooling and risk); (ii) matching (higher matching quality and quantity among workers and firms); and (iii) learning (knowledge generation, diffusion and accumulation). Complementarily, Rosenthal and Strange (2004) consider additional elements that generate agglomeration economies, such as local natural advantages, home market effects, consumption externalities, and rent seeking.

These mechanisms can be embodied in a broader perspective of transportation costs. Glaeser and Gottlieb (2009) argue that proximity reduces transportation costs of goods (supply and demand of intermediate goods are closer by), people (labour markets are more efficient in urban areas and service providers have greater access to their clients), and ideas (stronger knowledge exchange between people and firms).

It is worth mentioning that there may be agglomeration diseconomies in bigger cities associated with pollution, congestion and higher costs (of land, labour force and transportation), which can lead to negative externalities for firms and individuals (GLAESER; MARÉ, 2001). Moreover, the cost-benefit functions faced by different economic actors in large urban areas are expected to present a certain degree of heterogeneity, determining distinct locational choices (FUJITA; THISSE, 2012). Consequently, there will be push and pull factors related to agglomeration (dis)economies shaping the spatial distribution of firms and workers.

1.2 Industrial scope of agglomeration economies

The previous section emphasized the main mechanisms that generate agglomeration economies. Firms in a city may find the economic environment more or less suitable for fruitful interactions with other firms. Their location decision can be studied with analytical elements provided by the literature related to the industrial scope of agglomeration economies. In general terms, different sectors may benefit or not from local competition, specialisation and/or diversity.

The exchange of knowledge and ideas is a common element of the different mechanisms underlying agglomeration externalities, being related to the productivity shifts discussed above through the creation and adoption of new products and processes. This concept is central to studies investigating the industrial scope of agglomeration economies (Rosenthal and Strange, 2004), especially for the comparison between specialized, diversified and competitive urban structures. In this context, the MAR model² advocates that industrial specialization fosters the

² Marshall-Arrow-Romer, referring to the Marshallian externalities discussed above, combined with more recent advances in economic growth theory.

generation of knowledge spillovers, as the similarities between firms facilitate their interactions through the circulation of workers, imitation and business relations (GLAESER et al., 1992). Furthermore, a certain monopolistic power can be more suitable for innovation, as it increases the capacity of firms to internalize knowledge externalities (CAPELLO, 2014).

Alternatively, according to Jacobs (1969), the source of knowledge externalities is external to the industry, being related to industrial diversity. A higher variety of industries can generate more opportunities to create, share and combine ideas and practices, as well as to exchange skills. In this case, competition is viewed as an incentive for firms to innovate and to adopt new ideas, and usually industries that share a common science base can profit further from a diverse environment.

The third main theoretical perspective concerning the generation of knowledge externalities is based on Porter (1990). In his view, a competitive environment fosters economic growth, as it gives incentives for firms to innovate and to promote technical progress. This process is stronger in a specialized economy, as knowledge spillovers occur mainly within vertically integrated industries.

These different perspectives can be harmonized by considering that the maturity and the technological content of a certain industry can affect the way it benefits from the different types of agglomeration economies. Standardized manufacturing with well-established firms may derive higher benefits from the agglomeration of their own industry (specialisation), while innovative and high-technology industries will tend to locate in larger and more diverse urban areas (BEAUDRY; SCHIFFAUEROVA, 2009; HENDERSON, 2010).

With all these considerations in mind, it is important not only to understand how productivity is affected by city size, but also to what extent the industrial composition of an urban area may be more or less beneficial to firms in specific industries. In other words, agglomeration economies can be studied simultaneously by their direct effect on productivity and by their industrial scope. Nonetheless, the current empirical literature is still facing difficulties in finding consistent results, requiring a careful re-evaluation of concepts and methodologies (VAN OORT; LAMBOOY, 2014).

An additional aspect to be mentioned is that geographical proximity is a key concept for the generation of agglomeration economies (CAPELLO, 2014). However, there are different dimensions through which individuals and firms can interact, especially in relation to dynamic externalities and knowledge exchange. Section 1.3 will present a contribution to this issue and

consider how new technologies can change the role of space for the generation of these externalities.

1.3 Knowledge externalities, proximity and new technologies

Previously, Section 1.1 provided an initial discussion on the importance of knowledge creation and exchange to increase local productivity. In spite of the clear relevance these elements have for local competitive advantages (BEAUDRY; SCHIFFAUEROVA, 2009), there are a few issues regarding knowledge externalities that require further investigation. Among these, it is possible to highlight the reach of spatial spillovers (KRUGMAN, 2009) and the relative importance of different types of proximity (BOSCHMA, 2005). Therefore, it is not clear whether knowledge externalities really depend upon geographical proximity or if technological changes have already provided a solution to overcome this obstacle.

Firstly, there is a distinction to be made between different types of knowledge, namely tacit and codified knowledge. While the former refers to a person-specific knowledge, requiring direct contact between individuals to be transmitted, the latter can be transacted in markets and easily transferred between individuals (VAN OORT; LAMBOOY, 2014). Furthermore, tacit knowledge is acquired through learning-by-doing or learning-by-using, and its transmission requires proximity in different dimensions, such as spatial, cultural and social (BOSCHMA, 2005).

In this context, face-to-face (F2F) contact is usually associated with the transmission of tacit knowledge and is related to the different mechanisms of generation of knowledge externalities. The main advantages associated with F2F contact involve allowing the communication between different agents even with non-codifiable information, reducing the problems with incentives and lack of trust, the evaluation of potential partners and socialization, and improving performance by giving incentives for greater effort and motivating individuals (STORPER; VENABLES, 2004). Therefore, the spatial dimension has a crucial role in the exchange of knowledge, especially in its tacit version. The exchange of codified knowledge may also require spatial proximity, because the interpretation and assimilation of this kind of information can depend on tacit knowledge (HOWELLS, 2002).

The literature on spatial spillovers and agglomeration economies has focused on space as the main dimension to investigate the interaction between firms and workers. In recent years, however, there has been a shift in the perception of the types of proximity that are essential to

the promotion of knowledge spillovers and innovation. For instance, the endogenous growth literature highlights that innovations are determined by increasing returns, which are associated with spatial, relational and institutional proximity (CAPELLO; NIJKAMP, 2009). In this sense, there are different factors in motion to promote innovation, with proximity being considered a transitional concept between socio-economic and spatial interactions (STIMSON, 2014).

Therefore, proximity is no longer exclusively a geographical concept, being accompanied by other relevant dimensions (cognitive, organizational, social and institutional). In this sense, there is a high degree of complementarity between these different relevant dimensions, but even if geographical proximity can be substituted by a combination of other types of proximity in the learning process (BOSCHMA, 2005; TORRE; RALLET, 2005), there is still a great deal of knowledge exchange associated with space (CAMAGNI; CAPELLO, 2009). Combining different perspectives seems to be the most interesting way to proceed in order to model and estimate innovation and knowledge generation (CAPELLO, 2014).

Another relevant discussion in this subject relates to the role of geography in innovation systems under the rise of information and communication technologies (ICTs), that have emerged in the past few decades, accompanied by an increase in the globalization of economic activity (for example, CAIRNCROSS, 2001). These transformations are said to be changing the requirement of proximity in space by altering the way knowledge and information are produced and exchanged, based on the concepts of tradability and codification. The codification of information and meanings allows the transformation of complex and unfamiliar coordination tasks into routine activities, which can be placed in cheaper and remote locations. This reinforces the tendency towards de-agglomeration of activities with codified knowledge.

On the other hand, new activities are created through innovation, and usually the tasks related to them are complex and unfamiliar, requiring F2F contact and geographical proximity in general terms. Tacit knowledge will permeate these new activities, meaning that the knowledge required to conduct them is usually embodied in high-skilled workers (STORPER; VENABLES, 2004). Moreover, the quality of F2F interactions is higher than that of electronic ones, and they intersect in a complementary manner to lead to knowledge and information creation and diffusion. In this sense, digital technologies are usually able to maintain but not to create communities, while information may be easily diffused, but not necessarily understood (MORGAN, 2004).

Therefore, the production and exchange of knowledge and the process of innovation still rely significantly on spatial proximity. Technology has created new ways of promoting the interaction of individuals in a virtual manner, but the existence of tacit knowledge and trust issues make physical contact still important, at least for some of the interactions. Space, then, still seems to be relevant in the context of knowledge generation.

1.4 Static and dynamic agglomeration externalities

In addition to proximity, agglomeration economies also have a time dimension. Therefore, they can be classified into static and dynamic advantages, being ultimately defined in space and time. The previous sections highlighted the importance of space to produce agglomeration externalities in general, while time is key to define whether these externalities are static or dynamic.

In this context, static agglomeration economies are associated to an increase in productivity due to urban size in a specific moment of time, affecting the observed distribution of economic activity and the efficiency of already existing technologies (DE GROOT; POOT; SMIT, 2009). Furthermore, geography, infrastructure and population size are supposed to generate these static agglomeration advantages. They are enjoyed while workers are living in a big cities but are lost upon moving away (DE LA ROCA; PUGA; 2014).

However, agglomerations may have long-lasting effects, affecting the trajectory of productivity growth for each city and economic agent over time (COMBES; GOBILLON, 2015; CAMAGNI; CAPELLO; CARAGLIU, 2015). In this context, dynamic agglomeration economies are related to interactions between economic agents that change the path of growth of their productive capacity.

In general, knowledge externalities are supposed to generate these special conditions for productivity growth, being directly associated to dynamic agglomeration externalities. They help explain how cities change and generate location advantages over time for firms and workers, with a special focus on specialisation, competition and diversity at the local level. On the other hand, urbanisation and localisation economies refer to once-and-for-all static advantages (DE GROOT; POOT; SMIT, 2009).

Furthermore, dynamic advantages and knowledge externalities are also related to experimentation and learning in cities. They can be transformed into higher human capital, benefiting a worker even if he or she relocates (DE LA ROCA; PUGA, 2014).

The remaining of this dissertation will explore different aspects of agglomeration economies in Brazil. Chapter 2 presents an analysis of dynamic gains with a special focus on their industrial scope, in which specialisation, competition and diversity at the industry level generate these advantages. Then, Chapter 3 provides a comparison of static and dynamic advantages for workers in cities with different sizes. Finally, in Chapter 4, city size is supposed to generate unique local conditions for different levels of bargaining power of workers and firms, affecting local wage determination.

As mentioned at the beginning of this chapter, two concepts will be essential to understand the outcomes analysed in the empirical studies: spatial equilibrium and labour market equilibrium. In fact, they are completely interrelated and cannot be dissociated, as cities may compete among each other to attract workers and firms, aiming to provide the best environment for flourishing a productive labour market.

2. INDUSTRIAL SCOPE OF AGGLOMERATION ECONOMIES IN BRAZIL

2.1 Introduction

Brazil has recently shown an important decrease in personal and regional inequality. Nevertheless, its income inequality is still one of the highest worldwide (it has the 16th highest Gini index among 141 countries³), reaching 0.523 in 2013 down from 0.603 in 1995. Regional income disparities are also considerably high, demonstrated by the fact that average personal income was 1.78 times higher in the South-South-east than in the North-North-east in 2013.

Following the tendency towards urbanisation observed all over the world (GLAESER, 2011; UNFPA, 2007), there is a significant spatial concentration of population and economic activity in Brazil. In 2010, 84.4 per cent of the population lived in urban areas, occupying 1.07 per cent of the national territory⁴. Regarding economic concentration, in 2012, cities with more than 500,000 inhabitants held a 40.8 per cent share of total GDP, while housing 29.3 per cent of the total population⁵. Also, the correlation between the logarithm of wages and the logarithm of population density at the municipal level reached 0.06 in 2010 for the whole labour market, and 0.05 for the formal sector.

Moreover, as mentioned in the previous chapter, it is widely acknowledged that living costs are significantly higher in bigger cities, meaning that only with a positive relationship between wages and the city size is it possible to explain the existence of cities (DURANTON; PUGA, 2014; VAN OORT; LAMBOOY, 2014). In this context, the proximity of economic agents generates agglomeration economies by reducing transportation costs. Different theoretical approaches analyse these costs reductions under different perspectives – of goods, people, and ideas (GLAESER; GOTTLIEB, 2009).

Following the substantial empirical literature on agglomeration economies, this chapter aims to understand the nature and magnitude of regional disparities in Brazil by exploring the industrial scope of the extraordinary gains obtained in big cities. The strategy followed here presupposes that salaries are directly related to productivity, and investigates whether competition,

³ Considering the Brazilian Gini Index of 2012 and the most recent information for other countries, available at <https://www.cia.gov/library/publications/the-world-factbook/rankorder/2172rank.html>, accessed on 27/06/2014.

⁴ Based on data from the Demographic Census of 2010, IBGE (2013).

⁵ With information from the Municipal GDP, provided by IBGE.

specialisation, and/or diversity are more or less relevant to explain the regional residual variation in salaries in different sectors (which is not related to individual characteristics).

In comparison with the previous literature, different estimation strategies are provided here, controlling or not for individual unobserved characteristics constant in time (COMBES; DURANTON; GOBILLON, 2008a; COMBES et al., 2010; GLAESER and MARÉ, 2001; GROOT; DE GROOT; SMIT, 2014a). Moreover, a two-stage estimation on two different levels is considered (the first stage is at the individual level, while the second stage is at the sector-region-year level).

Concerning the empirical literature on agglomeration economies in Brazil, an analysis at the micro-level is provided, controlling for individual characteristics. In addition, the analysis presented here compares the potential effects of specific sources of externalities related to the industrial scope of agglomeration economies (diversity, specialisation, and competition) over sectors with different levels of technology and knowledge intensity. Finally, it is also possible to discuss how the formalisation of the Brazilian labour market may have affected all these relationships by analysing the formal sector in two different separate periods.

This chapter aims to contribute to the literature of static and dynamic agglomeration effects, by controlling for individual skills and comparing different estimation strategies. The focus is on the industrial scope of agglomeration economies, a strand of the literature that has recently attracted more attention because individual-level longitudinal data have become more easily available. Moreover, so far most empirical studies have focused on developed countries, but here instead an evaluation of the relative importance of the local industrial mix for different sectors in the context of a developing economy is provided (that of Brazil), which is another important contribution.

The next sections are organised as follows: Section 2.2 contains a review of the related literature; Section 2.3 discusses the methodological approach; Section 2.4 describes the data; and then, Section 2.5 presents the results. Finally, Section 2.6 draws the main conclusions from the analysis.

2.2 Urban wage premium and increasing returns

The relationship between city size and the returns of firms and workers has been examined based on the assumption that urbanisation reflects the gains from agglomeration. The next

subsections will briefly review the main issues investigated in the literature, and explore the advances of the empirical studies in this area.

2.2.1 Principles of agglomeration economies

As discussed in Chapter 1, the urban wage premium has been the subject of analysis of several studies aiming to identify how the density of economic activity affects the productivity of workers (HEUERMAN; HALFDANARSON; SUEDEKUM, 2010). There is usually a positive association between city size and salaries (Combes et al., 2010). However, cities present a trade-off between costs and benefits, which are directly related to dispersion and agglomeration forces (COMBES; DURANTON; GOBILLON, 2011; DURANTON; PUGA, 2014; GLAESER, 1998). The challenge is to isolate these effects from other explanatory factors of productivity differentials in space, while investigating the possibility of convergence and divergence between regions, sectors, educational levels, among other dimensions (LINDLEY; MACHIN, 2014).

Spatial wage differentials can be explained by three main sets of variables (COMBES; DURANTON; GOBILLON, 2008a): the skill composition of the local labour force; geographical characteristics and local factors of production; and interactions between firms or workers. The latter set is related to the generation of agglomeration economies.

In fact, under the conditions defined by the Spatial Impossibility Theorem (STARRET, 1978), increasing returns constitute one of the possible explanations for the existence of cities with trade flows between them. When there are positive externalities to the urban size, the competitive paradigm in space is no longer valid (COMBES; MAYER; THISSE, 2008b), thus creating incentives for the agglomeration of economic agents (FUJITA; THISSE, 2012; GLAESER, 1998).

Agglomeration economies are generated by the interactions between firms and individuals. Synthesising different contributions, Duranton and Puga (2004) and Puga (2010) identify three main micro-foundations: (i) the **sharing** of facilities and gains from individual specialisation and variety, and risk sharing, through labour pooling; (ii) the higher probability of a **matching**, with higher quality; and (iii) **learning**, related to knowledge generation, diffusion, and accumulation. These externalities are static when they explain the cross-sectional distribution of economic activity, and dynamic when related to productivity growth and knowledge spillovers (DE GROOT; POOT; SMIT, 2009; PESSOA, 2014). In fact, as mentioned in the

previous chapter, dynamic externalities are generated by the profitable interaction of workers and firms, which provide them the elements to increase their productivity in a long-lasting manner. According to Rosenthal and Strange (2004), they may be related to the city size (urbanisation economies, associated with Jacobian economies) or to the size of the particular economic sector (localisation economies, linked to Marshallian economies).

The micro-foundations discussed above come from a reinterpretation of Marshall (1890), who pointed out that proximity gains could be stronger between firms and workers within the same industry (labour market pooling, input sharing and knowledge exchange). Glaeser et al. (1992) summarised further developments in this direction with the MAR model (Marshall-Arrow-Romer), in which industrial specialisation could promote knowledge spillovers by favouring the interaction between firms with a similar structure. Moreover, it may be easier for firms to internalise innovation gains when they have a higher local monopolistic power (CAPELLO, 2014). Another perspective is brought by Jacobs (1969), according to whom knowledge externalities are fostered by the interaction between economic actors of different sectors. Therefore, the source of agglomeration economies is external to the industry, allowing the exchange of complementary knowledge, thus generating urbanisation economies. In this setting, competition is seen as an incentive for firms to innovate. Porter (1990) also defends this last aspect, but his theoretical framework advocates that a higher level of specialisation reinforces those gains (knowledge exchange is stronger in vertically-integrated industries).

Depending on the development stage of each industry in a certain country, they will benefit from different industrial structures in their locality. For instance, new sectors with a strong focus on innovation efforts may be located in larger and diversified urban areas where they can profit from a wide range of experiences and ideas (HENDERSON, 2010). Following De Groot, Poot and Smit (2009), it is possible to organise the sources of externalities and their expected effects over employment and productivity according to each theoretical view discussed above (see Table 2.1):

Table 2.1 - The effect of agglomeration externalities on employment and productivity.

Type		Measure	Effect on employment growth and productivity		
Static	Localisation externality	Geography, infrastructure	+		
	Urbanisation externality	Demand, population size	+		
			MAR (Marshall-Arrow-Romer)		
Dynamic	Knowledge externalities	Specialisation	+	-	+
		Competition	-	+	+
		Diversity	-	+	-

Source: De Groot, Poot and Smit. (2009).

This strand of the literature was initially synthesised during the 1960s and the 1970s (CAPELLO, 2014). However, more recently there has been a new wave of theoretical and empirical studies covering the relationship of productivity, geographical proximity and local industrial composition. From the perspective of the empirical literature, detailed databases based on micro-data at the individual level have allowed the brand new theoretical models to be tested by controlling for individual heterogeneity in order to assess regional-level relationships (ROSENTHAL; STRANGE, 2004). In the Brazilian context, the empirical literature using this recent approach is still very limited, as is discussed in the next subsection.

Therefore, the industrial scope of agglomeration economies can be further investigated by controlling for individual characteristics and other confounding factors. This strategy allows the identification of a proper local sector mix, the optimum city size, and an adequate degree of competition for different industries. This framework is now considered in the case of Brazil, a developing economy with a continental dimension and huge regional disparities.

2.2.2 Empirical studies on agglomeration externalities

The literature that relates productivity and different measures of agglomeration externalities has been reviewed elsewhere (BEAUDRY; SCHIFFAUEROVA, 2009; MELO; GRAHAM; NOLAND, 2009). In general, results are significantly heterogeneous, and one of the explanations for this is that the outcome related to productivity can vary. It can be either directly derived from the production function or indirectly derived (from salaries, rents, job creation, new establishments, among others), according to data availability. Rosenthal and Strange (2004) highlight the main difficulties embodied in the direct estimation of a productivity measure (capital and land may be hard to measure). On the other hand, rents are not observed for most of the countries in their whole extent.

Finally, salaries are easily observed at the individual level, with a significant periodicity. The main underlying assumption is that they equalise the marginal productivity of labour (under perfect competition), or that at least they are higher in places that are more productive (ROSENTHAL; STRANGE, 2004). An important drawback of this strategy is that elasticities in the labour market will determine the extent to which wages capture local productivity (MORETTI, 2011).

Apart from that, the heterogeneous results found in the literature can be associated with different measures for competition, diversity and specialisation (BEAUDRY;

SCHIFFAUEROVA, 2009). Melo, Graham and Noland (2009) also remark that the estimates of the elasticity of urbanisation economies vary among countries, data structure, geographical aggregation, and economic sector, in the range from 3 to 8 per cent (an increase of 1 per cent in urban density, for instance, is expected to increase wages from 3 to 8 per cent).

The theoretical framework most commonly applied to this sort of analysis is based on the wage equation of the urban economics literature (COMBES; MAYER; THISSE., 2008b). Glaeser and Maré (2001) conducted one of the first studies aimed at measuring the urban wage premium, while controlling for individual heterogeneity. They consider longitudinal individual data for the United States, and find evidence that there is an urban premium even when observed and unobserved individual characteristics are taken into account. In a similar setting, Mion and Naticchioni (2009) show that individual abilities and firm size explain a large share of the spatial heterogeneity of wages in Italy.

Combes, Duranton and Gobillon (2008a) find that controlling for the sorting of individuals (with individual fixed effects) reduces the estimated coefficient of urbanisation externalities by 40 to 50 per cent. However, apart from creating a black box for the sorting process, the inclusion of individual fixed effects may introduce a selection bias, as only those individuals who have migrated will provide the variation to estimate the agglomeration coefficient. Furthermore, pooled cross-sections will provide an upper bound for the estimates, while the panel estimation will generate a lower bound (GROOT; DE GROOT; SMIT, 2014a).

There is a strong concern for the potential endogeneity of urban size measures (endogenous quantity of labour), such as population density (COMBES et al., 2011). This is so because cities with higher wages may attract more individuals, leading to an increase in population and consequently, in city size. The approach proposed by Ciccone and Hall (1996) and adopted by Combes, Duranton and Gobillon. (2008a), Groot, De Groot and Smit. (2014a), and Graham et al. (2010), among others, is based on long time lags of the endogenous variable, while Combes et al. (2010) also suggest the inclusion of geographical characteristics as instruments. On the other hand, the individual fixed effects mentioned above aim to control for the endogenous quality of labour, namely, the sorting process of the labour force (COMBES; DURANTON; GOBILLON, 2008a; GLAESER; MARÉ, 2001).

As previously suggested, the industrial scope of agglomeration externalities can be evaluated in order to assess which theoretical framework seems more suitable to explain regional wage disparities. Glaeser et al. (1992) explore city-level data in the United States, finding results consistent with the framework presented by Jacobs (1969) – employment growth is encouraged

by urban diversity and local competition. Groot, De Groot and Smit (2014a) estimate a two-stage regression, explaining the spatial residual of a mincerian regression. The authors find an employment density elasticity of 4.8 per cent for NUTS-3 areas in the Netherlands, with a stronger indication of MAR externalities and small negative effects from competition and diversity measures.

Longitudinal data on the labour market in Brazil has become available only recently, and access is still very limited. Therefore, a large part of the empirical literature is based on individual-level cross-sections, or aggregated data at the area level. In an attempt to measure agglomeration externalities for the country, Henderson's (1986) estimates indicate that localisation economies seemed more important than urbanisation economies in the 1970s, meaning that manufacturing activities might be less productive in larger urban areas. Moreover, the mean of the coefficients estimated for different sectors is 0.046, ranging from 0.003 to 0.18.

Amaral et al. (2010) estimate a New Economic Geography (NEG) wage equation, and find a positive relationship between wages and market potential. Also based on an NEG model, Fally, Paillacar and Terra (2010) find that market potential and supply access (intermediary goods) are positively related to individual wages. Silva and Silveira-Neto (2009) explore the determinants of manufacturing employment growth at the state-sector level in Brazil between 1994 and 2004. Among the explanatory variables, there are: the average wage; average firm size; connections (a concentration index based on backward and forward linkages); a proxy for transportation costs; and dynamic externalities (specialisation and diversity). They find an indication that competition is relevant for employment growth (Jacobs's and Porter's theoretical arguments), as well as diversity (Jacobs).

Hierarchical wage equations are applied by Fontes, Simões and Oliveira (2010) to control for variables at different levels (individual and territorial unit of analysis). Based on Census data for 1991 and 2000 (cross-sections), the authors evaluate municipalities with at least 50,000 inhabitants, and find significant regional disparities even after controlling for observed individual heterogeneity. There is evidence of a positive effect of urban scale on wages, as well as of a positive relationship between the industrial concentration and the density of modern productive services with wage levels.

With an urban economics wage equation, Simões and Freitas (2014) apply municipal data to find that urbanisation economies are more relevant for sectors with high technological intensity, while sectors with low and medium technological intensity benefit more from medium sized

urban centres, which are relatively less diversified. Considering data from RAIS (Annual Report of Social Information, from the Ministry of Labour), Freguglia and Menezes-Filho (2012) find that, when controlling for individual heterogeneity, almost 63 per cent of the total spatial differential disappears, meaning that local policies should focus on human capital development in order to promote regional development.

The investigation of the industrial scope of agglomeration economies, while controlling for individual heterogeneity, has become the subject of analysis of recent studies, owing to an increase in the number of available longitudinal individual-level databases. There is scene to contribute to this literature, especially in the context of a developing country such as Brazil, where empirical studies have only recently started to cover these possibilities.

When compared to the literature, at least to what is the knowledge of the author, this is the first study to compare different theoretical perspectives regarding the industrial scope of agglomeration economies in Brazil while controlling for the longitudinal structure of the database at the individual level.

2.3 Analytical Framework

The theoretical framework adopted here is based on the wage equation developed by the urban economics literature. This approach is not new, especially in its aggregated regional version, which has been presented in a variety of studies ever since the 1970s. However, the recent wave of newly available individual-level longitudinal databases have allowed a proper control of individual heterogeneity. Following Combes, Mayer and Thisse (2008b), profit optimisation by a price-taking firm j may generate the following equilibrium wage:

$$w_j = \mu(1 - \mu)^{(1-\mu)/\mu} s_j \left(\frac{p_j A_j}{r_j^{1-\mu}} \right)^{1/\mu} = f(s_j, A_j, p_j, r_j). \quad (2.1)$$

In this case, the estimation can be conducted for individual i working in firm j , which requires detailed information at the individual level, while p_j and r_j capture, respectively, agglomeration and dispersion forces. The average unit value p_j of the good produced by firms can be raised by higher demand, weaker competition, or cheaper intermediate goods, leading to more agglomeration. On the other hand, r_j measures the effects of the prices of other production inputs, which can increase with congestion, and may provide a dispersion force. The term A_j is related to technological externalities (among others, knowledge and learning spillovers, and

existing technology associated with workers' abilities). Finally, it is necessary to control for individual skills s_j in order to capture the correct effect of local characteristics over salaries.

Within this context, when wages are regressed against a measure of city size, such as population density, the main underlying assumption is that this relationship happens through the elements discussed above. It is also important to control for all possible confounding variables. Moreover, the industrial mix must also be taken into consideration, especially because local wages can vary between sectors, which are heterogeneously affected by the degree of diversity, specialisation and concentration. These last factors are embedded in the term measuring technological externalities, namely A_j .

Combining all these elements, the next sections will discuss the estimation of the equation presented above in two stages. In the first stage (Equation 2.2), individual-level data is used to explain wages with individual characteristics and a dummy for each combination of region-sector-year, the spatial residual (GROOT; DE GROOT; SMIT, 2014a) – henceforth referred to as the spatial wage. This relationship can be written as:

$$\ln(w_{i,r,s,t}) = \beta_0 + \beta_1 age_{i,r,s,t} + \beta_2 age_{i,r,s,t}^2 + \sum_{edu} \beta_4 D_{i,r,s}^{edu} + \sum_{factor} \theta_{factor} Skill_{i,r,s,t}^{factor} + \sum_{firm_size} \beta_5 D_{i,r,s,t}^{firm_size} + \sum_r \sum_s \sum_t \lambda_{r,s,t} + \varepsilon_{i,t}. \quad (2.2)$$

Then this estimated spatial wage $\hat{\lambda}_{r,s,t}$ will be used as the dependent variable in the second stage (Equation 2.3), as it is associated with measures of agglomeration externalities and geographical characteristics.

$$\hat{\lambda}_{r,s,t} = \delta_0 + \delta_1 \ln(density_{r,t}) + \delta_2 area_{r,t} + \delta_3 specialization_{r,s,t} + \delta_4 diversity_{r,t} + \delta_5 competition_{r,s,t} + \sum_{geog} \delta_{geog} geography_{r,t} + \sum_s \delta_6 D_s^{sector} + \sum_t \delta_7 D_t^{year} + \sum_{macro-region} \delta_8 D_{mr}^{macro-region} + \xi_{r,s,t}. \quad (2.3)$$

Such a strategy in two stages is based on the argument that individuals residing in the same labour market and working in the same sector may share some specific characteristics that are not entirely captured by the controls included in Equation 2.2 and Equation 2.3. In this case, the error term of a single-stage estimation will be positively correlated across individuals from the same sector-region, generating downwardly-biased standard errors for the regional and sectoral level variables, which produce higher t statistics (COMBES; DURANTON; GOBILLON, 2008a; MOULTON, 1990, and BELL; NICKELL; QUINTINI., 2002, CARD, 1995 in a different context). A two-stage estimation procedure aims to overcome this issue.

A descriptive analysis of all variables included in the model will be presented in the next section. Among the individual characteristics considered in Equation 2.2: $w_{i,r,s,t}$ is the hourly wage; $age_{i,r,s,t}$ is measured in years; and $D_{i,r,s}^{edu}$ assumes value 1 when the individual has attained a certain educational level (edu = {illiterate, incomplete primary school, complete primary school to incomplete high school, complete high school to incomplete college, college degree or more}), or 0 otherwise. Moreover, $Skill_{i,r,s,t}^{factor}$ measures the degree of certain skills that are required for the job, following Maciente (2013)⁶. Finally, $D_{i,r,s,t}^{firm_size}$ represents a group of dummies for each firm size in which the individual works (aiming to control for technological patterns and competition differences between firms, GLAESER et al., 1992); and $\lambda_{r,s,t}$ is the sector-spatial wage.

Equation 2.3 presents a set of independent variables that aim to measure how city size and the industrial mix affect spatial wages. Urbanisation externalities are measured by the logarithm of employment density, and the area of the region⁷ is included to control for the scale effect (CICCONI; HALL, 1996, and COMBES et al., 2010). In addition, different measures of specialisation, diversity, and competition are considered in alternative specifications (see Table A.3 in the Appendix). Other controls include two-digit-sector dummies, year and macro-region dummies, as well as geographical characteristics (altitude and distance to the Equator line).

The estimation strategies have the following variations: in the first stage, an individual fixed effect may be included to control for the unobserved individual heterogeneity constant in time. Furthermore, in the second stage, the potential endogeneity of employment density is taken into account with instrumental variables (IV) techniques. The instruments considered here are the following: population density in 1940; distance to the coast; and dummies for sugar cane and coffee production, and gold exploration during the colonial period⁸. Table A.4 in the Appendix

⁶ See Table A.2 in the Appendix.

⁷ Regions are labour market areas (REGIC areas), which aggregate the municipalities in each region (482 regions containing all 5,565 municipalities), as defined by the Brazilian Institute of Geography and Statistics for 2007 (IBGE, 2013), taking into account all daily commuting and transportation connections among the municipalities. Taking labour market areas as the spatial unit of analysis is a standard procedure in the literature (see for instance GROOT; DE GROOT; SMIT, 2014a, and GLAESER; MARÉ, 2001).

⁸ These last three variables were generated and kindly provided by Naritomi et al. (2012). In the 1940s, the Brazilian economy was much less industrialised and productivity differentials were more related to agriculture. Therefore, population density in that decade is expected to be exogenous in relation to wages in the 21st century. Concerning distance to the coast, the colonisation process in Brazil was mostly concentrated on the coast, and a simple analysis of the distribution of cities in Brazil shows that there is still a high population concentration in that part of the country. Finally, the main economic activities developed during the colonial period helped to define the location of many urban centres, but are not necessarily related to the determinants of productivity nowadays.

provides a detailed description of all the variables considered, including methods of calculation and data sources.

The last methodological issue to be highlighted is that the regressions presented in Equations 2.2 and 2.3 are estimated separately for five aggregated sectors, defined in the Appendix (see Table A.5 for the aggregation criteria). The reason of for doing this is that it is possible to compare the effect of agglomeration externalities and the industrial mix over different development stages of manufacture and service industries, based on their technological and knowledge intensity. It also reduces the computational complexity of the estimation of the spatial wage⁹. The major drawback here is that the balanced panels consider only those individuals who do not move between these five aggregated sectors over time, but only within the industries belonging to each of them. Therefore, part of the individuals who move between sectors is not observed. As the group that moves is usually associated with a higher productivity, it is possible that the estimations will be downwardly-biased.

2.4 Data

The estimation of the wage equation discussed above is based on data of the formal Brazilian labour market. For this purpose, the main database considered here is the Identified RAIS (Annual Report of Social Information, from the Ministry of Labour), which consists of identified registration data of all formal firms and their employees, which focuses on the characteristics of the work contract. It provides annual information on all formally employed individuals in the private sector (or part of the public sector, depending on the type of contract), with a significant regional disaggregation (municipal level).¹⁰

Among the main advantages of this database, is that it has a longitudinal structure at both the firm and the individual level. Moreover, as it is a mandatory report, all firms are obliged to fill in the required information, covering the entire formal sector. This is also the reason why, differently from individual self-reporting surveys, there is a lower risk of wage under-reporting. However, a potential drawback of using this database is that it covers only the formal sector, and it does not include entrepreneurs (who are self-employed) or individuals in managerial positions. Another relevant detail is that the reporting process is more accurate for bigger firms,

⁹ In the case when the combination region-sector-year is considered for all industries simultaneously, the limit of Stata MP and R in a computer with RAM of 64G is reached.

¹⁰ The author is grateful to FIPE (The Economic Research Foundation Institute), Prof. Helio Zylberstajn and Eduardo Zylberstajn for allowing access to the database. All confidentiality requirements were respected.

which are usually located in larger urban areas, and sometimes firms with multiple branches may report incorrectly the municipality of each employee. Therefore, an analysis covering only the formal sector, with a better reporting for larger firms, may overestimate the urbanisation effect, as these bigger firms are usually more productive and located in large cities.

Following the process of formalisation of the labour market, the size of RAIS has increased significantly in the past decade. In 2004, 30.3 per cent of total jobs¹¹ were formal, while in 2012 this percentage increased to 39.3 per cent. It is also important to note that there is a significant regional heterogeneity in this process. In 2004, only 17.5 per cent of total jobs were formal in the North region of the country and 17.1 per cent in the North-east, while 39.4 per cent were formal in the South-east, 35.1 per cent in the South, and 28.6 per cent in the Centre-West. Then, in 2012, these percentages went up to 23.6 per cent in the North and 25.4 per cent in the North-east, while reaching 48.2 per cent in the South-east, 44.9 per cent in the South, and 40.2 per cent in the Centre-West.¹²

This issue is taken into account by comparing the relationships studied here over time (2004, 2008, and 2012). In order to explore the longitudinal structure of the database, a few steps were taken to generate a balanced panel for each period (2004-2008; 2008-2012; 2004-2008-2012), and the comparison between the first two data sets allows the discussion of the potential effect of formalisation on the characteristics of the labour force.¹³ Finally, database tractability was obtained by generating 20 per cent random samples for each of the five technological sectors.¹⁴

¹¹ Total jobs include public sector and military, formal and informal employees, self-employment, entrepreneurs, and unpaid work. The percentages discussed here refer to the representativeness of RAIS in the labour market.

¹² Information obtained from the National Household Sample Survey (PNAD), from the Brazilian Institute of Geography and Statistics (IBGE).

¹³ The original database contains individual information for each contract, meaning that individuals can appear more than once if they work for more than one firm. A few steps were taken in order to select the desired observations (ending up with only one contract per worker). Firstly, a filter selected only active contracts in December of each year for individuals working for private companies in permanent jobs. Then, contracts with missing individual IDs, wages equal to zero, or with less than 20 weekly hours were excluded. The next steps for individuals with multiple contracts included: keeping those with five or less contracts; dropping those with a different gender in each contract, and keeping the contracts with the highest number of weekly hours and with the oldest hiring date. Finally, for the remaining cases of multiple contracts, only one of them was randomly selected, leading to one contract per individual for the whole database. After all these procedures, and keeping only the manufacturing and service sectors, the database size ranged from 8.9 to 12.6 million between 2004 and 2012 (the increase is related to the expansion of the formal sector previously mentioned). Only those individuals who were observed in each period (2004 and 2008; 2008 and 2012; and 2004, 2008 and 2012 – depending on the model considered) were kept, with the same gender and birth date in every year. Finally, following a common practice in the related literature, the analysis was conducted for men who were 18 to 56 years old in 2004.

¹⁴ These samples are representative for the following characteristics: age group (less than 30 years old, 30 to 45 years old, 46 or more); population size (less than 100 thousand, 100 thousand to less than 500 thousand, 500 thousand to less than 1 million, 1 million or more); firm size (up to 4 employees, 5 to 9, 10 to 19, 20 to 49, 50 to 99, 100 to 249, 250 to 499, 500 to 999, 1,000 employees or more); macro-region; educational level (illiterate,

The descriptive statistics of the main variables of the first stage for the sample based on the period 2004-2008-2012 in 2012 are presented in Table 2.2, at the individual level. The average wage increases for sectors with higher technological or knowledge intensities. While Manufacturing high-tech (S3) and Services high-knowledge (S5) have relatively more workers with higher-education, Manufacturing medium-tech (S2) workers are more present in the groups with incomplete and complete middle school education. As expected, Manufacturing low-tech (S1) and Services less-knowledge (S4) are characterised by less qualified workers (up to incomplete primary school education).

In terms of the skills required for the job, Cognitive skills are more relevant for sectors with a higher intensity of knowledge or technology (S3 and S5). The same pattern is observed for Telecommunication, Attention, On-the-job experience, and Team-work skills. In the case of the service sectors (S4 and S5), they have more Assistance, Transportation, Artistic, Conflict management, Sales and Monitoring skills, while the manufacturing sectors (S1, S2 and S3) are associated with Maintenance and operation, Design and engineering, Accuracy and automation, and Independence skills. Age does not seem to change significantly among these groups of sectors, except for S4, which seems slightly higher.

In relation to their regional distribution, both service sectors (S4 and S5) have a similar distribution of workers in the sample in all regions, and the biggest disparities in this distribution are observed for the manufacturing sectors. S3 and S2 are relatively more concentrated in the South-east of the country (the richest region in terms of GDP), while S1 is more present in the North-east. Comparing the regional distribution of services and manufacturing, the South has a greater concentration of workers for S1, S2 and S3, while the Centre-West is relatively less important for S2 and S3.

Firm size also varies among sectors, as there seems to be a larger share of workers in bigger firms in S3 and S5, which indicates the need for scale to develop productive activities in these sectors, at least in 2012. Furthermore, the panel structure is extremely relevant to understand the results found in the next section. The share of individuals who do not change industry¹⁵ or REGIC area is very high, over 80 per cent for most sectors (except S5, in which it is slightly lower than 70 per cent). This indicates that the sample size necessary to obtain the variation to identify the spatial wage when individual fixed effects are included will be relatively small.

incomplete primary school, complete primary school to incomplete high school, complete high school to incomplete college, college degree or more); and technological sector.

¹⁵ Within the same technological sector.

Table 2.2 - Descriptive statistics of the main variables of the first stage of the model, 2012 (for the database containing individuals present in the whole period, 2004-2008-2012).

	S1		S2		S3		S4		S5	
	Manufacturing low-tech		Manufacturing medium-tech		Manufacturing high-tech		Services less-knowledge		Services more-knowledge	
	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
ln(real hourly wage)	2.36	0.66	2.67	0.69	3.08	0.73	2.24	0.58	2.60	0.92
age	40.26	8.93	41.03	8.63	40.86	8.49	43.36	9.17	41.38	8.72
Factor 1 - Cognitive skills	-0.63	0.74	-0.59	0.67	-0.34	0.77	-0.81	0.70	-0.23	0.98
Factor 2 - Maintenance and operation skills	0.47	0.93	0.69	0.87	0.48	0.95	-0.11	0.91	-0.51	0.77
Factor 3 - Assistance skills	-0.43	0.45	-0.47	0.44	-0.45	0.55	0.21	0.66	0.08	0.74
Factor 4 - Management skills	-0.16	0.90	-0.24	0.83	-0.21	0.96	-0.28	0.96	-0.05	1.17
Factor 5 - Design and engineering skills	0.03	0.69	0.20	0.75	0.39	0.93	-0.50	0.57	-0.42	1.06
Factor 6 - Transportation skills	-0.17	1.11	-0.17	1.00	-0.38	0.78	1.08	1.84	0.33	1.08
Factor 7 - Artistic skills	-0.32	0.68	-0.38	0.66	-0.40	0.67	-0.21	0.65	-0.16	0.83
Factor 8 - Accuracy and automation skills	0.05	0.72	0.12	0.69	0.02	0.73	-0.13	0.61	-0.05	0.86
Factor 9 - Supervised work skills	0.00	0.89	0.31	0.83	0.40	0.80	0.03	0.91	0.26	0.86
Factor 10 - Teaching and social science skills	-0.23	0.48	-0.18	0.44	-0.25	0.47	-0.20	0.50	-0.33	0.77
Factor 11 - Physical strength	0.15	0.80	0.16	0.82	-0.07	0.79	0.35	0.91	0.05	0.86
Factor 12 - Telecommunication skills	-0.29	0.86	-0.25	0.90	-0.02	0.94	-0.19	0.72	0.08	1.07
Factor 13 - Independence skills	-0.10	1.01	0.12	0.90	0.19	0.84	-0.62	1.22	-0.52	1.41
Factor 14 - Natural science skills	-0.16	0.81	-0.08	0.80	-0.15	0.79	-0.29	0.57	-0.12	0.87
Factor 15 - Attention skills	-0.19	0.81	-0.16	0.82	-0.06	0.89	-0.30	0.77	-0.04	0.91
Factor 16 - On-the-job experience	-0.32	0.90	-0.19	0.92	-0.11	0.89	-0.44	0.72	-0.10	0.84
Factor 17 - Conflict management skills	-0.30	0.83	-0.29	0.76	-0.26	0.75	0.28	0.76	0.50	1.04
Factor 18 - Team-work skills	-0.22	0.91	-0.13	0.85	0.00	0.74	-0.13	0.90	0.17	0.68
Factor 19 - Sales skills	-0.51	0.89	-0.58	0.82	-0.40	0.86	0.15	0.77	0.01	0.90
Factor 20 - Monitoring and compliance skills	-0.22	0.95	-0.11	0.82	-0.24	0.83	-0.08	1.21	-0.25	0.86
Factor 21 - Clerical skills	-0.38	1.13	-0.49	1.12	-0.17	1.04	-0.02	0.90	0.22	1.01
	S1		S2		S3		S4		S5	
Education attainment										
Illiterate	0.5%		0.3%		0.1%		0.4%		0.2%	
Incomplete primary school	27.8%		21.1%		10.9%		31.3%		11.7%	
Complete primary school - incompl. middle school	26.0%		23.7%		18.4%		30.7%		18.1%	
Complete middle school - incomplete college	39.2%		46.2%		50.2%		33.0%		41.3%	
College degree or more	6.6%		8.8%		20.5%		4.7%		28.8%	
Firm size										
Up to 4 employees	2.7%		2.0%		0.8%		11.2%		2.5%	
5 to 9	4.0%		3.7%		1.4%		15.8%		3.8%	
10 to 19	6.6%		7.6%		3.6%		10.9%		5.7%	
20 to 49	10.6%		12.4%		6.8%		14.8%		9.6%	
50 to 99	9.4%		12.1%		8.2%		9.6%		7.5%	
100 to 249	14.8%		15.7%		14.9%		11.1%		11.3%	
250 to 499	14.2%		15.3%		14.9%		8.8%		11.8%	
500 to 999	12.7%		13.8%		13.6%		8.4%		12.8%	
1,000 or more employees	25.0%		17.3%		35.8%		9.6%		35.1%	
Macro-region										
North	1.7%		1.3%		1.6%		2.1%		2.6%	
Northeast	14.7%		7.0%		2.8%		13.9%		13.2%	
Southeast	49.6%		70.2%		76.0%		66.8%		67.7%	
South	29.6%		20.0%		18.9%		12.8%		11.1%	
Center-West	4.4%		1.5%		0.7%		4.4%		5.4%	
Variation in the panel structure										
Individuals in the same industry and Area REGIC	85.4%		86.5%		83.7%		80.7%		69.1%	
Individuals in different industries	5.5%		4.1%		4.3%		7.2%		13.4%	
Individuals in different Areas REGIC	5.4%		6.9%		8.9%		6.7%		11.2%	
Individuals in different industries and Areas REGIC	3.7%		2.5%		3.2%		5.4%		6.3%	
Percentage of areas REGIC (total = 482)	89.2%		81.1%		56.0%		94.0%		88.4%	
Number of observations	45,902		34,487		37,840		72,014		58,350	

Source: Authors' own calculations.

Table A.6 in the Appendix provides a comparison of additional descriptive statistics for the first stage. Apart from the fact that individuals who change industry (within the same technological sector) and/or REGIC area are more educated, their mean salaries can be smaller than the full sample. This last element goes against the literature, which states that workers who move are positively selected and should be more productive.

Finally, it is not possible to observe individuals in the sample for all 482 REGIC areas of the country, because the 20 per cent sample restricts the analysis and the formal sector is not necessarily present in all areas of the country (especially when public administration is excluded). The share of observed areas is even smaller for S3, an indication that the high-technology manufacturing industries are more concentrated in specific parts of the country.

2.5 Results

In the previous section, it was mentioned that three different data sets would be analysed. Here the main results will be discussed in the case in which there are individuals present in the database for the years 2004, 2008 and 2012. At a certain point though, there will be a comparison using data for 2004-2008 and 2008-2012.

The first set of results refers to the first stage of the model (Equation 2.2), at the individual level (Table A.7). As mentioned in Section 2.3, there are two main alternatives considered here, a simple OLS estimation and a regression including individual fixed effects. When analysing the OLS estimation version, it is noticeable that most coefficients are statistically significant for all sectors, with the relative importance of age increasing with technology and knowledge intensity (when comparing the five sets of regressions, higher coefficients for age are found for S3 and S5). For each sector, age is also increasing at decreasing rates.

In the case of education, all coefficients are positive and increase in value with higher levels of educational attainment. In fact, individuals with a graduate or post-graduate degree obtain higher salaries than illiterate individuals (this difference ranges from 67 per cent to 83 per cent, depending on the sector). Only for the service sectors, are the two lowest educational levels non-significant. The estimated coefficients for each skill seem to follow the expected signs. For instance, Cognitive skills are positively associated with wages in all sectors, while Transportation skills are positive only for S4 and S5 (services). On the opposite side, the Physical strength coefficient has a negative sign for most sectors (non-significant only for S1),

which is in accordance with the literature. In the case of firm size, it is positively correlated to the individual wage, which means that bigger firms are supposedly more productive.

When individual fixed effects are considered (the last five columns of Table A.7), education variables are omitted, as they do not offer sufficient variation over time for each individual. Another relevant result is that many variables lose significance (i.e., age and some skills), as they are also captured by fixed effects. However, some elements are still present, such as the positive and significant coefficients for Cognitive skills, and increasing values with firm size. Both groups of regressions also include interaction dummies between REGIC areas, industries and years (as expressed in Equation 2.2). As mentioned in Section 2.3, the estimated coefficients of these dummies are directly used as dependent variables in the second stage, and are referred to as the spatial wage.

While the first stage aimed to control for individual heterogeneity, the second stage explores the relationship between the spatial wage and different measures of diversity, specialisation and competition, as well as of urban size. Therefore, it will be possible to investigate the industrial scope of agglomeration economies. Table 2.3 presents a brief descriptive analysis of the main variables included in the second stage, at the REGIC area-industry level for 2012.

Table 2.3 - Descriptive statistics of the main variables for the second stage of the model for 2012 (database 2004-2008-2012).

Dimensions	Variables	S1		S2		S3		S4		S5	
		Manufacturing low-tech		Manufacturing medium-tech		Manufacturing high-tech		Services less-knowledge		Services high-knowledge	
		mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
	Spatial wage (1st stage with OLS)	-0.154	0.426	0.438	0.837	-0.295	0.549	0.454	0.430	0.572	0.614
Urbanisation	Density (1,000 inhabitants / km ²)	27.5	68.4	31.4	72.3	48.1	94.6	27.2	67.3	32.2	77.1
Diversity	Inverse of Herfindhal Index	10.76	3.88	11.41	3.73	13.00	3.29	10.49	3.87	10.62	3.94
	Shannon entropy	2.77	0.35	2.83	0.30	2.96	0.22	2.74	0.35	2.76	0.36
	Share 5 biggest industries (except the own industry)	0.58	0.11	0.57	0.10	0.53	0.08	0.60	0.11	0.59	0.11
Specialisation	Location quotient	3.24	9.51	2.09	3.08	2.48	4.36	0.94	0.93	0.94	1.39
	Specialization degree	0.039	0.039	0.035	0.033	0.027	0.022	0.038	0.039	0.037	0.041
Competition	Competition degree	1.55	1.35	1.60	1.43	1.98	2.53	1.68	0.97	1.85	1.25

Note 1: The spatial wage presented here refers to the regression with an OLS estimation in the first stage.

Note 2: The unit of analysis is the combination REGIC area – industry.

Source: Authors' own calculations.

The urbanisation measure indicates that S3, S5 and S2, in that order, are associated with denser areas. The mean of diversity increases with technology intensity for the manufacturing sectors, and with knowledge level for the service sectors. This pattern is observed even for the diversity measure related to the share of the five biggest industries, because it is inversely related to diversity (a higher percentage indicates low diversity). In the case of specialisation measures,

notably for the degree of specialisation, this relationship is exactly the opposite (higher specialisation is more relevant for the lower technological and knowledge intensity sectors). The location quotient achieves a much lower value for the service sectors, probably indicating that services are more homogeneously spread across the country. The degree of competition shows a very similar pattern to that of the diversity measures. Finally, the mean spatial wage seems higher in S5, S4 and S2, but there is no clear pattern¹⁶. The maps below show its regional distribution in detail.

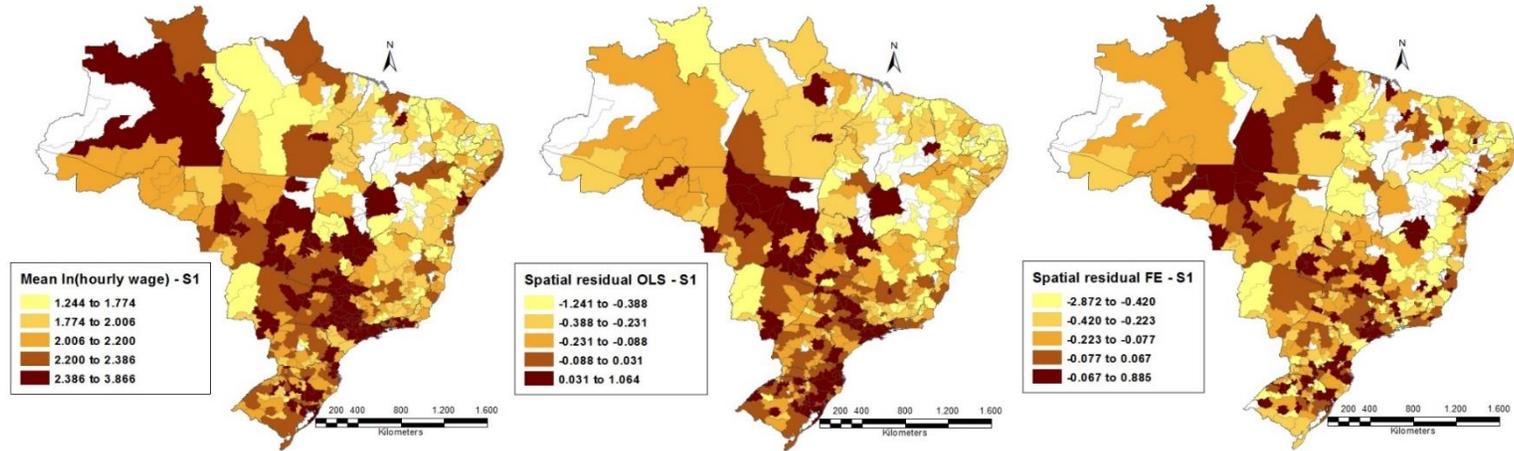
The set of maps presented below allows a number of different analyses. First, for each sector, the comparison of the first map on the left with the two others¹⁷ shows that when individual characteristics are taken into account, at least part of the regional distribution changes significantly. This is evidence that controlling for individual characteristics in this matter is essential for the analysis in the regional context. For instance, for S1 to S4, comparing the left and the centre maps, the Northern region of the country shows darker areas for the individual wage than for the spatial wage. For all sectors, the South-South-east of the country remains relevant in terms of having higher levels of the spatial wage (after controlling for individual characteristics). With regard to S3, there is a significant share of areas without information (44 per cent), due to a high regional concentration.

The inclusion of individual fixed effects in the first stage significantly changes the regional distribution of the spatial wage. As previously mentioned, in this case the spatial wage is calculated only for those individuals who moved between industries (inside the technological sector) and/or REGIC area over time (between 2004, 2008 and 2012). All maps are based on the 2012 data, which are associated with the last place in which these workers are observed. The fact that just a selected part of the sample of the workers who moved has been used to build the third column of maps explains why certain counterintuitive patterns emerge (such as darker areas in the countryside of the North-east, far from the coast) for most sectors.

¹⁶ The spatial wage can have negative values, as it is a deviation of the omitted coefficient of the combination REGIC area-industry-year.

¹⁷ The map in the centre refers to the spatial wage for the model with a pooled OLS estimation in the first stage, while the map on the right is related to the spatial wage obtained from the estimation with individual fixed effects in the first stage.

Sector 1 – Manufacturing: low-tech



Sector 2 – Manufacturing: medium-tech

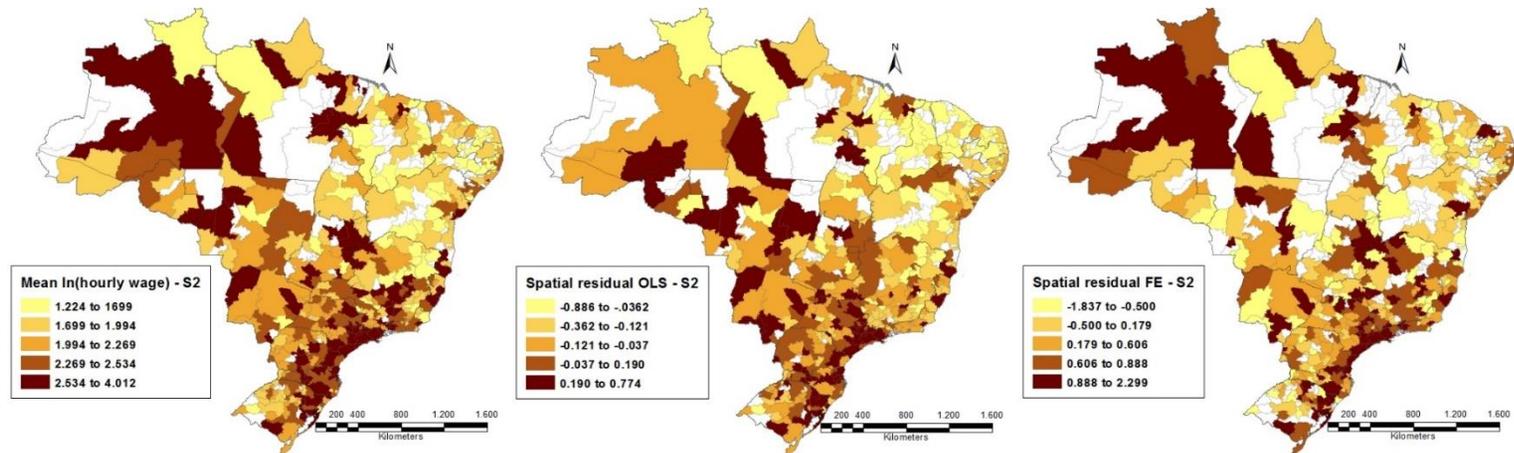
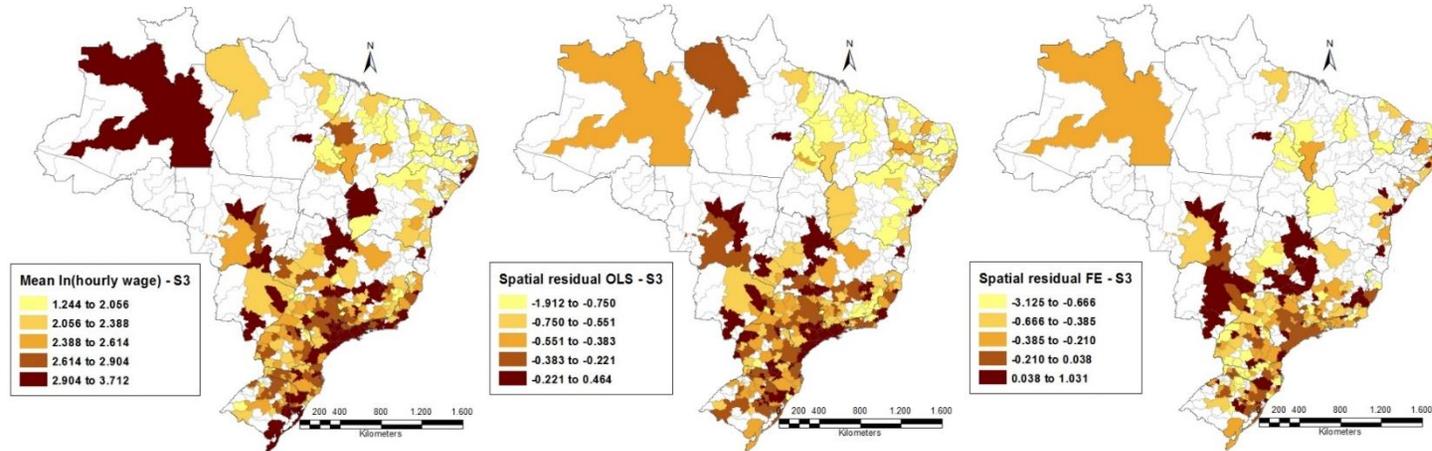


Figure 2.1 - Regional distribution of the mean logarithm of the hourly wage and the spatial wage (for OLS and FE in the first stage), by sector in 2012 (database 2004-2008-2012). (continues)

Sector 3 – Manufacturing: high-tech



Sector 4 – Services: less-knowledge

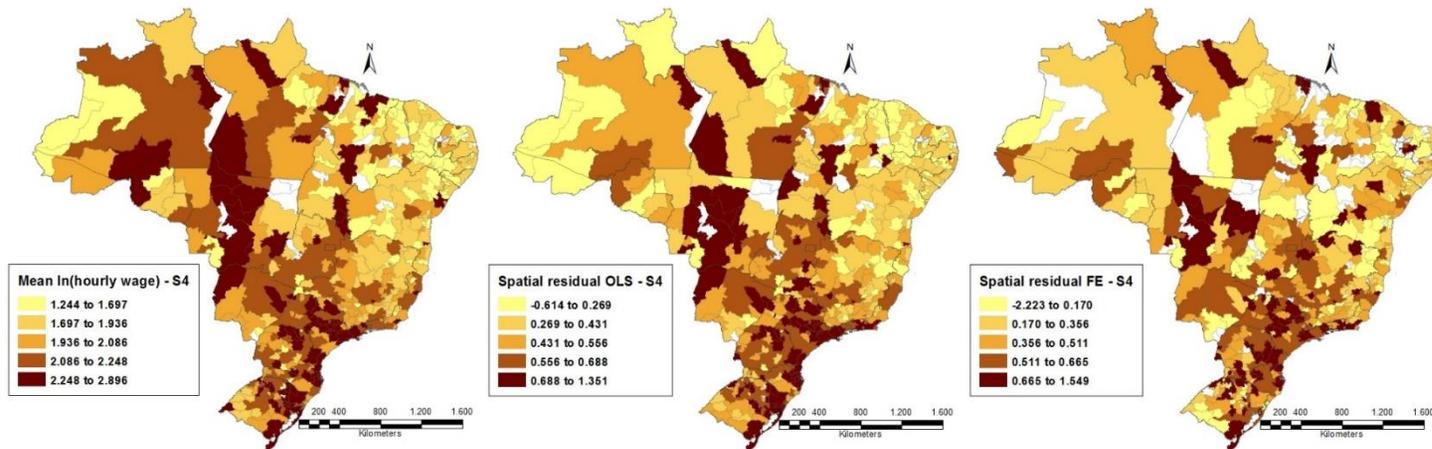


Figure 2.1 - Regional distribution of the mean logarithm of the hourly wage and the spatial wage (for OLS and FE in the first stage), by sector in 2012 (database 2004-2008-2012). (continues)

Sector 5 – Services: high-knowledge

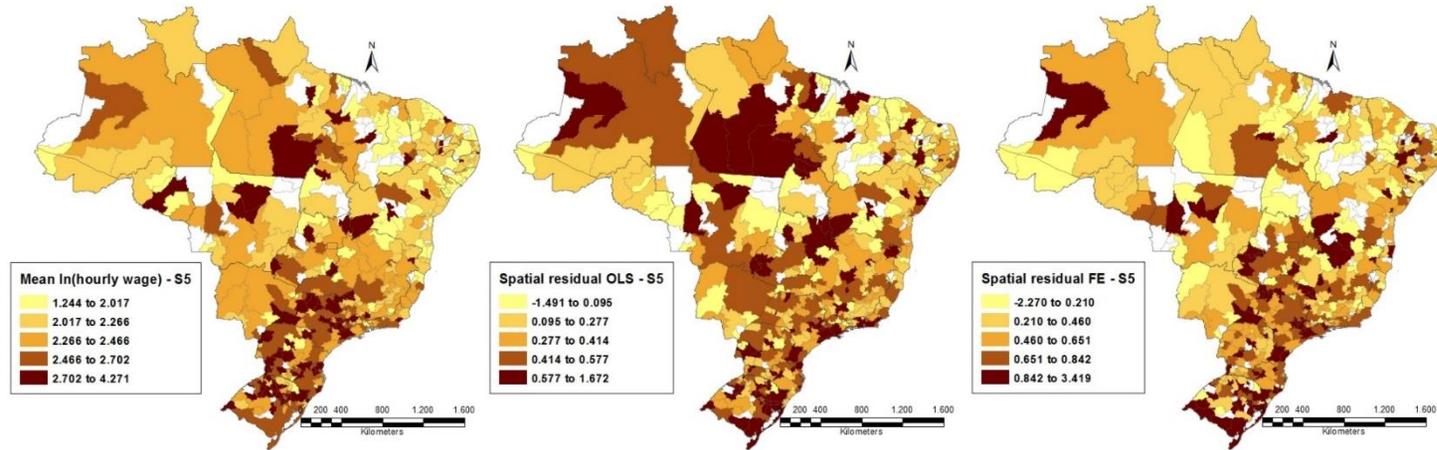


Figure 2.1 - Regional distribution of the mean logarithm of the hourly wage and the spatial wage (for OLS and FE in the first stage), by sector in 2012 (database 2004-2008-2012). (end)

Note 1: All maps refer to 2012, and, in the case of spatial wages, they are calculated as a simple average of each industry in the REGIC area inside the sector. In the case of the mean logarithm of the hourly wage, it is calculated as a simple average of the individual wages in the REGIC area-industry.

Note 2: The classes of each map are based on the division of quintiles of each variable.

Note 3: REGIC areas in white do not present any data in the sample and sector.

Source: Authors' own calculations.

Even if there is an indication of a possible spatial pattern in the data, for a number of reasons the possibility of estimating a spatial model was excluded. Firstly, the dependent variable in the second stage is not available for all REGIC areas. Moreover, by taking into account REGIC areas, many possible neighbourhood effects will be lost. This is because these areas are defined by daily commuting patterns, which already measure most of the relevant relations in the local labour market. If spatial effects were included here, they would possibly be capturing spurious relationships, which are absent in the specified model. These spatial effects could also capture part of the effects of local industrial composition. The main conclusion in this respect is that the spatial scale considered here minimises the need for spatial dependence analysis, which would be more justified on a smaller scale.

The dependent variable of the second stage is similar to the spatial wage depicted in the middle and right-hand maps (except that the maps show the average of the spatial wage over the various industries). Following Equation 2.3, a set of controls in the industry and area levels are included as explanatory factors, in addition to different combinations of measures for the degree of urbanisation, specialisation, diversity and competition. These different combinations, called *Compositions*, aim to identify a robust pattern in the results from the analysis of the industrial scope of agglomeration economies, even when the explanatory variables are changed.

Focusing initially on the five first columns of Table 2.4, which present the results for the second stage when there was OLS estimation in the first stage, there is a very robust indication for the urbanisation measure. Its coefficients are positive and statistically significant for all sectors, ranging from 0.0511 to 0.0940 in different *Compositions*. In all cases, the effects are ordered from the highest to the lowest as S3, S1, S5, S4 and S2. This can be considered as evidence that high-tech and low-tech manufacturing benefit more from urban scale. Another robust result is found for the diversity measures: there is a positive relationship for all sectors between more diverse areas and higher productivity (measured by spatial wages). The highest coefficients are observed for the manufacturing sectors, especially S3 (except in *Composition 4*, in which S1 and S2 have higher values).

Table 2.4 - Second-stage OLS regressions for the spatial wage, with and without individual fixed effects in the first stage.

		2nd stage with OLS (1st stage with OLS)					2nd stage with OLS (1st stage with individual fixed effects)				
		S1.	S2.	S3.	S4. Services	S5. Services	S1.	S2.	S3.	S4. Services	S5. Services
		Manufacturing	Manufacturing	Manufacturing	less-	high-	Manufacturing	Manufacturing	Manufacturing	less-	high-
		low-tech	medium-tech	high-tech	knowledge	knowledge	low-tech	medium-tech	high-tech	knowledge	knowledge
		S1	S2	S3	S4	S5	S1	S2	S3	S4	S5
Composition 1											
Urbanisation	ln(density)	0.0631***	0.0460***	0.0749***	0.0511***	0.0636***	0.0384***	0.191***	0.0318**	0.0711***	0.0600***
Diversity	ln(herfindhal)	0.0665***	0.108***	0.118***	0.0435**	0.0376*	0.0415*	0.233***	0.00680	0.0192	0.127***
Specialisation	ln(location quotient)	-0.00430	-0.0126	-0.0216	-0.00410	0.0135	-0.0186*	0.0649**	-0.0293	0.0282*	0.0225
Competition	ln(competition degree)	-0.0223**	-0.0656***	-0.0334*	-0.0478***	0.00284	-0.0388***	-0.122***	-0.0132	-0.0428**	-0.0155
	R ²	0.378	0.390	0.280	0.323	0.442	0.115	0.494	0.062	0.146	0.322
Composition 2											
Urbanisation	ln(density)	0.0788***	0.0564***	0.0901***	0.0577***	0.0678***	0.0565***	0.179***	0.0465***	0.0774***	0.0667***
Diversity	ln(shannon entropy)	0.345***	0.391***	0.400***	0.230***	0.0650	0.200**	0.630***	0.0254	0.109	0.272***
Specialisation	ln(specialisation degree)	0.0855***	0.0536***	0.0629***	0.0499***	0.0110	0.0628***	0.00789	0.0428*	0.0361***	0.0212
Competition	ln(competition degree)	-0.0151**	-0.0552***	-0.0134	-0.0463***	-0.00834	-0.0175*	-0.190***	0.0156	-0.0671***	-0.0346*
	R ²	0.393	0.393	0.284	0.327	0.441	0.121	0.490	0.063	0.147	0.320
Composition 3											
Urbanisation	ln(density)	0.0792***	0.0566***	0.0940***	0.0558***	0.0600***	0.0555***	0.164***	0.0452***	0.0750***	0.0556***
Diversity	ln(share 5 biggest sectors)	-0.222***	-0.285***	-0.165**	-0.206***	-0.190***	-0.150**	-0.749***	-0.0462	-0.124*	-0.405***
Specialisation	ln(specialisation degree)	0.0675***	0.0428***	0.0512***	0.0430***	0.0167	0.0533***	0.00718	0.0433*	0.0346***	0.0229
Competition	ln(competition degree)	-0.00751	-0.0516***	-0.0130	-0.0440***	-0.00544	-0.0118	-0.178***	0.0165	-0.0653***	-0.0261
	R ²	0.392	0.393	0.281	0.327	0.443	0.121	0.495	0.063	0.147	0.323
Composition 4											
Urbanisation	ln(density)	0.0786***	0.0581***	0.0926***	0.0587***	0.0654***	0.0554***	0.179***	0.0462***	0.0761***	0.0637***
Diversity	ln(herfindhal)	0.159***	0.158***	0.150***	0.0960***	0.0558**	0.105***	0.283***	0.0170	0.0677**	0.160***
Specialisation	ln(specialisation degree)	0.0907***	0.0564***	0.0641***	0.0507***	0.0168	0.0684***	0.0178	0.0438*	0.0411***	0.0310*
Competition	ln(competition degree)	-0.0123*	-0.0528***	-0.0118	-0.0454***	-0.00913	-0.0153*	-0.186***	0.0159	-0.0660***	-0.0358*
	R ²	0.398	0.396	0.287	0.328	0.442	0.123	0.492	0.063	0.148	0.322
Observations	N (REGIC * sector * year)	4912	3217	2301	5836	5208	4522	2895	2094	5378	4884

*: $\alpha = 0.10$; **: $\alpha = 0.05$; ***: $\alpha = 0.01$

Note 1: Additional controls are ln(area in square kilometres), ln(distance to the equator), ln(altitude), year dummies (2004 omitted), sector dummies, macro-region dummies (North omitted), constant.

Source: Authors' own calculations.

The specialisation measure considered in the first specification (location quotient), is non-significant for all sectors (and negative for most of them). However, all other *Compositions* include the degree of specialisation for this dimension, with positive coefficients (which are non-significant only for S5). Therefore, the location quotient seems unsuitable for the problem concerned, while the degree of specialisation is more suitable to capture the importance of this dimension for local productivity. Finally, the estimated coefficients of the degree of competition are negative in all cases. High significance is found only for S2 and S4, and S1 is slightly significant in most specifications. Therefore, spatial wages in S3 and S5 seem to be unaffected by local competition, while sectors with lower technological or knowledge intensity are negatively affected.

It is possible to compare the results presented in Table 2.4 with the theoretical frameworks summarised in Table 2.1. The MAR approach is associated with (+) specialisation, (-) diversity and (-) competition, while that of Jacobs is related to (-) specialisation, (+) diversity and (+) competition, and that of Porter is synthesised by (+) specialisation, (-) diversity and (+) competition. Hence, there is no clear pattern to indicate the more appropriate theoretical framework for each sector. In fact, for S5, it seems that only diversity is significant (and positive), suggesting at least partially that Jacobs's perspective is more adequate. For S1, S2 and S4, even if diversity is positive and significant, the combination of positive specialisation and negative competition coefficients indicate that Marshall could be more adequate to explain their patterns. In the case of S3, there are elements from both Marshall and Jacobs's perspectives. The result to be highlighted here is that the best possible industrial mix to foster productivity can vary for each sector, as it is associated with their technological intensity.

The last five columns of Table 2.4 depict the results for the spatial wage regression when individual fixed effects are included in the first stage. Therefore, only workers who have moved (between different REGIC areas and/or industries inside a technological sector) measure the spatial wage. Even if many of the coefficients change significantly from the previous analysis, in the case of density they are still positive and significant for all sectors. It is noteworthy that for S2 there is an important increase in the size of the elasticity (which was around 0.05 and went up to something around 0.16 and 0.19), but for the other sectors these elasticities assume a very similar value to the previous case. In the literature, most studies find that, when controlling for sorting (by including individual fixed effects), the estimated coefficient for urbanisation decreases significantly

(COMBES; DURANTON; GOBILLON, 2008a). Following Groot, De Groot and Smit (2014a), there is no consensus on whether the inclusion of individual fixed effects is the best strategy, because it may create a selection bias as the spatial wage will be measured only by migrants (who should be more associated with risk-taking, entrepreneurship, and adaptation, among other characteristics, which can be positively related to productivity).

In relation to the other variables that capture industrial composition, the signs are similar, but there are important changes in significance. Diversity measures are still positive, but non-significant for S3 and less significant for S4 in all specifications. The degree of specialisation becomes non-significant or slightly significant for S2, S3, and S5 in the different specifications. In the case of the degree of competition, elasticities are still negative but lose significance for S1, S3, and S5. Then, S4 gets closer to Marshall's explanation, while S1 and S5 spatial wages will be positively associated with diversity (Jacobs's perspective). Finally, the explanatory power of each set of models varies significantly. When the first stage is based on OLS, second stage variables are more relevant to explain the spatial wage, with an R^2 ranging from 0.28 to 0.44. On the other hand, the last five columns show a more heterogeneous pattern for this measure of fit: while the R^2 of S2 ranges around 0.49, the one for S3 is much lower (around 0.06). The fit of the model is also relatively high for S5, an indication that, for the remaining sectors (S1, S3, and S4), after controlling for individual heterogeneity, spatial wage variation is less associated with the local industrial composition.

The second issue explored here involves the potential endogeneity of the urbanisation measure (density). As discussed in Section 2.3, the inclusion of individual fixed effects aims to control for a sorting process, meaning that more productive individuals will migrate to bigger cities where they can find higher salaries, reinforcing city size and local productivity. However, it is also important to tackle the potential endogeneity of the quantity of labour (COMBES; DURANTON; GOBILLON, 2011). Table 2.5 presents a similar set of results as to those in Table 2.4, but with IV estimations. Once again, *Composition 1* seems less adequate, especially for the regressions with OLS in the first stage. The main conclusions seem to hold, especially for the density coefficients, which are still positive and significant. Focusing on the first five columns, once again (with OLS in the first stage), diversity seems to be positively associated with the spatial wage, even if for S5 in some specifications this coefficient is non-significant.

Table 2.5 - Second-stage regressions for the spatial wage, with and without individual fixed effects on the first stage, and with instrumental variables for ln(density).

		2nd stage with IV (1st stage with OLS)					2nd stage with IV (1st stage with individual fixed effects)				
		S1.	S2.	S3.	S4. Services	S5. Services	S1.	S2.	S3.	S4. Services	S5. Services
		Manufacturing	Manufacturing	Manufacturing	less-	high-	Manufacturing	Manufacturing	Manufacturing	less-	high-
		low-tech	medium-tech	high-tech	knowledge	knowledge	low-tech	medium-tech	high-tech	knowledge	knowledge
		S1	S2	S3	S4	S5	S1	S2	S3	S4	S5
Composition 1											
Urbanisation	ln(density)	0.0632***	0.0380***	0.0693***	0.0423***	0.0698***	0.0377***	0.184***	0.0582***	0.0568***	0.0561***
Diversity	ln(herfindhal)	0.0662***	0.122***	0.126***	0.0567***	0.0284	0.0425	0.245***	-0.0319	0.0406	0.133***
Specialisation	ln(location quotient)	-0.00423	-0.0163	-0.0235*	-0.00537	0.0156	-0.0189*	0.0615**	-0.0207	0.0260*	0.0211
Competition	ln(competition degree)	-0.0222**	-0.0698***	-0.0353*	-0.0540***	0.00851	-0.0392***	-0.125***	-0.00487	-0.0530***	-0.0191
	R ²	0.378	0.389	0.280	0.322	0.441	0.115	0.494	0.059	0.145	0.322
Composition 2											
Urbanisation	ln(density)	0.0971***	0.0574***	0.0900***	0.0597***	0.0800***	0.0681***	0.173***	0.0856***	0.0713***	0.0724***
Diversity	ln(shannon entropy)	0.272***	0.386***	0.400***	0.224***	0.0263	0.152*	0.655***	-0.131	0.130	0.254**
Specialisation	ln(specialisation degree)	0.0906***	0.0539***	0.0629***	0.0504***	0.0149	0.0660***	0.00629	0.0604**	0.0344**	0.0230
Competition	ln(competition degree)	-0.0135**	-0.0552***	-0.0134	-0.0453***	-0.00136	-0.0166*	-0.190***	0.0140	-0.0703***	-0.0313*
	R ²	0.391	0.393	0.284	0.327	0.441	0.120	0.490	0.058	0.147	0.320
Composition 3											
Urbanisation	ln(density)	0.0942***	0.0511***	0.0932***	0.0488***	0.0694***	0.0650***	0.147***	0.0847***	0.0620***	0.0524***
Diversity	ln(share 5 biggest sectors)	-0.179***	-0.303***	-0.167**	-0.228***	-0.162**	-0.122*	-0.798***	0.0355	-0.165**	-0.414***
Specialisation	ln(specialisation degree)	0.0745***	0.0405***	0.0508***	0.0407***	0.0203*	0.0579***	0.000175	0.0655***	0.0301**	0.0217
Competition	ln(competition degree)	-0.00778	-0.0516***	-0.0130	-0.0472***	-0.000994	-0.0121	-0.178***	0.0143	-0.0713***	-0.0277
	R ²	0.391	0.393	0.281	0.327	0.442	0.121	0.495	0.058	0.147	0.323
Composition 4											
Urbanisation	ln(density)	0.0956***	0.0592***	0.0945***	0.0590***	0.0776***	0.0663***	0.175***	0.0838***	0.0697***	0.0693***
Diversity	ln(herfindhal)	0.140***	0.157***	0.148***	0.0956***	0.0437	0.0927***	0.288***	-0.0112	0.0741**	0.154***
Specialisation	ln(specialisation degree)	0.0975***	0.0569***	0.0651***	0.0509***	0.0216*	0.0728***	0.0158	0.0654***	0.0387***	0.0332*
Competition	ln(competition degree)	-0.0108*	-0.0528***	-0.0118	-0.0452***	-0.00213	-0.0145*	-0.186***	0.0148	-0.0695***	-0.0325*
	R ²	0.397	0.396	0.287	0.328	0.441	0.122	0.492	0.058	0.147	0.322
Observations	N (REGIC * sector * year)	4912	3217	2301	5836	5208	4522	2895	2094	5378	4884

*: $\alpha = 0.10$; **: $\alpha = 0.05$; ***: $\alpha = 0.01$;

Note 1: Additional controls are ln(area in square kilometres), ln(distance to the equator), ln(altitude), year dummies (2004 omitted), sector dummies, macro-region dummies (North omitted), constant.

Note 2: Instruments for ln(density) are ln(density in 1940), ln(distance to the coast), dummy of sugar cane production in colonial times, dummy of gold prospection in colonial times, dummy of coffee production in colonial and imperial times.

Source: Authors' own calculations.

There is no clear order of effects among sectors, except to highlight the fact that S3 has the highest elasticities. For competition, the coefficient of S5 is non-significant and the one for S1 becomes less significant. These results also indicate that specialisation is relatively more important for S1 and then S3, while the elasticities of competition show a stronger negative effect for S2 and S4. Therefore, the inclusion of instrumental variables does not change, in a relevant way, the main conclusions drawn from Table 2.4, which is evidence of robust results.

Another robustness check involves comparing these models for different time periods. Instead of analysing all years simultaneously (2004-2008-2012), it is possible to break them down into two groups: individuals who are observed in 2004-2008, and individuals present in 2008-2012. The sample will not necessarily be composed by the same individuals as in the previous case, because now the requirement is that the individual is observed in the database in only two instead of three years.

Brazil has seen a significant formalisation process over the last decade, leading to an increase in the original database of this study (see Section 2.4). It is not possible to analyse the whole labour market simultaneously, owing to the lack of identified information at the individual level for the informal sector. However, it is necessary to investigate whether and how this formalisation may have changed productivity determinants over the country. The aforementioned comparison of two periods provides elements to address this issue at least partially, and, at least indirectly, controls for the regional heterogeneity in the formalisation process identified in Section 2.4. Table A.8 in the Appendix provides the main results for the models with IV in the second stage and OLS in the first. This set of results can be compared with the first five columns of Table 2.5.

In fact, significance does not seem to be affected in a relevant way by the split of the database, meaning that the relationships explored here are present during the whole period. The main change refers to the size of the main elasticities, which decrease from 2004-2008 to 2008-2012. This is especially true in the case of the urbanisation measure, with the highest decreases happening for S2, S4 and S1. A possible explanation for this result is that this formalisation process was stronger in more remote areas, places in which initial informality was higher. It is expected that the urban size will be smaller for these new areas with incoming workers, reducing density elasticities in relation to the spatial wage. For the other variables there is no clear pattern, with increasing or decreasing coefficients, depending on the sector. However, the fact that signs, significance, and relative size do not change significantly is another indication of the robustness of the results.

In summary, the main results found here seem to be reasonably robust to different specifications and time spans. The most suitable industrial mix for each sector can vary, meaning that not necessarily only one theoretical framework can explain the economy as a whole. There seems to be a heterogeneous effect for different sectors, a fact that must be taken into account when proposing public policies aiming to foster productivity. Moreover, private sector agents should balance the most relevant factors for their industry when choosing its location.

2.6 Conclusion

This chapter has aimed to contribute to the literature on static and dynamic agglomeration effects, controlling for individual skills, and comparing different estimation strategies. Focusing on the industrial scope of agglomeration economies in Brazil, a large set of results covering different concerns usually present in this literature is provided.

The analysis was based on identified registration data covering all formal firms and their employees in Brazil in three particular years (2004, 2008 and 2012). After all proper cleaning procedures, a sample was selected, with information for almost the whole country. Then the two-stage estimation procedure allowed the following variations: a first stage for the logarithm of the hourly wage at the individual level, with and without individual fixed effects; and a second stage for the spatial wage (obtained as the estimated interaction dummies area-sector-year in the first stage), with and without instrumental variables for employment density. In addition, the analysis was divided in two time periods, 2004-2008 and 2008-2012, in order to assess whether the formalisation process seen in Brazil in this period could have affected the results.

With a separate analysis for each of the five broad sectors (S1 – Manufacturing low-tech; S2 – Manufacturing medium-tech; S3 – Manufacturing high-tech; S4 – Services less-knowledge; and S5 – Services high-knowledge), the main conclusion is that the best possible industrial mix to foster productivity can vary for each sector, being associated with their technological intensity. An unambiguous and universally valid theoretical framework can apparently not explain the results found for the different sectors.

Synthesising the theoretical alternatives, the MAR approach is associated with (+) specialisation, (-) diversity, and (-) competition; Jacobs's approach is related to (-) specialisation, (+) diversity, and (+) competition; and Porter's approach is summarized by (+) specialisation, (-) diversity, and (+) competition. In the simplest estimation (with OLS in the

first and second stages), for S5 it seems that only diversity is significant (and positive), suggesting that the Jacobian perspective is more appropriate. For S1, S2, and S4, even if diversity is positive and significant, the combination of positive specialisation and negative competition coefficients indicates that the MAR framework could be more adequate to explain these patterns. In the case of S3, there are elements from both Marshall's and Jacobs's perspectives. These results seem to be robust to different specifications and estimation strategies.

Another important aspect to be highlighted is that, when splitting the database into two periods (2004-2008 and 2008-2012), even if the main conclusions remain unchanged, the size of the elasticities for urbanisation economies is much smaller in the second period, for all sectors. This can be evidence of the fact that the recent formalisation process was more concentrated in the remote areas of the country (with higher initial informality). Therefore, the urban size in these new areas is expected to be smaller, reducing density elasticities in relation to the spatial wage for the newcomers in the formal labour market.

Finally, the urbanisation economies coefficient (the logarithm of employment density) is positive and significant for all sectors, ranging from 0.0511 to 0.0940 in different specifications, under the simplest estimation (using OLS in the first and the second stages). Ordering the effects between the sectors from the highest to the lowest, we find the following sequence: S3, S1, S5, S4 and S2. This can be considered as evidence that the high-tech and the low-tech manufacturing sectors benefit more from urban scale in Brazil, followed by services associated with higher knowledge.

Consequently, different city sizes can be more appropriate for different industries. The local industry mix can also vary for each type of activity, which means that sometimes diversity is more relevant, while on another occasion specialisation can be more important, as well as competition. This implies that both public policy and private sector actions must consider this heterogeneity when searching for the most appropriate incentives and locational strategies to increase productivity in a certain region or specific production unit.

3. SHOULD I STAY OR SHOULD I GO? SELECTION ON MIGRATION AND LEARNING IN CITIES IN BRAZIL

3.1 Introduction

Cities are more important than ever for economic growth and life in society. As discussed before, people living, working, and thinking together in dense areas is a sign of the success of urban agglomerations (GLAESER, 2011). In the United States, Glaeser and Maré (2001) find that in metropolitan areas, workers earn 25% more than their non-urban counterparts do. From this perspective, increasing urbanisation is interpreted as a sign of gains from agglomeration (net from costs). In other words, the trade-off between agglomeration economies and congestion costs generates cities (BEHRENS; DURANTON; NICOUD, 2014).

As discussed in the previous chapters, there are three main sources of regional wage disparities in any given moment of time (COMBES; DURANTON; GOBILLON, 2008a): (i) the composition of the local labour market; (ii) the availability of local non-human endowments that can increase productivity; and (iii) agglomeration economies. The third item is related to wage differences that follow from close proximity between firms and workers/consumers, thick labour markets and knowledge spillovers (DURANTON; PUGA, 2004). Contributing to this discussion, Behrens, Duranton and Nicoud (2014) divide (i) into two different mechanisms that take place in large cities: a sorting process of more talented individuals, and the selection of the most profitable firms, due to tougher competition. According to them, natural advantages become less important in the modern world.

Most of the literature has been focusing on the static effects of agglomeration economies. In general, the urban wage premium can range from 1% to 11% depending on the sample and the country. The mechanism behind the positive relationship of urban size and wages relies heavily on the assumption that wages, under imperfect competition, are supposed to be higher in places that are more productive (DURANTON, 2014a). Thus, agglomerations may generate economic advantages that are not completely internalized by firms and workers (as discussed in Chapter 1), which may be reflected in higher productivity in larger urban areas.

In addition to their static and instantaneous effect, agglomeration economies can have a long-lasting dynamic impact on productivity (COMBES; GOBILLON, 2015). Different processes may take place over time: people can learn from each other (DE LA ROCA; PUGA, 2014), or they can migrate in search of better opportunities, and return to the original city due to

unmatched expectations (DE LA ROCA, 2015; BEHRENS; DURANTON; NICOUD, 2014; PAPAGEORGIOU, 2014). These two last processes are confounding factors to the measurement of the dynamic effects of agglomeration economies. They are related to the aforementioned sorting and selection processes and affect the supply of different skills and qualifications in each local labour market over time. On the other hand, learning can influence the way individuals benefit from agglomeration over time.

The empirical literature of agglomeration economies in a developing economy such as Brazil is still very limited, based on estimations at the regional level or with individual cross-sections. As far as the author is aware, there is no study covering dynamic agglomeration economies in the Brazilian context, or for developing countries (DURANTON, 2014b). Therefore, this chapter aims to fill this gap in the empirical research in Brazil, as well as to contribute to the methodological discussion of the literature. This is done by first providing conclusions on the way individuals make their locational choices. Then, a second investigation aims to explore how they benefit over time from agglomeration advantages. Two different models developed elsewhere will provide the basic reduced-form equations to explore these issues.

The first set of results covers an evaluation of the main factors that affect initial and return migration. This is important because it provides a better understanding of the sorting process that may be a confounding factor for the correct measurement of the urban wage premium. Then, the second set of results aims to understand how agglomeration gains vary over time for workers in different-sized cities. A large longitudinal database of administrative reports covering the period from 1995 to 2008 is employed (RAIS-MIGRA), following individuals in their jobs over time, in different firms and cities.

The next sections are organised as follows: Section 3.2 contains a literature review on static and dynamic externalities, selection in initial and return migration and on the process of learning in cities. Then, Section 3.3 discusses the theoretical frameworks; Section 3.4 presents the data; Section 3.5 provides the main descriptive statistics, Section 3.6 explores the results; and, finally, Section 3.7 draws the main conclusions from this analysis.

3.2 Literature review

Agglomeration economies may appear in diversified and/or specialised labour markets, with different degrees of interaction between workers and firms (see Chapter 1). Over time, they can be strengthened by knowledge spillovers, which increase the productive capacity of a certain

area. It is in this context that dynamic agglomeration externalities should be understood, as they represent changes over time in local productivity, derived from this increasing capacity for innovation, creating new ways of production, and reducing costs.

In addition to dynamic agglomeration economies, the process of sorting of more productive workers into larger cities can also foster local productivity over time. This means that sorting can be a confounding factor of the advantages derived from agglomerations over time. The next subsections will briefly discuss the main aspects investigated by the literature regarding selection in initial and return migration, and dynamic agglomeration externalities.

3.2.1 Static versus dynamic externalities

As mentioned in Chapter 1, static agglomeration economies are defined as once-and-for-all effects of urban size on productivity, with no further impacts in the following periods. Nonetheless, it is also possible that the benefits of agglomeration last longer than one period, with a nonlinear effect on wages and heterogeneous impacts on individuals with different observed and unobserved characteristics. In this context, dynamic agglomeration economies aim to measure the long-lasting effect or permanent impact of urban density on productivity (COMBES; GOBILLON, 2015). In other words, static advantages relate to the idea that large cities are more productive than smaller ones in a specific moment of time, while dynamic advantages are associated with different trajectories of productivity for each city size over time (CAMAGNI; CAPELLO; CARAGLIU, 2015).

As discussed in Chapter 2, the urban economics literature provides strong evidence of the existence of an urban wage premium. This type of analysis can be found, for example, in Combes, Duranton and Gobillon (2008a), Glaeser and Maré (2001), Melo and Graham (2009), Mion and Naticchioni (2009), Ciccone and Hall (1996), Combes et al. (2010), Groot, De Groot and Smit (2014a), Andersson, Klaesson and Larsson (2014), and Groot and De Groot (2014b). Literature reviews are provided elsewhere (PUGA, 2010, and ROSENTHAL; STRANGE, 2004). The elasticity of wages with respect to population density can range from 1% to 11% depending on the sample and the country.

Many studies were recently conducted in different countries because longitudinal individual data on labour market outcomes was made available. However, as discussed in Chapter 2, the literature of agglomeration economies in Brazil is still very restricted, usually based on cross-sectional analysis or aggregated data at the regional level. So far, there has been no study

covering dynamic agglomeration economies considering wage evolution over time in the Brazilian context or for an emerging economy, at least according to the author's knowledge.

Among the possible channels that generate permanent or long-term impacts of agglomeration economies, local productivity growth can be driven by technological spillovers, and individuals may learn more and faster in bigger cities (LUCAS, 1988). With regard to the latter aspect, working in bigger cities increases the opportunities of learning (DE LA ROCA; PUGA, 2014). Glaeser (1999) argues that the rate of interaction between people is accelerated by urban density, leading to an increase in human capital accumulation. This effect is particularly significant for high-skilled people, usually more concentrated in big cities. Yankow (2006) points out that this learning hypothesis requires that individuals stay for a while in the city, meaning that there is no expectation of immediate wage gains when moving to a bigger urban concentration or any sudden wage loss when workers move to smaller cities.

Regarding the empirical estimation of dynamic agglomeration economies, it is still not clear which is the best strategy to analyse and separate them from static effects, due to data limitations. Subsection 3.2.3 will detail these issues and present the most recent empirical studies.

3.2.2 Migration and selection bias

Migration is very relevant to the spatial redistribution of workers and helps to explain the skill composition of cities over time. Even though migration is expected to equalise real regional wage differentials, it can actually reinforce regional disparities, as migrants are usually more skilled and less risk-averse than stayers are (GREENWOOD, 1997). For developing countries, special attention has been given to rural–urban movements and to the interaction of migration and formal and informal sector outcomes (LUCAS, 1997).

In Brazil, numerous studies have investigated the relationship between internal migration and regional inequality by comparing the wages of migrants and stayers, or of return and permanent migrants. Santos Junior, Menezes Filho and Ferreira (2005) suggest that migrants seem to form a positively selected group, with higher earnings in the destination than individuals both at the origin and at the destination. On the other hand, Ramalho and Queiroz (2011) find that return migrants are negatively selected in relation to permanent migrants, but those more qualified obtain wage increases by returning (probably because they find a more suitable occupation for their qualification). Both these studies are based on individual cross-sections.

The proper study of migration should include controls for unobserved and observed individual characteristics, as well as the comparison of individual outcomes over time, and this requires longitudinal databases. In Brazil, the main source for this type of data is RAIS-MIGRA.¹⁸ Based on a sample of individuals from this database, Taveira and Almeida (2014) show that migration flows of qualified migrants between municipalities in Brazil are mostly driven by higher expected income in the destination, population, GDP per capita, the degree of industrialisation and better amenities. Furthermore, a spatial panel technique indicates that the characteristics of neighbouring areas are also relevant in this migration decision. Freguglia, Gonçalves and Silva (2014) note that after controlling for individual fixed effects, wage differentials between origin and destination states are very relevant in determining the migration of skilled workers. In addition, these migrants look for states with higher prosperity, higher population density, better urban amenities and higher dynamism.

The results found by Freguglia and Procópio (2013) indicate that after controlling for individual unobserved characteristics, changing jobs and interstate migration are the most important factors in determining regional wage differentials. Furthermore, the effect on wages of changing firms is lower than the effect of moving to a different municipality.

However, these studies do not necessarily focus on the relationship between migration and the sorting process that runs parallel to the extraordinary gains derived from large agglomerations. Papageorgeou (2014) and De La Roca (2015) attempt to develop dynamic models of occupational choice that are associated with the different experiences migration can provide to workers. They show that the first decision to migrate differs between skill levels and can be related to the size of the city in both the origin and the destination.

More specifically, Papageorgeou (2014) finds that individuals who have migrated to bigger cities in the United States are more likely to change jobs with a higher frequency, but over time this difference disappears. Furthermore, workers in big cities have a lower propensity to migrate. The positive effect on wages of moving to another city and switching occupation is higher than the one obtained from changing city and staying in the same 3-digit occupation.

According to De La Roca (2015), the sorting of more productive workers into bigger cities in Spain may be related to higher earnings in these areas. Their findings suggest that there is a selection by observable characteristics in initial migration, resulting from the fact that the

¹⁸ This database is provided by the Ministry of Labour as a derived product from the identified RAIS (which was used in Chapter 2), and will be described in Section 3.4, and has limited access due to confidentiality issues.

probability of migration is positively associated with initially more productive workers (with higher education attainment and higher occupational skills before moving). The second round of sorting (return migration) is associated with low realized earnings in the first destination (in the case where the first destination is a big city), while high-skilled workers and individuals with high educational attainment are less likely to return.

Therefore, within the context of agglomeration economies, it is important to investigate not only the decision to migrate, but also who benefits the most from agglomeration, and who manages to stay in the destination or has to return. This is crucial in characterising the composition of local labour markets. Berry and Glaeser (2005) show that there is an increasing clustering of skilled workers in cities with originally high skill levels. In the same vein, Elvery (2010) finds that the skill mix employed by firms in smaller cities is lower than the one observed in larger cities. Therefore, large urban areas seem to concentrate a relatively higher percentage of skilled workers, partially because of this migration process.

In summary, larger cities seem to attract and select more skilled individuals. The cost–benefit analysis of living in a big city, with all the costs involved, may not prove positive for low-skilled individuals or for those who are negatively surprised by the income obtained in the destination. Therefore, not only is city size related to the attraction of more qualified workers, but it can also induce less-skilled individuals to leave large agglomerations.

3.2.3 Learning in cities

The last subsection emphasised the importance of investigating selection in initial and return migration. These phenomena are very relevant to understanding the changing skill composition of cities. According to Glaeser and Resseger (2010), agglomeration effects are stronger in cities with higher skills. They speculate that the link between human capital and agglomeration economies appears because workers learn more and acquire more skills in big and high-skilled cities, or because technological change can be faster in these places. These perspectives raise the question of how knowledge externalities evolve, and who benefits the most from these gains over time.

The literature on the static urban wage premium has explored at length the main estimation issues and strategies to identify agglomeration economies. As discussed before, there is a fairly well-developed empirical literature on this matter, even though it is still difficult to give a clear interpretation of the results, due to weak links between estimated specifications and theoretical

models (COMBES; GOBILLON, 2015). On the other hand, only recently dynamic agglomeration economies have started to be investigated, requiring longitudinal individual data and representing a further development from static agglomeration estimates.

De La Roca and Puga (2014) conducted one of the most complete empirical studies in dynamic agglomeration economies so far. Their main goal was to separate the possible reasons why firms would pay higher salaries in bigger cities: (i) static advantages; (ii) sorting of more productive workers; (iii) dynamic advantages (cities facilitate learning and experimentation). Therefore, they estimated a Mincerian wage equation, including individual fixed effects to capture the sorting process and the number of years of experience each individual has of working in different city sizes. This model was estimated using a very large Spanish panel data (the whole dataset comprises the period from 1981 to 2009, and the authors evaluate workers' wages from 2004 to 2009). Two restrictions applied to the model discussed in the previous sections are that the authors consider only a few groups of city sizes (instead of allowing a continuous range of city sizes) and that the individual heterogeneity effect is the same in the dynamic and the static contexts.

Carlsen, Rattso and Stokke (2013) adopt a similar strategy. Firstly, they focus on the size of the static urban wage premium across education groups, finding increasing elasticities according to educational attainment. However, when unobserved fixed effects are controlled for, these differences disappear, and the common agglomeration elasticity for static effects is around 0.03 (half of the elasticity obtained without controlling for unobserved characteristics). Their data for Norway also allow the estimation of dynamic agglomeration economies, following the strategy of De La Roca and Puga (2014). They separate the effect of experience obtained in large cities from experience obtained elsewhere, and calculate different coefficients for the years worked in the top-ten high wage sectors in comparison to other sectors. Their conclusions are that the aggregated wage premium is not affected by the inclusion of worker experience history. However, the medium-term premium for highly educated workers in high wage sectors is positively affected.

Using a different approach, D'Costa and Overman (2014) consider a large panel (1998–2008) of British workers, aiming to evaluate whether the sorting of high-ability workers can explain urban wage premiums and whether this wage premium is immediately given to workers or it is paid over time through faster wage growth. Their conclusion is that both learning and sorting matter in understanding the effects of cities on wage growth. The main restriction imposed by

these authors is that even if they estimate the specification with first differences, allowing for a distinct heterogeneous effect in the static and the dynamic cases, they have to exclude movers from the analysis (therefore, this study does not measure between-city dynamic effects).

A few years earlier, Wheeler (2006) estimated the impact of density on annual wage growth, calculating the within-job and between-job components of this growth, with data for young male workers in the United States. The author's main finding is that there is a positive relationship between wage growth and city size. However, when they control for individual fixed effects, there is no evidence of this urban premium on wage growth.

Dynamic agglomeration gains can be perceived as a reflection of occupational progression. Gordon (2015) shows that the high living costs in large metropolitan areas can only be met by more ambitious individuals. Furthermore, occupational advancement usually results from a combination of the exposure to learning opportunities and the capacity to profit from them (experience and dynamic human capital, respectively). In this context, ambition is supposed to be a key factor, and among the migrants' group, the share of more ambitious individuals seems to be higher. These individuals will seek and enjoy learning and escalating opportunities, which are more likely to appear in the diversified environment of a large urban centre.

The strategy followed by Rattso and Stokke (2011) is slightly different. They look for regional income divergence that could have been caused by migratory movements from the periphery to cities over more than three decades in Norway. The comparison of regional income distribution over time indicates that what actually happened was a process of convergence, contradicting the hypothesis of agglomeration economies generated by migration.

Finally, Yankow (2006) regresses the wage growth against changes in the location of work, separating the mobility effect from the growth effect. The author's conclusions regarding male young workers in the United States is that when they move into cities, they experience a significant wage growth (6 percentage points) in comparison to individuals that stayed in a rural area (with a symmetric effect for out-of-the-city migrants). The main drawback in this author's strategy is that it is not possible to control for sorting on unobservable variables, a limitation that D'Costa and Overman (2014) try to overcome.

In summary, there are different possibilities to estimate dynamic gains from agglomeration economies, but none of them covers all the relevant issues simultaneously. On the one hand, it is possible to identify this effect on the level of wages with movers (DE LA ROCA; PUGA, 2014) or with stayers (D'COSTA; OVERMAN, 2014), over wage growth (YANKOW, 2006;

WHEELER, 2006), controlling or not for individual unobserved characteristics (sorting). Here, different strategies will be compared in order to provide as complete an analysis as possible. It is clear, however, that this literature still has to overcome some important limitations.

3.3 Theoretical frameworks and empirical strategies

The literature on agglomeration economies has been mostly focused on partial equilibrium relationships in the labour market. From this perspective, the interactions between workers and firms can be analysed with a focus on the supply or on the demand side of the labour market. Section 3.2 discussed the main contributions in the literature to the analysis of selection in migration and of dynamic agglomeration advantages. Now, this section will briefly describe the models that will inspire the empirical analysis of these two subjects in the Brazilian formal labour market.

Here, the sorting process will be studied through the analysis of the migration decision and the importance of individual self-selection in this process. In this context, individuals decide to supply their work in different-sized cities. Such an analysis will provide elements to understand which type of worker is attracted to larger agglomerations. Next, these same workers are faced with the decision of whether or not to stay in the destination. Return migration becomes an additional element to define the sorting of workers. It may determine their resilient behaviour when facing adversities and higher life costs in these larger urban areas, as well as identify who benefits the most from bigger agglomerations.

After that, from the perspective of the demand side of the labour market, wages are evaluated as a reflection of the productivity achieved by firms and how much they are willing to pay to workers of a certain type. Embodying the possibility of learning in bigger cities, these wages may follow a path that will depend on the qualification of the worker, their previous experience in different city sizes and their own unobserved characteristics.

3.3.1 Selection in initial and return migration

The conceptual framework presented here presupposes that there is a pool of heterogeneous workers initially located in low-density cities L (see DE LA ROCA, 2015). Workers differ only by their skill level s_i . Each worker rents a house of standard quality and spends the remaining

income on a numeraire good. In L , housing costs are normalized to zero and the utility of individual i is:

$$U_i^L = s_i \quad (3.1)$$

When workers move to a city of type H (high density), they pay R for rent, receive a skill shock of $\delta_i \sim U[0, 2\delta]$ and have a productivity gain of α from working there.

$$U_i^H = \alpha(s_i + \delta_i) - R \quad (3.2)$$

Given that there is a moving cost C , worker i migrates from L to H if and only if $\alpha(s_i + \delta_i) - R - C > s_i$. Thus, workers with low skills will stay in L in equilibrium if:

$$s_i \leq \frac{R + C - \alpha\delta_i}{\alpha - 1} \quad (3.3)$$

On the other hand, initial migration from H to L will be pursued by workers with skill levels respecting:

$$s_i \leq \frac{R - C - \alpha\delta_i}{\alpha - 1} \quad (3.4)$$

Then, De La Roca (2015) computes the aggregated effects given the uncertainty of δ_i and the possibility of returning to the initial location by paying an additional moving cost C . The main predictions of the model regarding the selection in initial and return migration from L to H are the following:

- Workers with low s_i do not migrate from L to H ;
- Workers with intermediate s_i may migrate from L to H , and depending on the realisation of δ_i , they stay in H or return to L in a second period;
- Workers with high s_i migrate to H and do not return.

Furthermore, workers that are initially in H will observe their realisation of δ_i and decide whether or not to move to L . Then, selection in initial migration from H to L will respect the following conditions:

- Workers with low s_i migrate from H to L and do not return;
- Workers with intermediate s_i migrate from H to L only if they get a bad outcome of δ_i in H ;
- Workers with high s_i do not migrate to L .

This simple framework provides the main expected migration movements between low-density (L) and high-density cities (H). With these elements, it is possible to investigate the main factors that affect workers' decision to supply their labour in cities of different sizes.

Therefore, following De La Roca (2015), it is possible to specify a single-exit discrete duration model, defined for the population that may migrate for the first time in each period. Whenever an individual migrates, he or she will be dropped from the population of interest. Each unit of analysis will be considered over time, with a different number of observed characteristics of the individual himself and the place he or she lives. The model to be estimated is based on the probability of migrating at time t , conditional on the fact that the individual has not migrated before t :

$$h(t) = P[T = t | T \geq t] = F[\beta_0(t) + \beta'_1 x(t)] \quad (3.5)$$

In this specification, T is the year of migration, F is a logistic cumulative probability function, $x(t)$ is the vector of observed characteristics, β'_1 is a vector of parameters and $\beta_0(t)$ is a duration parameter. The log-likelihood function of a logit model aggregates over time the exit probabilities in each t , with a migration indicator $Y_t = \mathbf{1}(T_i \geq t \geq e_i)$ that equals 0 in every year except the one prior to migration, in which it equals 1, and it is given by:

$$L(\beta) = \sum_{t=1}^{T_i} \left\{ \sum_{i=1}^N \mathbf{1}(T_i \geq t \geq e_i) [m_i Y_{ti} \log h_i(t) + (1 - m_i Y_{ti}) \log(1 - h_i(t))] \right\} \quad (3.6)$$

where m_i equals 1 when a migration movement is observed and 0 otherwise, e_i is the year of entry in the sample (which will be the same for the whole sample here), and T_i is the number of years until first migration happens. Then, $\hat{\beta}$ will be the maximum likelihood estimator of $L(\beta)$.

Apart from the direct estimation of the first migration decision, this strategy will be the basis for the analysis of the return migration. In this case, the sample will be comprised of individuals who have already migrated once.

3.3.2 Dynamic advantages from agglomeration economies

In Chapter 2, the framework adopted by the urban economics literature was responsible for generating the basic productivity specifications that later resulted in the reduced form models

estimated there (COMBES; MAYER; THISSE, 2008b). It is possible to depart from that model to obtain dynamic effects in the following way.

Firstly, Equation 2.1 from Chapter 2 can be rewritten to encompass the association of individual wages with a composite effect of local productivity, $B_{r,t}$ and the skill level of individuals, $s_{i,t}$. The components of $B_{r,t}$ are the following: pecuniary externalities that appear through local markets (p_j and r_j , referring to agglomeration and dispersion forces respectively), and pure local externalities (that do not appear through a market mechanism and include A_j , a measure of technological externalities).¹⁹

$$w_{i,t} = B_{r,t}s_{i,t} \quad (3.7)$$

The inclusion of skills is essential to capture productivity, which is supposed to be directly associated to wages. Although a component of local labour skills could represent the fact that high skills are overrepresented in large cities, it would capture part of the agglomeration effect. Therefore, individual skills seem to be more appropriate, as they are not determined by agglomeration economies. In addition, due to the great difficulty of estimating the effects embedded in $B_{r,t}$ separately, most studies in the literature quantify the overall effect of local productivity on wages (COMBES et al., 2011). Furthermore, the relationship between local characteristics and nominal wages is enough to capture the net effect of agglomeration on productivity.

In addition to static effects, in which agglomeration economies have a once-and-for-all impact on productivity, they may also have a long-lasting or even permanent effect. One of the possibilities to evaluate this hypothesis is based on the inclusion of the previous experience of workers in the same region or in other places. This term can account for all the elements included in $B_{r,t-k}$, with $k > 0$, and internalised by the worker.

Departing from Equation 3.7, Combes and Gobillon (2015) show how dynamic agglomeration effects can be included in the model. Following De La Roca and Puga (2014), it is possible to assume that the log hourly wage of worker i in region r at time t depends on:

$$w_{i,r,t} = \theta_r + \mu_i + \sum_{j=1}^R \delta_{j,r} \text{exper}_{i,j,t} + X'_{i,t} \boldsymbol{\beta} + \varepsilon_{i,r,t} \quad (3.8)$$

¹⁹ Scitovsky (1954).

where θ_r is a region fixed effect (which later on will be related to a measure of agglomeration²⁰), μ_i is a worker fixed effect, $exper_{i,j,t}$ represents the experience acquired in region r up until time t , $X_{i,t}$ is a vector of time-varying individual and job characteristics (which include a skill measure), β is a vector of parameters and $\varepsilon_{i,r,t}$ is the error term. An initial estimation static agglomeration effect could be based on the following:

$$w_{i,r,t} = \theta_r + X'_{i,t}\beta + \eta_{i,r,t} \quad (3.9)$$

where individual fixed effects and the term related to previous experience are omitted, and $\eta_{i,r,t}$ is the error term. A pooled OLS estimation of the static urban wage premium embedded in $\theta_{r,t}$ in this restricted model is likely to be biased upwards if individuals with high unobserved ability sort themselves into bigger cities, or if individuals with more valuable previous experience work in bigger cities. From Equation 3.9, a strategy to address the problem of the sorting of workers is to include individual fixed effects (COMBES et al., 2011; COMBES; DURANTON; GOBILLON, 2008a; COMBES; GOBILLON, 2015; GLAESER; MARÉ, 2001).

$$w_{i,r,t} = \theta_r + \mu_i + X'_{i,t}\beta + \zeta_{i,r,t} \quad (3.10)$$

As mentioned before, the static fixed-effects structure of Equation 3.10 may still generate an upward-biased estimation of the static urban wage premium, because the term related to experience is still omitted. Therefore, dynamic effects derived from previous experience in the same region or in other places should be included, if the formulation presented in Equation 3.8 is correct.

When dynamic agglomeration economies are taken into account, Equation 3.8 will be the basic model to be estimated. However, it is possible to consider variations that will provide a better understanding of who actually profits the most from an additional year of life in each type of city. In this sense, the term related to the experience in region r will interact with a measure of the individual skill level.

In addition to static individual fixed effects, it is possible that unobserved worker characteristics influence wage growth. Thus, Equation 3.8 is transformed to include an interaction of an individual fixed effect with experience:

²⁰ The relationship of density and the spatial wage θ_r will be measured as an average value over the whole period of analysis (following DE LA ROCA; PUGA, 2014). This simplification is necessary because time effects are partially captured by experience (which increases over time at the same speed years pass by).

$$w_{i,r,t} = \theta_r + \mu_i + \sum_{j=1}^R \delta_{j,r} \text{exper}_{i,j,t} + \delta_i \text{exper}_{i,j,t} + X'_{i,t} \boldsymbol{\beta} + \varepsilon_{i,r,t} \quad (3.11)$$

In summary, many different alternative models are estimated in this context of dynamic agglomeration economies. Firstly, Equations 3.9 and 3.10 provide an initial evaluation of static agglomeration effects, controlling or not for the process of sorting of workers. Then, Equation 3.11 assesses the importance of dynamic agglomeration effects. These three models generate an estimative of $\hat{\theta}_r$, the region fixed effect that can be defined as a spatial wage (which synthesizes all local characteristics that are relevant for the individual wage). So, this spatial wage is the dependent variable of a second stage,²¹ in which a vector of local characteristics G_r , a time dummy, local area and density²² are the independent variables:

$$\hat{\theta}_r = \lambda_t + \varphi \log \text{density}_r + \log \text{area}_r + G_r \tau + \xi_r \quad (3.12)$$

This formulation is more similar to the one proposed by D'Costa and Overman (2014). They state that De La Roca and Puga (2014) assume that $\mu_i = \delta_i$ to estimate Equation 3.11 directly with an iterative process. Even though De La Roca and Puga (2014) follow the two-stage procedure described above that results in the estimation of Equation 3.12, over here the estimation will be based in only one stage, in order to compare the results directly with the approach described by D'Costa and Overman (2014). Another advantage of this strategy is that it will be possible to measure static and dynamic agglomeration effects at the same level (individual). To do so, Equation 3.12 is substituted into Equations 3.8, 3.9 and 3.11. In the case of Equation 3.11, it becomes the following:

$$w_{i,r,t} = \varphi \log \text{density}_{r,t} + \log \text{area}_r + \mu_i + \sum_{j=1}^R \delta_{j,r} \text{exper}_{i,j,t} + X'_{i,t} \boldsymbol{\beta} + \varepsilon_{i,r,t} \quad (3.13)$$

Alternatively, D'Costa and Overman (2014) obtain an estimation of the urban wage premium by evaluating the effect of agglomeration on wage growth. Before presenting their alternative for Equation 3.11, static effects based on the wage growth are obtained by first-differencing Equation 3.8:

²¹ This approach deals with the problem of bias in standard errors discussed in Chapter 2 and presented by Moulton (1990).

²² Most empirical studies consider employment instead of population as a measure of the agglomeration because it is a better measure of local economic activity and is easily available for many years (COMBES; GOBILLON, 2015). Ciccone and Hall (1996) defend that density should be included in the place of total employment in this regression, because it deals better with the heterogeneity of the spatial size of the geographical units considered.

$$\Delta w_{i,r,t} = \Delta \theta_{r,t} + \sum_{j=1}^R \delta_{j,r} \text{exper}_{i,j,t} - \sum_{j=1}^R \delta_{j,r,t-1} \text{exper}_{i,j,t-1} + \Delta X'_{i,t} \boldsymbol{\beta} + \Delta \varepsilon_{i,r,t} \quad (3.14)$$

When the worker does not move, this equation becomes:

$$\Delta w_{i,r,t} = \delta_{r,r} + \Delta X'_{i,t} \boldsymbol{\beta} + \Delta \varepsilon_{i,r,t} \quad (3.15)$$

On the other hand, for a worker who moves, the expression becomes much more complicated, mixing static and dynamic effects. Equation 3.15 provides an estimation of $\delta_{r,r}$ (being in r and staying in r), which represents the value of one additional year spent at r . This approach is not superior to the direct estimation of Equation 3.8. However, if unobserved workers fixed effects are allowed to affect wage growth, it is possible to estimate them in a way that $\mu_i \neq \delta_i$ (static individual fixed effects are different from dynamic individual fixed effects). In this case, Equation 3.15 becomes:

$$\Delta w_{i,r,t} = \delta_i + \delta_{r,r} + \Delta X'_{i,t} \boldsymbol{\beta} + \Delta \varepsilon_{i,r,t} \quad (3.16)$$

It is worth noting that D'Costa and Overman (2014) only exclude from the sample the observations that refer to the exact year the individual has moved. Therefore, by keeping movers, they are still able to identify city effects for them.

Consequently, when there are dynamic agglomeration effects and the individual unobserved heterogeneity can affect the wage growth path, the preferred strategy will be the one adopted by D'Costa and Overman (2014), which is expressed in Equation 3.16. These results will be compared to Equation 3.13, a restricted version that does not include the possibility of interaction of individual fixed effects with the wage growth path.

3.4 Data

The main database considered here is RAIS-MIGRA (Annual Report of Social Information – Migration, from the Ministry of Labour), which consists of identified registration data of all formal firms and their employees in the Brazilian labour market, focusing on the characteristics of the contract. This database provides longitudinal data for all formally employed individuals in the private sector (or part of the public sector, depending on the type of contract), with a

significant regional disaggregation (municipal level), from 1995 to 2008. The RAIS-MIGRA is comparable to the identified RAIS used in Chapter 2²³.

One of the main advantages of this database is that it is a mandatory report, covering the entire formal sector. Because of this, unlike individual self-reporting surveys, there is a smaller risk of wage under-reporting. Nonetheless, the fact that it only covers the formal sector generates a potential drawback in the analysis, as the informal sector in Brazil is very relevant (RAIS represents less than 30% of the workforce in the initial years of the period analysed, reaching around 35% in 2008). Moreover, the reporting process is supposed to be more accurate in the case of bigger firms, which are usually located in larger cities.

The regional unit of analysis is the REGIC area (Area of Influence of Cities),²⁴ which is a better measure of a labour market area. In fact, this level of analysis is more suitable for the empirical analysis of agglomeration economies, as the estimation is not supposed to be affected by daily commuting (which is very different from moving permanently to another area to look for a job). A few steps were conducted²⁵ in order to achieve a database representing the dynamics of a competitive labour market, in the manufacturing and service sectors. The total number of observations from 1995 to 2008 before balancing the panel is 43,874,819. The balanced panel, excluding individuals with inconsistent age over time, will be composed of 6,749,778 observations (482,127 individuals in 14 years). Due to limitations of computer processing, a sample of 10%²⁶ of this database was generated (48,000 individuals observed over 14 years).

²³ The main difference between these two databases is that the process of selecting valid observations is done by the Ministry of Labour in the case of RAIS-MIGRA, and with steps defined by the author in the case of the identified RAIS.

²⁴ REGIC areas (482) aggregate municipalities based on their interaction, mostly associated with daily commuting flows, transportation links and interconnectedness in general terms (IBGE, 2013). From 1995 to 2008 some new municipalities appeared, requiring an adaptation of the original REGIC areas to encompass these new municipalities according to the municipalities they were originated from. The author can provide details of this aggregation process on demand.

²⁵ There was an initial selection of active contracts in December of each year for male individuals working for private companies in permanent jobs. Then contracts with wages equal to zero, or with less than 20 weekly hours were excluded, as well as contracts with missing information on educational level and occupation. The next step conducted was the selection of individuals aged 18 to 36 years in 1995, with no contradictory age information over time. Finally, only individuals working in the manufacturing and service sectors were kept in the database (sectors 15 to 36, 55 to 74 and 76 to 99, at the 2 digits-CNAE – National Classification of Economic Activities 1.0 classification).

²⁶ These samples are representative for the following characteristics in the initial year of each database: age group (less than 25 years old, 25 to 29 years old, 30 to 36); density of employment in the REGIC area (less than 1, 1 to less than 2, 2 to less than 5, 5 to less than 10, 10 to less than 20, 20 to less than 50, 50 or more workers per km²); firm size (up to 4 employees, 5 to 9, 10 to 19, 20 to 49, 50 to 99, 100 to 249, 250 to 499, 500 to 999, 1,000 employees or more); educational level (illiterate, incomplete primary school, complete primary school to incomplete high school, complete high school to incomplete college, college degree or more); and sector (CNAE 1.0 at the 2 digits level).

One important drawback of the RAIS-MIGRA database is that the classification of occupations is constant over time, and it is not sufficiently disaggregated to be harmonised. Therefore, skills that were considered in Chapter 2 will not be available here, and education attainment will be the only variable capturing the qualification of workers and the complexity of their jobs. In addition, sectors were classified into eight large groups: manufacture; food and accommodation; transport and communication; finance, insurance, pensions and other services; real estate, rents and services to companies; education; health and social services; personal services and other sectors.

Finally, the period of analysis considered for migration models runs from 1995 to 2007 (2008 is omitted as it is just used to define whether the individual has moved to another REGIC area from 2007 to that year). On the other hand, all the estimations regarding static and dynamic agglomeration externalities will be based on the period from 2000 to 2008, and the information from 1995 to 1999 will be used to construct the variable of previous work experience. Table A.9 in the Appendix provides a detailed description of all the variables considered, including methods of calculation and data sources.

3.5 Descriptive statistics

This section will present a basic analysis of the main descriptive statistics of the variables that are relevant for the study of the determinants of a migration decision and for the analysis of static and dynamic agglomeration economies.

3.5.1 Migration decision

The first aspect to be highlighted here is that this analysis will be limited by the fact that the sample covers only individuals who are working in the formal sector. It is not possible to know what happens to an individual if he or she is not reported in the database in a certain year (they can either be unemployed, out of the labour force, working in the public sector or in the informal sector, or may even be employers). It is also not possible to determine the REGIC area in which the individual is living; only the place he or she is working. Therefore, migration decisions covered here will most likely be the ones in which the individual already has a job offer in another location, and they may represent a selected sample of all migratory movements.

Table 3.1 – Descriptive statistics of individuals according to their migration status over the period 1995–2008.

	Non-migrants	1st migration	2nd migration non-return	2nd migration return
ln(hourly wage) - origin	2.64	2.77	3.08	2.74
ln(hourly wage) - destination		2.83	3.14	2.76
Income quantile (origin)				
p25	25.1%	21.0%	19.3%	24.3%
p50	25.0%	21.0%	14.8%	22.1%
p75	25.0%	23.9%	18.5%	23.1%
p100	24.8%	34.0%	47.5%	30.5%
Age	34.2	32.2	34.5	33.6
Tenure in the job (years)	8.4	6.8	5.3	5.8
Education attainment				
Less than 8 years of schooling	34.6%	28.5%	17.3%	29.8%
8 to 10 years of schooling	26.2%	22.9%	17.7%	24.1%
11 to 14 years of schooling	29.5%	32.8%	35.4%	32.7%
15 years of schooling or more	9.7%	15.8%	29.6%	13.5%
Macroregion				
North	1.6%	1.1%	1.8%	1.3%
North-east	9.7%	9.7%	10.2%	10.1%
South-east	65.2%	63.0%	60.7%	67.3%
South	20.3%	22.4%	20.5%	16.3%
Centre-west	3.2%	3.7%	6.8%	5.1%
Sector of activity				
Manufacturing	57.9%	49.0%	34.2%	44.4%
Food and accommodation	2.7%	1.8%	1.5%	2.0%
Transport and communication	10.0%	11.0%	12.6%	11.2%
Finance, insurance, pensions and other services	6.1%	13.9%	21.7%	12.3%
Real estate, rents and services to companies	14.2%	18.4%	25.6%	22.1%
Education	2.0%	1.3%	0.9%	0.5%
Health and social services	3.0%	1.2%	0.9%	1.4%
Personal services and other sectors	4.3%	3.3%	2.6%	6.0%
Firm size				
Up to 4 employees	2.8%	2.0%	2.1%	1.8%
5 to 9	4.8%	3.7%	6.5%	2.7%
10 to 19	5.5%	5.7%	9.7%	4.6%
20 to 49	8.7%	10.3%	13.6%	8.0%
50 to 99	8.4%	9.3%	9.4%	8.0%
100 to 249	14.4%	16.4%	17.3%	15.9%
250 to 499	15.1%	15.4%	13.3%	15.1%
500 to 999	14.9%	15.6%	12.7%	16.8%
1,000 or more employees	25.6%	21.7%	15.5%	27.2%
N	611,693	7,741	1,171	3,389

Source: RAIS-MIGRA.

Following the literature²⁷ and the information available in the database, there is a set of variables that can be considered as explanatory factors for the decision to migrate (initial or return). They are the following: age group, the size of the firm in which the worker is employed in $t - 1$, education attainment,²⁸ employment density at the origin, sector of employment in $t - 1$, macroregion of work in $t - 1$, and number of years living in the origin from 1995 onwards. In

²⁷ De La Roca (2015), Papageorgeou (2014), Freguglia (2014).

²⁸ It would be interesting to include a variable measuring the occupation the individual had in $t - 1$, but the database presents a limitation: from 1995 to 2002, occupations were classified according to the CBO 1994 (Código Brasileiro de Ocupações – Brazilian Code of Occupations), and from 2003 to 2008 they were based on the CBO 2002. Furthermore, it is not possible to make them compatible because codes are disaggregated only at the three-digit level until 2002, and at the four-digit level from 2003 onwards.

addition, wage in $t - 1$ and the quartile of the annual wage distribution in which the individual was positioned in $t - 1$ will be considered as alternatives to education attainment.

As mentioned in the previous section, the REGIC area will be the spatial unit of analysis. In relation to migration, this means that to start working in another REGIC area, the worker has to pay a significant cost. The reason for this is that these areas are supposed to encompass most daily commuting displacements from the municipality of residence to the municipality of work. Therefore, migration is defined by working in a different REGIC area than in the previous year. The main descriptive statistics of the database regarding migration decisions are presented in Table 3.1.

In general terms, first-time migrants receive a higher salary in the origin than non-migrants, are more qualified, more concentrated in the sectors of finance and real estate, and working originally in medium-sized firms. Then, comparing second-time migrants, those who return to their original REGIC area receive a lower salary in the first location (the location after the first movement, 2.74), and end up receiving a lower salary in the destination than second-time migrants that move to a different REGIC area (2.76 against 3.14). Furthermore, the latter are more qualified (almost 30% of individuals with 15 years of schooling or more, while among migrants who return this group represents only 13.5%). Second-time migrants who return usually worked in bigger firms, and were relatively more concentrated in the South-east (as the first place they have moved into). They are also relatively more concentrated in the manufacturing sector.

In addition, following Table 3.2, the distribution of workers according to employment density in the origin and in the destination does not change significantly, except for a decrease in the concentration of workers who live in REGIC areas with a density of 50 or more after they migrate for the first time (from 49.3% to 40.4%). In the opposite direction, second-time migration returning to the original place of work increases the concentration of workers in more dense REGIC areas (from 41.5% to 50.4%). On average, first-time migration takes 4.28 years to happen, while second-time migration happens after 2.53 years for workers who move to a second destination and 1.02 years for those who return to the origin. The average distance travelled to migrate does not change much between each group.

3.5.2 Dynamic agglomeration economies

The estimation of dynamic agglomeration economies will be conducted in one stage, and the main descriptive statistics at the individual level are presented in Table 3.3. In general, average wages and density at the place of work increased in real terms over time. Consequently, wage growth was positive for most of the years analysed. Furthermore, individuals in the sample became more educated over time, almost doubling the percentage of individuals with College degree or more between 2000 and 2008. This sample also concentrates relatively more individuals working in manufacturing, in big firms and in the South-east.

Table 3.2 – Descriptive statistics of the employment density at the origin and destination (1995-2007 when applicable).

	Non-migrants	1st migration	2nd migration non-return	2nd migration return
Density in the origin (employees per km ²)				
Less than 1	1.8%	3.7%	3.3%	2.5%
1 to less than 2	1.6%	2.5%	4.1%	6.0%
2 to less than 5	5.1%	5.7%	8.9%	7.4%
5 to less than 10	8.9%	10.8%	9.6%	9.1%
10 to less than 20	11.7%	9.8%	14.2%	8.9%
20 to less than 50	21.7%	23.1%	20.1%	24.7%
50 or more	49.3%	44.4%	39.9%	41.4%
Density in the destination (employees per km ²)				
Less than 1		3.0%	4.2%	1.8%
1 to less than 2		4.4%	4.3%	1.8%
2 to less than 5		7.5%	8.5%	3.8%
5 to less than 10		9.0%	11.1%	11.0%
10 to less than 20		11.3%	15.0%	10.9%
20 to less than 50		23.8%	22.3%	21.3%
50 or more		41.0%	34.7%	49.4%
Average distance for the migration (in km)		435.7	462.0	456.4
Years before the 1st migration (from 1995 on)		4.29		
Years after the 1st and before the 2nd migration			2.49	0.92
N	611,693	7,741	1,171	3,389

Density is measured in workers/km².

Source: RAIS-MIGRA.

The experience in each city size is calculated in the following way: from 1995 on, every time the individual is observed in a certain city size group, he or she will get one more year of experience in this city group in the following year. As the database is restricted to evaluate individuals between 2000 and 2008, they are supposed to have at the first year a sum of 5 years of experience in different city sizes. The average experience in each city size will grow every year because of the way this variable is calculated.

Furthermore, as the sample is balanced at the individual level, people will get older over time (there are no new individuals entering the database). After some time, the age group comprising 18 to 24 years old loses its share in the sample.

Table 3.3 – Descriptive statistics of the main variables for the regressions of dynamic agglomeration externalities.

	2000	2001	2002	2003	2004	2005	2006	2007	2008
ln(hourly wage)	2.64	2.64	2.63	2.67	2.72	2.75	2.80	2.84	2.87
Wage growth		0.00	-0.01	0.04	0.05	0.03	0.05	0.04	0.03
ln(density)	3.90	3.94	3.99	4.01	4.07	4.12	4.18	4.24	4.29
Experience in cities with density < 5 (in years)	0.53	0.62	0.70	0.78	0.85	0.92	0.99	1.05	1.12
Experience in cities with 5 <= density < 20 (in years)	1.05	1.26	1.47	1.67	1.88	2.08	2.28	2.48	2.67
Experience in cities with density >= 20 (in years)	3.43	4.12	4.83	5.55	6.27	7.00	7.73	8.47	9.21
Education attainment									
Incomplete primary school	35.9%	34.1%	32.7%	31.2%	29.9%	28.5%	27.3%	26.2%	25.0%
Comp. primary school - incompl. middle school	27.1%	26.8%	26.2%	25.6%	25.1%	24.2%	23.8%	23.2%	22.8%
Complete middle school - incomplete college	28.7%	29.9%	31.0%	31.3%	32.5%	33.9%	34.7%	35.4%	36.1%
College degree or more	8.3%	9.2%	10.1%	11.9%	12.6%	13.5%	14.2%	15.2%	16.1%
Age group									
18 to 24	3.6%	1.3%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
25 to 29	21.6%	18.4%	14.6%	10.5%	6.7%	3.6%	1.3%	0.0%	0.0%
30 to 34	32.0%	31.0%	29.2%	26.7%	24.4%	21.5%	18.3%	14.5%	10.3%
35 or more	42.7%	49.3%	56.1%	62.7%	69.0%	74.9%	80.4%	85.4%	89.7%
Firm size									
Up to 4 employees	2.7%	2.7%	2.7%	2.7%	2.8%	2.7%	2.7%	2.6%	2.7%
5 to 9	4.7%	4.8%	4.8%	4.9%	4.8%	4.9%	4.8%	4.8%	4.7%
10 to 19	5.5%	5.5%	5.6%	5.6%	5.5%	5.4%	5.4%	5.5%	5.4%
20 to 49	8.8%	8.7%	8.7%	8.7%	8.5%	8.5%	8.4%	8.2%	8.2%
50 to 99	8.5%	8.8%	8.6%	8.6%	8.4%	8.2%	8.2%	8.0%	7.9%
100 to 249	15.0%	14.8%	14.9%	14.5%	14.0%	13.9%	13.8%	13.5%	13.4%
250 to 499	15.3%	14.9%	14.8%	15.1%	15.0%	14.8%	14.7%	14.2%	14.1%
500 to 999	15.0%	15.9%	15.5%	15.2%	14.9%	15.0%	14.6%	14.3%	14.3%
1,000 or more employees	24.6%	24.0%	24.5%	24.7%	26.2%	26.6%	27.4%	28.8%	29.2%
Sector of activity									
Manufacturing	57.3%	57.2%	57.1%	57.1%	57.4%	57.5%	57.9%	58.1%	57.8%
Food and accommodation	2.7%	2.6%	2.6%	2.6%	2.6%	2.6%	2.6%	2.6%	2.6%
Transport and communication	10.1%	10.1%	10.1%	10.0%	10.1%	10.1%	10.2%	10.2%	10.2%
Finance, insurance, pensions and other services	6.3%	6.3%	6.2%	6.2%	6.2%	6.2%	6.1%	6.3%	6.2%
Real estate, rents and services to companies	14.5%	14.6%	14.8%	14.9%	14.5%	14.5%	14.0%	13.9%	14.0%
Education	1.9%	1.9%	1.9%	2.0%	2.0%	2.0%	2.1%	2.0%	2.1%
Health and social services	3.0%	3.0%	2.9%	2.9%	2.9%	3.0%	3.0%	2.9%	3.0%
Personal services and other sectors	4.3%	4.3%	4.3%	4.3%	4.3%	4.2%	4.2%	4.1%	4.1%
Macro-region									
North	1.6%	1.6%	1.6%	1.6%	1.6%	1.6%	1.6%	1.6%	1.6%
North-east	9.7%	9.8%	9.7%	9.7%	9.7%	9.8%	9.8%	9.8%	9.7%
South-east	65.2%	65.2%	65.3%	65.2%	65.2%	65.1%	65.1%	65.0%	65.0%
South	20.2%	20.1%	20.2%	20.3%	20.3%	20.3%	20.3%	20.4%	20.4%
Centre-West	3.3%	3.3%	3.2%	3.2%	3.2%	3.2%	3.2%	3.3%	3.3%
N	48,000	48,000	48,000	48,000	48,000	48,000	48,000	48,000	48,000

Source: RAIS-MIGRA.

3.6 Results

This section will discuss the main determinants of the migration decision and the effects of static and dynamic agglomeration economies over individual wages.

3.6.1 Migration

This subsection will discuss the main results found in the analysis of selection in initial and return migration and in the measurement of dynamic agglomeration economies. The first set of results refers to the estimation of the determinants of first time migration. As noted previously, the database is composed only of employed individuals, which means that migrants will necessarily have to find a job at the destination. In addition, the variable of first-time migration is defined as follows: equals 0 if the individual has never moved; equals 1 in the year the individual moves. After this first move, the individual is dropped from the database.

Table 3.4 – Logit models for the probability of first-time migration, 1995 to 2007.

(continues)	Model 1 Logit	Model 2 Logit	Model 3 Logit	Model 4 Logit	Model 5 Logit	Model 6 Logit
ln(density in the origin)	0.844*** (0.011)	0.834*** (0.011)	0.826*** (0.011)			
Density in the origin (employees per km ²) - omitted less than 1						
1 to less than 2				0.605*** (0.061)	0.597*** (0.060)	0.589*** (0.059)
2 to less than 5				0.369*** (0.032)	0.372*** (0.032)	0.364*** (0.032)
5 to less than 10				0.430*** (0.034)	0.443*** (0.035)	0.430*** (0.034)
10 to less than 20				0.325*** (0.027)	0.330*** (0.028)	0.318*** (0.027)
20 to less than 50				0.358*** (0.030)	0.354*** (0.029)	0.343*** (0.028)
50 or more				0.307*** (0.027)	0.299*** (0.027)	0.286*** (0.026)
Education attainment - omitted less than 8 years of schooling						
8 to 10 years of schooling	1.164*** (0.038)			1.170*** (0.038)		
11 to 14 years of schooling	1.589*** (0.050)			1.596*** (0.050)		
15 years of schooling or more	2.622*** (0.106)			2.603*** (0.105)		
ln(hourly wage) - origin			1.523*** (0.026)			1.507*** (0.025)
Income quantile (origin) - omitted p25						
p50		1.087** (0.040)			1.085** (0.040)	
p75		1.386*** (0.052)			1.378*** (0.052)	
p100		2.219*** (0.086)			2.180*** (0.085)	
Age	1.053*** (0.019)	1.037** (0.018)	1.030* (0.018)	1.053*** (0.019)	1.038** (0.018)	1.031* (0.018)
Age squared	0.999*** (0.000)	0.999*** (0.000)	0.999** (0.000)	0.999*** (0.000)	0.999*** (0.000)	0.999** (0.000)
Tenure in the job	0.969*** (0.002)	0.960*** (0.002)	0.960*** (0.002)	0.970*** (0.002)	0.961*** (0.002)	0.961*** (0.002)
Macroregion - omitted North						
North-east	2.597*** (0.317)	2.840*** (0.348)	3.103*** (0.381)	3.082*** (0.389)	3.305*** (0.419)	3.593*** (0.456)
South-east	2.832*** (0.336)	2.910*** (0.346)	2.975*** (0.354)	3.306*** (0.406)	3.316*** (0.408)	3.378*** (0.417)
South	2.703*** (0.320)	2.885*** (0.342)	2.915*** (0.346)	3.539*** (0.447)	3.704*** (0.469)	3.742*** (0.475)
Centre-west	1.946*** (0.249)	1.929*** (0.247)	1.966*** (0.252)	2.403*** (0.319)	2.333*** (0.310)	2.379*** (0.317)

(end)	Model 1 Logit	Model 2 Logit	Model 3 Logit	Model 4 Logit	Model 5 Logit	Model 6 Logit
Sector of activity - omitted Manufacturing						
Food and accommodation	1.090 (0.098)	1.240** (0.113)	1.345*** (0.122)	1.075 (0.097)	1.215** (0.110)	1.312*** (0.119)
Transport and communication	1.510*** (0.061)	1.585*** (0.065)	1.564*** (0.063)	1.507*** (0.061)	1.576*** (0.064)	1.554*** (0.063)
Finance, insurance, pensions and other services	2.946*** (0.122)	3.175*** (0.130)	3.058*** (0.126)	2.879*** (0.119)	3.115*** (0.127)	3.005*** (0.124)
Real estate, rents and services to companies	2.121*** (0.076)	2.353*** (0.086)	2.364*** (0.086)	2.105*** (0.076)	2.328*** (0.085)	2.335*** (0.085)
Education	0.702*** (0.072)	0.850 (0.086)	0.847 (0.086)	0.702*** (0.072)	0.847 (0.086)	0.845* (0.086)
Health and social services	0.471*** (0.049)	0.566*** (0.059)	0.565*** (0.059)	0.463*** (0.048)	0.557*** (0.058)	0.556*** (0.058)
Personal services and other sectors	1.004 (0.068)	1.106 (0.074)	1.105 (0.074)	0.987 (0.066)	1.087 (0.073)	1.087 (0.073)
Firm size - omitted up to 4 employees						
5 to 9	1.048 (0.106)	1.025 (0.104)	1.004 (0.102)	1.048 (0.106)	1.025 (0.104)	1.002 (0.102)
10 to 19	1.370*** (0.132)	1.363*** (0.131)	1.319*** (0.127)	1.385*** (0.133)	1.382*** (0.133)	1.339*** (0.129)
20 to 49	1.655*** (0.150)	1.648*** (0.150)	1.575*** (0.143)	1.672*** (0.152)	1.670*** (0.152)	1.597*** (0.145)
50 to 99	1.824*** (0.167)	1.810*** (0.166)	1.713*** (0.157)	1.840*** (0.168)	1.833*** (0.168)	1.737*** (0.159)
100 to 249	1.925*** (0.170)	1.911*** (0.169)	1.784*** (0.158)	1.938*** (0.171)	1.931*** (0.171)	1.806*** (0.160)
250 to 499	1.762*** (0.156)	1.740*** (0.155)	1.618*** (0.144)	1.782*** (0.158)	1.765*** (0.157)	1.645*** (0.147)
500 to 999	1.852*** (0.164)	1.824*** (0.162)	1.698*** (0.151)	1.855*** (0.165)	1.835*** (0.163)	1.712*** (0.153)
1,000 or more employees	0 (0.136)	1.470*** (0.130)	1.378*** (0.122)	1.548*** (0.136)	1.478*** (0.130)	1.387*** (0.123)
Years without migrating (from 1995 on) - omitted 0						
1 year	0.774*** (0.032)	0.795*** (0.032)	0.764*** (0.031)	0.765*** (0.031)	0.783*** (0.032)	0.752*** (0.031)
2 years	0.569*** (0.026)	0.599*** (0.028)	0.559*** (0.026)	0.557*** (0.026)	0.584*** (0.027)	0.545*** (0.025)
3 years	0.456*** (0.023)	0.491*** (0.025)	0.458*** (0.024)	0.441*** (0.023)	0.472*** (0.024)	0.441*** (0.023)
4 years	0.475*** (0.025)	0.527*** (0.028)	0.496*** (0.026)	0.464*** (0.025)	0.511*** (0.027)	0.481*** (0.026)
5 years	0.410*** (0.023)	0.467*** (0.026)	0.430*** (0.024)	0.411*** (0.023)	0.464*** (0.027)	0.428*** (0.025)
6 years	0.410*** (0.024)	0.477*** (0.028)	0.441*** (0.025)	0.404*** (0.024)	0.466*** (0.027)	0.431*** (0.025)
7 years	0.294*** (0.019)	0.353*** (0.023)	0.328*** (0.022)	0.288*** (0.019)	0.342*** (0.023)	0.318*** (0.021)
8 years	0.286*** (0.020)	0.354*** (0.024)	0.324*** (0.022)	0.279*** (0.019)	0.342*** (0.024)	0.314*** (0.022)
9 years	0.279*** (0.020)	0.353*** (0.025)	0.318*** (0.022)	0.271*** (0.019)	0.340*** (0.024)	0.306*** (0.022)
10 years	0.450*** (0.028)	0.584*** (0.037)	0.521*** (0.033)	0.435*** (0.028)	0.559*** (0.036)	0.498*** (0.032)
11 years	0.419*** (0.028)	0.557*** (0.037)	0.488*** (0.032)	0.404*** (0.027)	0.531*** (0.036)	0.465*** (0.031)
12 years	0.352*** (0.025)	0.480*** (0.035)	0.414*** (0.030)	0.344*** (0.025)	0.464*** (0.034)	0.401*** (0.029)
Pseudo R ²	0.0537	0.0530	0.0539	0.0541	0.0532	0.0540
N	564,306	564,306	564,306	564,306	564,306	564,306

*Odds ratios (exponentiated coefficients) are reported for each model. Additional controls: constant, percentage of workers in each of the economic sectors in the whole labour market of each REGIC area. The reference category is stayers, and once individuals move they are dropped from the sample. Density is measured as the number of workers per km². Standard errors are presented in parentheses. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Source: Author's own calculations.

The logit models in column 1 of Table 3.4 indicate that workers with 15 years of schooling or more are 1.62 times more likely to migrate than workers with less than 8 years of schooling. In addition, a 10% increase in the logarithm of density in the origin reduces the probability of outmigration by 0.5%.²⁹ Furthermore, individuals in the South-east are 1.83 times more likely to outmigrate from their original REGIC area than workers in the Northern region of the country. Each additional year of age increases slightly the probability of outmigration (5.3%), while an additional year of tenure at the job decreases this probability by 3.1%. Moreover, workers in the financial and real estate sectors are much more likely to migrate than workers in other sectors.

One final result shown in column 1 of Table 3.4 is that if workers stay in a certain place for longer, the probability that they will outmigrate becomes even lower. Therefore, the chance that a worker who is being observed in a specific REGIC area will migrate after 5 years is 59.0% lower than the chance of a worker who has just arrived deciding to migrate. This finding is also driven by the fact that workers are observed for the first time in 1995 (but they could have been working before that). As time passes by, the remaining individuals in the sample (those who have not migrated yet) will become more and more selected with characteristics of stayers.

In column 2, education attainment is substituted by the position the worker occupied in the distribution of wages of the sample in the previous year. Thus, workers who were in the second quartile had a chance of migration that was 8.7% higher than workers in the lower quartile. Moreover, workers in the upper quartile present a chance of migrating that is 122% higher than workers in the lower quartile. When the individual wage is considered in the place of education attainment (column 3), a very similar result is found – individuals with higher wages will be more likely to outmigrate (a 10% increase in the logarithm of the hourly wage increases the probability of migration in 1.9%). Finally, columns 4 to 6 show that when different density groups are considered (instead of the observed local density), the probability of outmigration decreases for denser cities. The other main results remain unchanged.

Therefore, workers who migrate for the first time seem to be positively selected (higher education attainment, higher initial wage, working in sectors with higher knowledge intensity), but the size of the local labour market decreases the probability of migration, as expected. This means that migration movements in the country are expected to happen from smaller to larger cities, where job opportunities are easier to find and wages are supposedly higher.

²⁹ This value is calculated as $10\% * (0.844 - 1)/e$.

The results presented in Table 3.4 referred to all possible migration movements. However, in Table 3.5 the estimations are measuring the probability that a worker will move to a REGIC area with a density of at least 10 workers per squared kilometre, with an increase in local density from the previous location. This set of models aims to investigate the main factors that increase the chance that a worker will move to a denser urban area.

Table 3.5 – Logit models for the probability of first-time migration towards a REGIC area with a density higher than 10 workers/km² and with an increase in density, 1995 to 2007.

	Model 1 Logit	Model 2 Logit	Model 3 Logit	Model 4 Logit	Model 5 Logit	Model 6 Logit
ln(density in the origin)	0.580*** (0.011)	0.570*** (0.011)	0.563*** (0.011)			
Density in the origin (employees per km ²) - omitted Less than 1						
1 to less than 2				0.453*** (0.057)	0.437*** (0.055)	0.429*** (0.054)
2 to less than 5				0.300*** (0.032)	0.300*** (0.032)	0.291*** (0.031)
5 to less than 10				0.335*** (0.033)	0.342*** (0.034)	0.330*** (0.033)
10 to less than 20				0.286*** (0.029)	0.284*** (0.029)	0.274*** (0.028)
20 to less than 50				0.283*** (0.029)	0.270*** (0.028)	0.260*** (0.027)
50 or more				0.122*** (0.014)	0.115*** (0.014)	0.109*** (0.013)
Education attainment - omitted Less than 8 years of schooling						
8 to 10 years of schooling	1.233*** (0.060)			1.229*** (0.060)		
11 to 14 years of schooling	1.690*** (0.080)			1.677*** (0.079)		
15 years of schooling or more	3.514*** (0.214)			3.415*** (0.207)		
ln(hourly wage) - origin			1.874*** (0.047)			1.831*** (0.046)
Income quantile (origin) - omitted p25						
p50		1.312*** (0.073)			1.278*** (0.071)	
p75		1.973*** (0.112)			1.903*** (0.108)	
p100		3.449*** (0.201)			3.292*** (0.193)	
Pseudo R2	0.0906	0.0929	0.0952	0.0819	0.0837	0.0857
N	559,925	559,925	559,925	559,925	559,925	559,925

*Odd ratios (exponentiated coefficients) are reported for each model. Additional controls: constant, percentage of workers in each of the economic sectors in the whole labour market of each REGIC area, firm size, macroregion, years before migration, sector of activity, age, age squared and tenure at the job, years before migration since 1995. The reference category is stayers, and once individuals move they are dropped from the sample. Individuals who move at any moment to another REGIC area with a density lower than 10 or who move to a REGIC area with a lower density than the previous place of work are excluded from the sample. Density is measured as the number of workers per km². Standard errors are presented in parentheses. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Source: Author's own calculations.

Comparing these results, it is noticeable that local density in the origin becomes even more important as a barrier to new migration movements (now a 10% increase in the logarithm of

employment density will decrease the probability of migration to a large city in 1.5% - which is compared to a decrease of 0.5% in the previous table). This is explained by the fact that workers who are already in large urban areas have a lower incentive to move to a similar REGIC area in terms of the complexity of the local labour market.

Furthermore, the average income in the origin becomes an even more important factor: workers in the upper quartile are 2.45 times more likely to migrate to a dense REGIC area than workers in the lower quartile (column 2). The results presented in column 1 indicate that more educated workers are relatively more likely to migrate to large urban areas, especially when compared to all possible initial migration movements. Noteworthy, more educated workers have a higher propensity to migrate to larger cities than to migrate to any city (their coefficients in Table 3.5 are higher than the ones observed in Table 3.4).

Tables 3.4 and 3.5 provide the conclusion that there is a relevant process of self-selection in first-time migration in the formal labour market in Brazil. This selection becomes even more important when only migration to denser urban areas is taken into account. Therefore, the way incentives are distributed in space makes it more likely that regional inequalities will increase, as cities with a higher percentage of skilled workers will attract more of these high-qualified individuals.

In addition to that, it is noticeable that, in general, migrants are more skilled and have higher wages in the origin than stayers. Therefore, even if a worker is moving to a less dense city, he or she is more likely to be more productive than a similar worker who stayed in the original REGIC area. Migration is a relevant instrument of selection of risk-takers, entrepreneurs, young and skilled workers.

After analysing the determinants of first-time migration, it is possible to identify which are the main characteristics of workers who decide to migrate for a second time. These migrants have the option of returning to the original REGIC area (“return”) or of going to another place (“move on”). Consequently, it is necessary to estimate a multinomial logit with three options: staying in the second place of work, moving to a third place, or returning to the original place of work. Table 3.5 presents a comparison of these events for workers who have already migrated once. Therefore, workers who have moved to a REGIC area during the previous years and decided to stay there afterwards compose the base group.

Table 3.6 – Multinomial logit models for the determinants of second-time migration (for individuals who have already migrated once), 1996 to 2007.

	Model 1		Model 2		Model 3		Model 4	
	Return	Move on Return	Move on Return	Move on Return	Move on Return	Move on Return	Move on	Move on
ln(density in the origin)	0.744*** (0.014)	0.916** (0.033)	0.746*** (0.015)	0.928** (0.033)				
Density in the origin (employees per km ²) - omitted less than 1								
1 to less than 2					2.270*** (0.364)	0.908 (0.216)	2.178*** (0.353)	0.898 (0.214)
2 to less than 5					0.981 (0.155)	0.996 (0.218)	0.956 (0.152)	0.992 (0.219)
5 to less than 10					0.753* (0.118)	0.787 (0.173)	0.738* (0.116)	0.787 (0.174)
10 to less than 20					0.419*** (0.068)	0.748 (0.167)	0.404*** (0.066)	0.738 (0.166)
20 to less than 50					0.566*** (0.092)	0.727 (0.171)	0.534*** (0.088)	0.691 (0.163)
50 or more					0.285*** (0.047)	0.670* (0.163)	0.280*** (0.047)	0.680 (0.166)
Education attainment - omitted less than 8 years of schooling								
8 to 10 years of schooling	0.878** (0.046)	1.159 (0.118)			0.897** (0.047)	1.167 (0.119)		
11 to 14 years of schooling	0.685*** (0.035)	1.357*** (0.126)			0.699*** (0.036)	1.361*** (0.126)		
15 years of schooling or more	0.517*** (0.034)	1.941*** (0.197)			0.509*** (0.034)	1.940*** (0.196)		
ln(hourly wage) _t at the 2nd location			0.490*** (0.019)	0.870** (0.055)			0.492*** (0.019)	0.866** (0.055)
ln(hourly wage) at the 1st location			1.813*** (0.072)	1.712*** (0.111)			1.793*** (0.071)	1.722*** (0.111)
ln(distance in the first migration)			1.070*** (0.022)	1.087** (0.038)			1.059*** (0.022)	1.090** (0.038)
Adjusted R ²	0.0527		0.0629		0.0560		0.0658	
N	31,403		31,353		31,403		31,353	

*Relative risk ratios (exponentiated coefficients) are reported for each model. Additional controls: constant, percentage of workers in each of the economic sectors in the whole labour market of each REGIC area, firm size, macrorregion, years before migration, sector of activity, age, age squared and tenure at the job, years before migration since 1995. The reference category is first-time migrants who become stayers, and once individuals move for the second time, they are dropped from the sample. Individuals who move at any moment to another REGIC area with density lower than 10 or who move to a REGIC area with lower density than the previous place of work are excluded from the sample. Density is measured as the number of workers per km². Standard errors are presented in parentheses. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.
Source: Author's own calculations.

The main results of Table 3.6 are the following: workers who decide to return to their original REGIC areas are less educated than those who decide to stay or to move to another REGIC area (having 15 years of schooling or more decreases the probability of returning in 48.3%). Furthermore, wages in the second location (after the first migration) are negatively associated to the probability of moving again, but less so for moves in direction to a third location. Local density is negatively associated to the probability of leaving a certain REGIC area,

independently of the direction of this movement. Finally, the larger the distance the worker had to cover in the first migration, the higher the probability he or she will move to another place.

Therefore, second migration seems to reinforce the effects of first-time migration. Workers are positively selected to migrate to a third place, but those who decide to return to the original REGIC area are less qualified, with lower salaries in the place they were living after the first migration. A final aspect to be studied regarding selection in initial and return migration refers to the correlation between wages in the origin and migration decisions. It is not possible to discuss causality here, because wages in $t - 1$ are correlated to migration movements in t .

Table 3.7 presents these correlations. In relation to first-time migration (Model 1), workers who have decided to migrate to a dense city (10 or more workers per squared kilometres) used to receive a higher wage in the origin. Even workers who have migrated to a less dense city (less than 10 workers per squared kilometres) received in average a higher wage than stayers. In Model 2, the wage two years before migrating is included, and the coefficients are now supposed to capture the correlation between migration and wage growth. In fact, workers who have decided to migrate to a less dense city had seen a decrease in wages prior to taking this decision, while those who have migrated to a dense city had seen an increase in their wages.

Table 3.7 – Partial correlation of the ln of hourly wages in t-1 and migration decisions in t, 1996 to 2007.

	1st migration		2nd migration	
	Model 1	Model 2	Model 3	Model 4
Migration towards a dense city	0.151*** (0.010)	0.009* (0.005)		
Migration towards a less dense city	0.037*** (0.008)	-0.008* (0.005)		
Migration towards a dense city - return			0.031*** (0.016)	0.007 (0.011)
Migration towards a less dense city - return			-0.196*** (0.016)	-0.029** (0.011)
Migration towards a dense city - no return			0.169*** (0.028)	0.017 (0.015)
Migration towards a less dense city - no return			0.017*** (0.024)	0.005 (0.013)
ln(hourly wage) _{t-2}		0.869*** (0.001)		0.885*** (0.003)
Adjusted R ²	0.5361	0.8869	0.5647	0.9106
N	564,306	516,306	31,403	23,999

* Additional controls: constant, percentage of workers in each of the economic sectors in the whole labour market of each REGIC area, firm size, macrorregion, years before migration, sector of activity, age, age squared and tenure at the job, years before migration since 1995. Individuals who move at any moment to another REGIC area are excluded from the sample. Density is measured as the number of workers per km². Standard errors are presented in parentheses. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Source: Author's own calculations.

Regarding the second migration, the decision to migrate to a dense city is related to higher wages in the previous period, even if workers are returning to their original REGIC area. Furthermore, in the case workers decide to migrate to a third place, their wages in the previous period were higher, and those who decide to go to a dense city were supposed to have the higher original wages.

Even if these results are not indicating any causality, they seem to show that workers who migrate were originally better off in their first location than stayers, and workers who migrate for the second time to a different location are positively selected once again. Migration in the formal labour market in Brazil seems to have a very important role in the sorting of more skilled individuals in large urban areas.³⁰

3.6.2 Dynamic agglomeration economies

The literature review presented in Section 3.2.2 indicates that there is not a well defined strategy to measure dynamic agglomeration economies. Here, two different possibilities will be tested. The first is based on the estimation of Equation 3.13, with a special focus on the effects of work experience in cities over the level of wages. On the other hand, the second approach is based on the investigation of the main factors that affect the wage growth (Equation 3.16).

A first aspect to be highlighted in Table 3.8 is that the process of sorting seems to be extremely relevant to explain what would be otherwise accredited to the urban wage premium. The elasticity of employment density in relation to individual wages decreased from 0.091 to 0.006 when individual fixed effects were taken into account (Models 1 and 2). This is an indication that a significant share of the positive relationship between wages and local density can actually be explained by the migration of self-selected workers to large urban areas, as discussed in the previous subsection.

Moreover, years of experience in each type of city are included in Model 3. It is possible to find positive and significant coefficients for all of them, with a higher coefficient for the years worked in less dense cities (density < 5). When these years of previous experience are combined with the present place of work (Model 4), individuals working in less dense cities who had previous experience in denser areas will benefit the most from these gains.

³⁰ This result indicates that the sorting process should be taken into account whenever individual wages are evaluated to measure agglomeration economies. In Chapter 2, controlling for individual characteristics seems to suffice this requirement.

Table 3.8 – Estimation of the dynamic and static city-size earnings premia (dependent variable is the logarithm of the hourly wage), 2000 to 2008.

	OLS	FE	FE	FE
	Model 1	Model 2	Model 3	Model 4
ln(density)	0.091*** (0.001)	0.006*** (0.001)	0.005*** (0.001)	0.009*** (0.002)
ln(area)	0.027*** (0.001)	0.006*** (0.002)	0.006*** (0.002)	0.004* (0.002)
Experience in cities with density < 5 (in years)			0.034*** (0.001)	0.039*** (0.003)
Experience in cities with density < 5 (in years) ²				0.000* (0.000)
Experience in cities with 5 <= density < 20 (in years)			0.031*** (0.000)	0.028*** (0.001)
Experience in cities with 5 <= density < 20 (in years) ²				0.000*** (0.000)
Experience in cities with density >= 20 (in years)			0.026*** (0.000)	0.029*** (0.001)
Experience in cities with density >= 20 (in years) ²				0.000*** (0.000)
Experience in cities with 5 <= density < 20 (in years)*Now in city with density < 5				0.018*** (0.002)
Experience in cities with 5 <= density < 20 (in years)*Now in city with density >= 20				0.001** (0.000)
Experience in cities with density >= 20 (in years)*Now in city with density < 5				0.008*** (0.002)
Experience in cities with density >= 20 (in years)*Now in city with 5 <= density < 20				0.000 (0.001)
Education attainment - omitted less than 8 years of schooling				
8 to 10 years of schooling	0.198*** (0.002)			
11 to 14 years of schooling	0.595*** (0.002)			
15 years of schooling or more	1.313*** (0.003)			
Region dummies	Yes	Yes	Yes	Yes
Age group dummies	Yes	No	No	No
Firm size dummies	Yes	Yes	Yes	Yes
Sector dummies	Yes	Yes	Yes	Yes
Worker fixed effects	No	Yes	Yes	Yes
R ²	0.5207			
R ² within		0.1375	0.1382	0.1386
R ² between		0.1773	0.1102	0.1474
N	432,000	432,000	432,000	432,000

* All models include a constant term. Density is measured as the number of workers per km². Standard errors are presented in parentheses. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Source: Author's own calculations.

In other words, workers that are now in cities with density lower than 5 and have worked in cities with density higher than 5 and lower than 20 are likely to obtain an average wage increase of 1.8% for each additional year of previous experience. In the same way, if previous experience

was obtained in areas with density higher than 20, one additional year of experience increases wages in cities with density lower than 5 in 0.8%. In conclusion, previous work experience is more valuable in less dense cities, with a higher effect for the years worked in cities with medium density level.

The second set of results regarding the estimation of dynamic agglomeration advantages comprises regression models inspired in Equation 3.16, with the individual wage growth between t and $t - 1$ as the dependent variable and a set of individual characteristics in t as explanatory variables. In this specification, it is possible to control for unobserved worker heterogeneity in relation to dynamic city-size earnings premium (Models 7 and 8 of Table 3.9).

The urban wage growth premium is defined as the relation between local density and individual wage growth. From Model 1 in Table 3.9, there is an indication of convergence of wages between cities according to their density. In fact, as aforementioned, wages are higher in denser cities. Therefore, a negative coefficient for higher density cities in relation to wage growth represents the fact that wages grow faster in less dense cities. The inclusion of additional observed controls at the individual level strengthens this effect (Model 2). However, the estimation with workers fixed effects (Model 3) reduces the significance of the coefficients related to density. Model 4 has the same specification as Model 3, but now the years in which workers have moved are excluded from the sample. Therefore, migrants stay in the sample, but their wage growth is calculated only between years in which they stay in the same place. The main results do not change significantly from Model 3 to Model 4.

Models 5 to 8 follow the same specifications of Models 1 to 4, but they also include previous work experience by city density levels. Focusing the analysis in Model 8, the dynamic wage growth premium obtained from working in other cities is positive and significant, and it is also higher for the experience obtained in denser cities. In this case, the static wage growth premium becomes non-significant. One additional year working in cities with density higher than 20 generates a variation of 0.7% in individual wage growth, while the experience accumulated in cities with density lower than 5 results in a variation of 0.4% in individual wage growth.

This analysis can be complemented by the evaluation of how this previous experience may have heterogeneous effects over wage growth according to the place the individual is currently working. Therefore, in Table 3.10, following the same preferred specification as before (Model 4 of Table 3.9, with workers fixed effects and excluding moving years), previous experience accumulated in cities with density lower than 5 have a positive and significant effect only in

cities with similar density level or cities with high density level (20 or more). In addition, the experience accumulated in cities with medium density level is relevant to explain wage growth only in cities with the same or higher density levels. Finally, the experience accumulated in denser cities has a positive and significant effect over wage growth only in high-density cities. Once again, static agglomeration effects over wage growth become non-significant.

Table 3.9 – Regression for the urban wage growth premium (dependent variable is the individual wage growth in one year), 2001 to 2008.

	OLS Model 1	OLS Model 2	FE Model 3	FE Model 4	OLS Model 5	OLS Model 6	FE Model 7	FE Model 8
Density ≥ 5 & density < 20	-0.003* (0.002)	-0.006*** (0.002)	-0.011* (0.006)	-0.014** (0.007)	-0.011*** (0.003)	-0.012*** (0.003)	-0.008 (0.006)	-0.010 (0.007)
Density ≥ 20	-0.008*** (0.002)	-0.011*** (0.002)	-0.004 (0.007)	-0.002 (0.008)	-0.010*** (0.003)	-0.012*** (0.004)	-0.002 (0.007)	0.003 (0.009)
Experience in cities with density < 5 (in years)					-0.027 (0.038)	-0.036 (0.038)	0.001 (0.001)	0.004*** (0.001)
Experience in cities with $5 \leq$ density < 20 (in years)					-0.026 (0.038)	-0.036 (0.038)	0.004*** (0.001)	0.006*** (0.001)
Experience in cities with density ≥ 20 (in years)					-0.027 (0.038)	-0.036 (0.038)	0.004*** (0.000)	0.007*** (0.000)
Education attainment (incomplete primary school omitted)								
Comp. primary school - incompl. middle school		0.002 (0.001)	-0.006** (0.003)	-0.010*** (0.004)		0.002 (0.001)	-0.006** (0.003)	-0.010*** (0.004)
Complete middle school - incomplete college		0.006*** (0.001)	-0.006* (0.004)	-0.010** (0.004)		0.006*** (0.001)	-0.006* (0.004)	-0.010** (0.004)
College degree or more		0.020*** (0.002)	0.006 (0.005)	0.004 (0.006)		0.020*** (0.002)	0.006 (0.005)	0.004 (0.006)
Age group (18 to 24 omitted)								
25 to 29		-0.039*** (0.010)	-0.035*** (0.012)	-0.039*** (0.012)		-0.039*** (0.010)	-0.035*** (0.012)	-0.040*** (0.012)
30 to 34		-0.056*** (0.010)	-0.043*** (0.012)	-0.047*** (0.012)		-0.056*** (0.010)	-0.043*** (0.012)	-0.047*** (0.012)
35 or more		-0.070*** (0.010)	-0.044*** (0.012)	-0.047*** (0.013)		-0.070*** (0.010)	-0.044*** (0.012)	-0.047*** (0.013)
Firm size dummies	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Region dummies	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Sector dummies	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Year dummies	Yes							
Worker fixed effects	No	No	Yes	Yes	No	No	Yes	Yes
R ²	0.006	0.009	0.007	0.008	0.006	0.009	0.007	0.008
N	384,000	384,000	384,000	329,250	384,000	384,000	384,000	329,250

* All models include a constant term. Wage growth is calculated as $\ln w_t - \ln w_{t-1}$. Density is measured as the number of workers per km². Standard errors are presented in parentheses. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Source: Author's own calculations.

These results indicate that previous experience has a relevant impact on wage growth only in cities with at least the same density level. Therefore, dynamic advantages are expected to be more relevant for migrants who move to denser cities. It is noteworthy that Model 4 of Table 3.10 excludes individual unobserved characteristics that may affect wage growth. Therefore, the process of sorting of high-skilled workers into denser cities is not affecting these results.

Finally, there is no evidence of a pure wage growth effect once controls for previous experience are included in the specification (coefficients for the local density level are not significant in Model 4). This result is in accordance to D'Costa and Overman (2014).

Table 3.10 – Regression for the urban wage growth premium with previous experience and current workplace (dependent variable is the individual wage growth in one year), 2001 to 2008.

	OLS Model 1	OLS Model 2	FE Model 3	FE Model 4
Density ≥ 5 & density < 20	-0.025*** (0.008)	-0.028*** (0.008)	0.007 (0.033)	-0.044 (0.050)
Density ≥ 20	-0.029*** (0.007)	-0.030*** (0.007)	-0.011 (0.035)	-0.059 (0.057)
Experience in cities with density < 5 (in years)*Now in city with density < 5	-0.029 (0.038)	-0.039 (0.038)	0.001 (0.001)	0.004*** (0.001)
Exp. in cities with density < 5 (in years)*Now in city with $5 \leq$ density < 20	-0.026 (0.038)	-0.036 (0.038)	0.000 (0.003)	0.008 (0.005)
Exp. in cities with density < 5 (in years)*Now in city with density ≥ 20	-0.025 (0.038)	-0.034 (0.038)	0.008** (0.004)	0.014** (0.006)
Exp. in cities with $5 \leq$ density < 20 (in years)*Now in city with density < 5	-0.022 (0.038)	-0.032 (0.038)	0.013*** (0.004)	0.009 (0.006)
Exp. in cities with $5 \leq$ density < 20 (in years)*Now in city with $5 \leq$ density < 20	-0.026 (0.038)	-0.036 (0.038)	0.004*** (0.001)	0.006*** (0.001)
Exp. in cities with $5 \leq$ density < 20 (in years)*Now in city with density ≥ 20	-0.025 (0.038)	-0.036 (0.038)	0.006*** (0.002)	0.009** (0.003)
Exp. in cities with density ≥ 20 (in years)*Now in city with density < 5	-0.027 (0.038)	-0.037 (0.038)	0.006 (0.004)	0.005 (0.006)
Exp. in cities with density ≥ 20 (in years)*Now in city with $5 \leq$ density < 20	-0.027 (0.038)	-0.037 (0.038)	0.002 (0.002)	0.003 (0.004)
Exp. in cities with density ≥ 20 (in years)*Now in city with density ≥ 20	-0.026 (0.038)	-0.037 (0.038)	0.004*** (0.000)	0.007*** (0.000)
Education attainment (incomplete primary school omitted)				
Comp. primary school - incompl. middle school		0.002 (0.001)	-0.006** (0.003)	-0.010*** (0.004)
Complete middle school - incomplete college		0.006*** (0.001)	-0.006* (0.004)	-0.010** (0.004)
College degree or more		0.020*** (0.002)	0.005 (0.005)	0.004 (0.006)
Age group	No	Yes	Yes	Yes
Firm size dummies	No	Yes	Yes	Yes
Region dummies	No	Yes	Yes	Yes
Sector dummies	No	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes
Worker fixed effects	No	No	Yes	Yes
R ²	0.0063	0.0089		
R ² within			0.0073	0.0082
R ² between			0.0055	0.0013
N	384,000	384,000	384,000	329,250

* All models include a constant term. Wage growth is calculated as $\ln w_t - \ln w_{t-1}$. Density is measured as the number of workers per km². Standard errors are presented in parentheses. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Source: Author's own calculations.

In conclusion, individual wage level is expected to be more affected by previous work experience in less dense cities. However, future wage growth will be higher for individuals who migrate to higher density cities or for those who stay in cities of at least the same density group. This result indicates that apart from the sorting process, workers who move to bigger cities are expected to obtain higher salaries in the future.

3.7 Conclusion

This chapter has aimed to investigate static and dynamic advantages of working in agglomerated areas. The literature on agglomeration economies highlight the importance of the process of sorting of high-skilled workers into larger cities, which is usually captured by the inclusion of individual fixed effects. There are two main problems with this strategy (GROOT; DE GROOT; SMIT, 2014a): this sorting process becomes a black box (it is not actually explained or understood), and the variation to estimate agglomeration effects will come solely from migrants.

Therefore, following the most recent literature, a migration model was estimated for Brazil with data from RAIS-MIGRA. The main conclusions from this analysis were that high-educated workers are more likely to migrate to work in another place, as well as individuals living in less dense cities. Furthermore, tenure at work and the time working in a certain REGIC area reduce the probability of outmigration. Therefore, the way incentives are distributed in space makes it more likely that regional inequalities will increase, as cities with a higher percentage of skilled workers will attract more of these high-qualified individuals. In addition, second migration seems to reinforce the effects of first-time migration. Workers are positively selected to migrate to a third place, but those who decide to return to the original REGIC area are less qualified, with lower salaries before the second migration in comparison to individuals who decide to stay).

Finally, the analysis of the correlations of wages in the origin in $t - 1$ with the migration decision in t indicate that workers who have decided to migrate to a dense city in first-time migration (10 or more workers per squared kilometres) used to receive a higher wage in the origin. Even workers who have migrated to a less dense city (less than 10 workers per squared kilometres) received in average a higher wage than stayers did.

These results provide a sign that workers self-select into more agglomerated areas, and this process happens both in first as well as second-time migration decisions. In addition to that, the

estimation of static agglomeration economies indicate that the inclusion of workers fixed effects capture most of the effect of local density over wages. Therefore, the sorting process explains a large share of the urban wage premium.

The inclusion of previous work experience in different cities aims to capture dynamic agglomeration advantages. In this setting, static agglomeration advantages are no longer significant. The main conclusion is that when years of previous experience are combined with the current place of work, individuals working in less dense cities who had previous experience in denser areas will benefit the most from these gains.

Moreover, the estimation of dynamic agglomeration advantages controlling for the possible worker heterogeneity in these gains indicates that previous experience has a relevant impact on wage growth only in cities with at least the same density level of the current place of work. Therefore, dynamic advantages are expected to be more relevant for migrants who move to denser cities.

In conclusion, workers who move to bigger cities are expected to obtain higher salaries in the future. This effect is supposedly net of the sorting process, as individuals fixed effects are controlled for. Combining these results with the conclusions from migration models, large cities attract high-skilled workers and provide them the conditions to obtain higher wage growth over time. Consequently, urban agglomeration economies are likely to increase wage inequality between cities. This process reinforces the attraction of more qualified workers to large urban areas, which will increase local productivity in these places.

Even though this movement towards spatial concentration of skilled workers could be expected to increase congestion costs in big cities, two other processes are supposed to happen at the same time. First, as noted in the main migration results, less skilled workers are more likely to return to their original REGIC areas, creating an outmigration flow. Furthermore, cities are reinventing themselves to bear with a higher demand for public services and land. Cities can become more productive, as it has already happened over time. Maybe new ways of organising work relations in space and provide public services with higher productivity could be possible solutions to face the increasing human flows to large cities.

However, despite all these factors, raising life cost and congestion still seems to be an issue in large cities. Therefore, public policy should aim to make cities smarter, reorganizing incentives for the locational choice of firms and workers, and adopting new technologies to increase sustainability of the urban life in different dimensions.

4. A COMPREHENSIVE ANALYSIS OF THE WAGE CURVE IN BRAZIL: NON-LINEARITIES, URBAN SIZE, AND THE SPATIAL DIMENSION

4.1 Introduction

The wage curve represents an empirical negative relation between wages and local unemployment rates at the regional level (BLANCHFLOWER; OSWALD, 1994). This result is expected, as wage flexibility plays a relevant role in adjusting local economies to adverse shocks. Moreover, causality is expected to flow from unemployment size to the level of wages (BLANCHFLOWER; OSWALD, 1995). Many empirical studies have corroborated the original formulation of the wage curve by finding negative and significant relationships in different environments.

Within the context of a developing country, with an expressive informal sector and significant heterogeneity in terms of population density, the wage curve can be an expression of the degree of flexibility in the labour market, given the relative bargaining power of workers and firms and the different costs faced by firms in each sector. City size is one additional factor in this equation, by increasing the number of opportunities for a matching between firms and workers. The investigation of this subject provides an opportunity to further understand how urbanisation economies affect local labour markets, as different city sizes can imply that firms and workers may have more or less precautionary behaviour.

The higher sectorial diversity usually found in bigger cities is one of the factors that contributes to these special conditions, by accommodating uncorrelated random sector-specific shocks and reducing uncertainty (GORDON; MCCANN, 2000). Furthermore, urban areas concentrate diverse and complementary economic activities under intense competition (GLAESER, 2011). In this setting, the interaction of economic agents provides different locational advantages, among which it is possible to highlight the higher probability of a successful matching in the labour market (MARSHALL, 1890). Therefore, not only is the local economic environment supposed to be more resilient to economic shocks, but there is also a broader range of opportunities for firms and workers to become adjusted to adverse economic situations.

Moreover, in the context of developing countries, an additional element becomes relevant, namely the existence of an informal sector and its consequences in labour market outcomes (FIELDS, 2005). In this sense, exploring different aspects of the wage curve for a developing country such as Brazil would provide the opportunity to understand additional effects of

agglomeration economies in the labour market and their impacts on different labour market arrangements.

This analysis contributes to the discussion developed in the previous chapters by providing a comparison of the formal and the informal sectors in Brazil, and how their interaction is affected by the intensity of agglomeration. Furthermore, different aspects of the wage curve in Brazil are investigated, with varying databases, regional aggregation, time periods, and data structure. Finally, an analytical framework based on different bargaining power levels for workers in the formal and the informal sector is proposed, with a direct implication of urban size over the relative power of workers and firms.

This chapter is organised as follows: Section 4.2 presents a literature review on the wage curve, Section 4.3 explores the empirical literature of the wage curve in different countries and in Brazil in particular, and Section 4.4 discusses the literature on labour market segmentation and proposes a theoretical framework of analysis. Furthermore, the empirical strategy is presented in Section 4.5, Section 4.6 provides a description of the databases and the main results, and Section 4.7 presents the main conclusions of the chapter.

4.2 The wage curve

The wage curve represents an empirical relation between the level of wages and the unemployment rate in different regions. Considering unemployment on the x-axis and wages on the y-axis, this relation is characterised by a downward sloping curve (BLANCHFLOWER; OSWALD, 1995, 2005). A similar concept but with relevant differences is the Phillips curve, which is concerned with the relationship between the level of unemployment and wage growth. Moreover, while the former has microeconomic foundations and can be discussed in a regional context, the latter is usually considered within a macroeconomic perspective.

A theoretical challenge associated with the wage curve is its regularity between countries and periods, despite the significant institutional differences among them. The wage curve can be interpreted as a measure of the flexibility of the labour market (BLANCHFLOWER; OSWALD, 1995), and several mechanisms have been proposed to explain it, based on imperfect competitive structures:

- Bargaining power: a higher unemployment rate can reduce the bargaining power of workers, as it may be harder for them to find another job in a depressed economic environment. On the other hand, a lower unemployment rate means that firms will have more

difficulty finding a replacement that matches their needs. In the first case, workers may lose their ability to demand a higher share of the surplus, while just the opposite occurs in the second case, resulting in higher wages;

- Efficiency wage: firms pay a certain wage aiming to increase labour productivity and workers choose the effort expended at work. In this context, wages influence productivity, but when the unemployment rate is higher, shirking may have a greater penalty if it takes a long time to find a job; consequently, workers will put more effort in the job, even if the payment is low. Then, according to Card (1995), group-specific unemployment rates should be more relevant (for instance, for individuals with different skill levels). Moreover, high and low-amenity regions will be located in different parts of the wage curve;
- Labour contracts: considering the maximisation of the joint welfare of employers and employees, unemployed workers contribute less to the surplus of this combination. Therefore, with a higher wage, it will be interesting to have more individuals employed (leading to a lower unemployment rate);
- Turnover costs: labour markets are perceived as ‘thin’ by workers and the wage curve represents the degree of monopsonistic power of firms, which is generated either by turnover costs or horizontal job differentiation based on geographic location. Then, instead of the perfectly elastic supply curve of the competitive model, firms face an upward-sloping supply curve in the local labour market (MANNING, 2003).

Some of these mechanisms have not necessarily become structured theoretical models, and most importantly, they have not been empirically validated (NIJKAMP; POOT, 2005). An exception to this pattern is the model by Campbell and Orszag (1998) based on the monopsonistic competition structure generated by turnover costs, further confirmed empirically by Morrison, Papps and Poot (2006). Alternatively, Sato (2000) provided a simple search and matching framework combined with local productivity differentials. Their main conclusion was that regions with higher productivity would have higher wages and lower unemployment rates in equilibrium. These regions also present a higher concentration of people, and consequently, significant congestion costs that ensure the persistence of spatial inequalities.

The wage curve is observed by comparing local labour markets. In this respect, there are two possible ways to see the interaction of spatial units and urban size: on the one hand, neighbouring areas can increase accessibility to job offers; on the other hand, agglomeration size may increase local opportunities and influence the relationship of wages and the unemployment rate. Accessibility to jobs can affect the balance of the bargaining power of

workers and employers (LONGHI; NIJKAMP; POOT, 2006), generating a positive effect over the local wage level (wages offered in regions with higher accessibility to jobs should be higher). In addition, the unemployment elasticity to pay is expected to be negatively affected by a higher level of accessibility, meaning that more integrated regions are supposed to show a lower elasticity.

Furthermore, as discussed in the previous chapters, agglomeration economies may increase local productivity, leading to higher wages and lower unemployment rates (Sato, 2000). Therefore, they may be essential to determine the location of different local labour markets in the wage curve. It is also possible that local labour markets are subject to common shocks, with spatially autocorrelated residuals, and spatial proximity can lead to a direct negative effect of unemployment in the neighbouring areas over local wages (LONGHI, NIJKAMP; POOT, 2006).

With these considerations in mind, space and agglomeration size are essential elements in the analysis of the wage curve. The effects of agglomeration are not restricted to the size of the labour market, but also to higher local productivity, which result from the different mechanisms discussed in the introductory chapter. These effects can be heterogeneous depending on the degree of formalisation of work relationships. Therefore, formal and informal workers may have their bargaining power affected by local market size and higher local productivity to a different extent. This issue will be further explored in the following sections.

4.3 Empirical literature

The wage curve has been investigated in the literature as a measure of the flexibility of the labour market in relation to economic downturns, with the unemployment rate synthetizing the business cycle. As mentioned before, many studies have found a significant and negative relationship between wages and the unemployment rate, which is usually explained by market imperfections in the labour market. However, in developing countries there is an additional dimension to be taken into account: the labour market is partially comprised by an informal sector, with more flexible work arrangements.

The empirical literature has focused on the analysis of this relationship in a comparison between countries or regions, controlling for individual characteristics and other microeconomic aspects. Nijkamp and Poot (2005) provided a meta-analysis of the coefficients estimated in more than 200 empirical studies, finding that there is a recurrent negative relationship between wages and

the local unemployment rate. Correcting for publication and aggregation bias, this elasticity is estimated to be around -0.07.

The most recent advances in the empirical literature have involved a stronger effort to control for worker heterogeneity based on observed and unobserved characteristics. This was made possible by the recent spur of surveys with a longitudinal structure at the individual level, as in Baltagi, Blien and Wolf (2009, 2012) for Germany, and Turunen (1998) for the United States. The latter disaggregated the wage curve by year and worker characteristics, finding very different coefficients depending on the group analysed. In addition, there has been an effort in the literature to control for the possible endogeneity of the unemployment rate (instrumented by its time lag).

The inclusion of agglomeration effects and an explicit spatial dimension has followed the most recent advances in spatial econometrics techniques (BALTAGI; BLIEN; WOLF, 2012; BALTAGI; ROKICKI, 2014b; ELHORST; BLIEN; WOLF, 2007; LONGHI; NIJKAMP; POOT, 2006). Most studies have found that unemployment in surrounding areas also affects the equilibrium in the local labour market, as workers are supposedly aware of job opportunities in these areas. In the absence of spatial spillovers, the local unemployment elasticity is likely to be overestimated (BALTAGI; ROKICKI, 2014b). Longhi, Nijkamp and Poot (2006) noted that the unemployment elasticity is lower in regions that interact more with other regions, meaning that more isolated areas have a less flexible labour market. Similar studies were conducted by Ramos, Nicodemo and Sanromá. (2015) for Spain, Fingleton and Palombi (2013) in the UK, Baltagi and Rokicki (2014b) for Poland, among others.

As mentioned above, in the context of developing countries, usually a large share of the labour market is unregulated, based on informal relations and does not follow the legislation on the matter. This has important implications for the bargaining relationships in the labour market, as the informal sector represents an additional status apart from formal jobs and unemployment. Therefore, economic shocks may be absorbed in this intermediate sector, with wages varying significantly given a change in the local unemployment rate. In Colombia, Ramos, Duque and Surinach (2010) found that the elasticity of individual wages to the unemployment rate is much higher in the case of informal workers. Baltagi, Baskaya and Hulagu (2013), who estimated different wage curves for the formal and the informal sectors in Turkey, provided a similar result. When controlling for the potential endogeneity of unemployment (FE-2SLS with the lagged value of the unemployment rate as an instrument), they found a coefficient of -0.071 for

the formal sector and -0.263 for the informal sector (and -0.107 when both were considered together). Higher wage flexibility was also found for women, young individuals, with low tenure and low education.

Labour market dualism and wage differentials between workers in the formal and informal sectors are very relevant for the degree of wage flexibility in these sectors. However, Menezes-Filho, Mendes and Almeida (2004) showed that when individual observed and unobserved characteristics are taken into account, most of the differences between wages in the formal and the informal sectors in the Brazilian labour market disappear. Controlling for these determinants, the wage level in the informal sector is actually higher than the one in the formal sector (probably as a compensation for providing less benefits). Therefore, there is an indication of a self-selection process of more qualified workers in the formal sector, and any analysis trying to assess labour market outcomes under the perspective of sector dualism should control for it.

Regarding the interaction of the formal and informal sectors, many empirical and theoretical studies have discussed the existence of market segmentation, which can be defined as the existence of a wage gap between jobs in these two sectors. In this sense, controlling for worker, sector and firm observed heterogeneity, a job in the formal sector is supposed to pay a higher salary. With PME data on a more recent period (2003-2008), Corseuil and Foguel (2012) estimated a model for the Brazilian labour market with a similar structure to a wage curve (controlling for individual fixed effects), but included an interaction of the unemployment rate with a dummy variable identifying whether the job is formal or informal. Their interpretation for this result was that the unemployment rate controls for the business cycle and its effect over individual wages varies according to the moment of the economy (expansion or retraction). In an expansion, they find that the formality rate goes up and the wage gap between formal and informal workers increases. Botelho and Ponczek (2011) interpreted this wage gap as a segmentation phenomenon, strongly associated with the strict regulation over certain aspects of labour contracts in Brazil.

Focusing again on the wage curve, most studies in Brazil have focused on a comparison of the level of wage flexibility in the formal and the informal sectors, as well as between rural and urban areas. Barros and Mendonça (1997) applied this methodology to a pseudo-panel combining gender, age groups, educational levels, and metropolitan regions with data from the Monthly Employment Survey (PME-IBGE), from 1982 to 1994. Their conclusion was that the degree of wage flexibility in Brazil was considerably high during that period (a slope of -5),

similar to the one observed in industrialised economies, but varied significantly over time (affected by the economic instability). They also found that a lower wage flexibility for more qualified workers, a similar result found by Reis (2006) – who considered state-level data from 1990 to 1999 from the National Household Survey (PNAD). For a similar period (1989-1993), with data from PNAD, but evaluating the wage curve on macroeconomic terms (the ratio between the growth of real wages and the growth of the unemployment rate), Amadeo and Camargo (1997) found a slope of -10. When adding informal employment variation to unemployment variation, this slope fell to -4.5.

Wage determination in Brazil has changed significantly over the past few decades. According to Carneiro and Henley (1998), at the beginning of the 1980s, the State provided a centralised wage indexation, weakening labour unions. Between the mid-1980s and the beginning of the 1990s, in spite of an increasing spur on the role of unions, extremely high inflation rates led the State to keep pursuing significant control over wage indexation. Finally, in the mid-1990s unions were able to bargain for higher real wages, developing a collective bargaining structure more similar to the one observed in Western Europe. These developments had important implications on the flexibility of wages to economic downturns over time.

Still in the Brazilian context, Souza and Machado (2004) used PNAD data from 1981 to 1999, following a two-stage estimation strategy (individual data in the first stage and macroregion-year data in the second). Their findings suggested that labour markets in urban areas (-0.23) are much more flexible than those in rural areas (-0.06). Silva and Monsueto (2012) considered a similar framework for the period 2002-2009, also with PNAD data. They found an elasticity of wages to unemployment of -0.0474, with an indication that low-education segments are more flexible.

Santolin and Antigo (2009) provided an evaluation of a dynamic wage curve considering averages among wage groups, gender, race, and schooling years, in six metropolitan regions from 1997 to 2005 (with PNAD data). According to them, wage flexibility appeared to be higher for the whole group of workers (-0.15) than for only formal workers (-0.05). Furthermore, the coefficient of the lagged wage was close to zero, indicating that unemployment rates seem to be more related to the level of wages than to their change over time.

A more recent piece of evidence on the wage curve relation in Brazil was presented by Baltagi Rokicki and Souza (2014a), analysing PNAD data from 2002 to 2009. Their estimations

controlled for individual observed heterogeneity and state-level fixed effects, instrumenting the unemployment rate by its lag. When the formal and the informal sectors were considered separately, only the latter presented significant results (-0.251, while the coefficient was -0.08 when both sectors were considered together). The authors also compared workers in the formal and the informal sectors with similar probabilities of being formal (below the median), finding that the elasticity of unemployment reached -0.129 and -0.305 respectively. Furthermore, they also found that men were more responsive to the unemployment rate (-0.13), while in the case of women the coefficient was not significant (women were more risk-averse, looking for more stable jobs).

Summing up the results found in the literature, there is an indication of the existence of a wage curve in Brazil, with different degrees of wage flexibility according to educational level, age, gender, size of the labour market and formalisation of the job. Furthermore, over time wage flexibility has decreased, accompanied by an increase in economic predictability and macroeconomic stability of the country, the formalisation of the labour market and an increase of the educational level of the labour force. However, there is room to explore further aspects of the wage curve, in special spatial dependence and agglomeration economies, the control of individual unobserved heterogeneity, and controlling for a possible selection bias of the participation in the labour market.

4.4 Labour market dualism in Brazil and a theoretical framework

There are sharp differences in the way individuals interact in the labour market of developing economies according to sector, location, firm size, educational level, among other characteristics. In this context, the literature on wage dualism has investigated whether similar workers in different sectors (formal and informal) receive different payments (SOARES, 2004). Labour market segmentation can be generated by different mechanisms, among which there are: i) firms' heterogeneity (technology, market power, size, sector), and ii) institutional aspects (unions and/or labour legislation). The latter may lead to segmentation if it imposes effective restrictions on the labour market and if there is an inefficient structure of enforcement to follow these rules, applied only to a share of the agents (PERO, 1992; ULYSSEA, 2006).

According to Estevão and Carvalho Filho (2012), the main characteristics of the formal sector in Brazil were originally established by the Labour Law Consolidation (Consolidação das Leis do Trabalho – CLT), a set of labour laws from 1943. Over the years, some changes have been

made to this legislation, especially with the new Constitution, in 1988. A job according to these regulations is characterised in the following way: every worker has a booklet (work card) recording his or her work history; workers and firms must contribute to the social security system and respect certain working conditions related to minimum wages, maximum hours of work and payment for overtime work, pre-paid vacations, pre-paid maternity leave, pre-notification of firing, costs to unjustified demission. Public servants and military are not usually hired under the CLT, but their working relations are based on a contract signed with the government that guarantees similar rights, with more stability.

On the opposite side, informal jobs do not have the benefits and guarantees mentioned above, and can be defined as legal but unregulated activities (ULYSSEA, 2010). Their existence is explained by the significant cost of hiring a formal worker (mostly related to social security contributions) and for the relevant costs of formalising an enterprise. There are different types of informal jobs: high-skilled and low-skilled self-employment (for example, consultants in the first case, and street hawkers in the second case), and individuals hired under informal relations by regulated or unregulated firms. When there is no information on contracts, social security status can be an indicator of informality (HENLEY; REZA ARABSHEIBANI; CARNEIRO, 2009). Another possible characterisation is that it is more difficult to achieve a matching between firms and workers in the informal sector due to the lack of a clear channel to exchange information between economic agents and due to high search costs (EL BADAOU; STROBL; WALSH, 2010).

Agénor (2007) identified three main segments that characterise labour markets in developing countries: rural sector (self-employed individuals and unpaid family workers), informal urban sector (low-skilled self-employed persons and small enterprises, mainly providers of services and other non-tradable goods, with limited access to credit), and formal urban sector (medium and large enterprises producing tradable and non-tradable goods, employing workers with different skills). The informal sector can involve paid labour without a formal contract, subject to high job insecurity, underemployment and low productivity, a few benefits, and high flexibility. On the other hand, employers in the formal sector must provide a number of benefits, and wages can be highly affected by the minimum wage legislation, unions and productivity considerations.

The author continued describing the main institutions and regulations in the labour market of developing countries, which are usually relevant in the differentiation between the formal and the informal sectors. Among these, it is possible to highlight (AGÉNOR, 2007):

- Minimum wage laws: their effects on the labour market will depend on the degree of law enforcement and the frequency of adjustment (encompassing real gains or losses in an inflationary context). Minimum wages may represent a lower bound for wages in the formal sector, mostly affecting payments of unskilled workers.
- Hiring and firing regulations: this sort of legislation aims to protect workers hired in the formal sector, including restrictions for firms to fire workers without a ‘proper reason’, the requirement of a notification period before dismissal, large severance arrangements, and complicated administrative procedures that increase the cost of layoffs and plant closures.
- Nonwage costs and unemployment benefits: the former includes social security contributions and nonwage benefits (housing, health care, pensions, subsidised transportation, subsidised meals, and family allowances), whose cost is partially shared by firms with workers through lower wages. The cost of unemployment benefits is also shared, and may lead to an increase in the level and the duration of unemployment, apart from discouraging human capital accumulation.
- Indexation practices: may induce a high degree of real wage rigidity, combined with a reduction of the effect of monetary shocks over output and employment. They may also distort the expected effect of policy-induced changes in relative prices. Finally, indexed wages can lead to inflation persistence. Wage rigidity in one sector may coexist with full wage flexibility in another sector – leading to market segmentation.
- Bargaining structures and trade unions: wage bargaining arrangements can be related to centralised collective bargaining, with significant government intervention, possibly resulting in wage segmentation. Trade unions are usually tied with political parties and the state to obtain legislative protection and redistributive policies. If the collective bargaining process is less concentrated, the local labour market will be more flexible.

Therefore, firms aiming to hire a worker respecting labour regulation face a higher cost, in terms of contributions as well as the cost of dismissal. Their adjustment capacity is reduced under adverse economic situations, implying that wage sensitivity to a higher unemployment rate will be lower. Being in the informal sector has the advantages of lowering hiring costs. However, there are some drawbacks related to the risk of being caught, especially for firms that

are not respecting labour legislation and self-employed individuals who are not paying the appropriate taxes. Informal labour relations may also lead to difficulties in obtaining credit and participating in certain governmental programmes. Large firms are more likely to get caught for not paying taxes, which implies that the formal sector will usually be associated with large firms (EL BADAQUI; STROBL; WALSH, 2010).

In relation to the empirical literature investigating labour market dualism, earlier findings did not include all relevant observed and unobserved heterogeneity and possible selection biases in an appropriate manner (SOARES, 2004). Controlling for worker heterogeneity and firm size seems to tackle most of the formal sector net wage premium in relation to the informal sector (EL BADAQUI; STROBL; WALSH, 2010; MENEZES-FILHO; MENDES; ALMEIDA, 2004). Therefore, whenever possible it is important to deploy methods involving selection models, longitudinal databases and instrumental variables, in addition to other techniques that may enhance the control for endogeneity issues and confounding factors.

The characteristics of the Brazilian labour market are typical of a developing country, with significant unemployment and wage dualism (FIELDS, 2005). These two features are essential when modelling the economic relations in the labour market. There is also a certain degree of heterogeneity between jobs in the informal sector, with a special distinction for jobs based on self-employment and wage employment (FIELDS, 1990, 2005). Furthermore, self-employed workers can be divided between low and high-skilled, and it can be hard to distinguish them in the labour market, as both usually provide services in various activity sectors. Depending on the self-employment category, it may even be preferred to a formal job. As a conclusion, the focus here will be on salaried workers³¹, and the main difference between formal and informal workers will rely on the existence or not of a contract between workers and firms. While the self-employment category is not considered here, this is not a relevant drawback of this study, as the informal sector can be viewed also as an unregulated microentrepreneurial sector associated with voluntary self-employment (MALONEY, 2004).

The fact that labour markets in developing economies are supposedly segmented, with particular characteristics, imply that their relationships of employment and unemployment are not necessarily the same as the ones observed in advanced economies (AGÉNOR, 2007).

³¹ Another specificity of self-employed individuals is that they possibly have a stronger entrepreneurial spirit, something which could differentiate them from wage employees.

Hence, the wage curve relationship may not be similar to what has been found in previous studies for developed economies.

The model discussed in the following section aims to represent the relationships in local labour markets that generate the wage curve under potential labour market dualism. It allows for distinct flexibility degrees observed in different sectors in a dual economy of a developing country. This is achieved by combining turnover costs in an efficiency wage model and a varying bargaining degree. These two mechanisms have an impact in the observed relationship between wages, productivity and unemployment (BOOTH, 2014).

4.4.1 Theoretical framework: institutional costs, bargaining power and efficiency wages

The previous subsection showed that firms in the formal sector may be subject to “labour codes” that regulate hiring and firing, and impose higher costs to them in terms of higher taxes and benefits that must be paid to workers (FIELDS, 2005). Under these conditions, wages may not adjust completely to the business cycle. The main characteristics of dual labour markets indicate that bargaining power can vary significantly between these sectors. This may be generated by a higher presence of unions in the formal sector, as well as higher turnover costs that can be found in this sector. There is also the risk of a shortage of high-skilled workers necessary for specific tasks in the formal jobs sector, which decrease the propensity of dismissal by firms in this sector. Therefore, one possible way of modelling these differences is to consider that workers in the informal sector have less bargaining power than workers in the formal sector (BLIEN et al., 2013, consider differences in the bargaining power as the key to explain the wage curve observed for different regimes of collective bargaining).

The model developed by Barth et al. (2002) is adapted here to the context of a developing country with a formal and an informal sector. This model combines two mechanisms that generate the relationship of the wage curve: efficiency wage and wage bargaining. The efficiency wage model encompass endogenous turnover costs (CAMPBELL; ORSZAG, 1998), depicted in Equation 4.1, which are dependent on the probability that the worker quits (negatively related to the unemployment rate). Therefore, under higher unemployment rates, turnover costs are supposed to be lower. Replacement costs per worker (C) are given by the multiplication of hiring costs (H) and the probability that the worker quits (q):

$$C_r = H_r q_r \quad (4.1)$$

Extending the model proposed by Barth et al. (2002) from a single point economy to a spatial setting, migration is possible and quitting is affected by wages in other local labour markets (MORRISON; PAPPS; POOT, 2006). The elasticity of quits to the expected relative pay is $-\eta$ in the local labour market and $-\mu$ in neighbouring areas. Wages and the unemployment rate of other local labour markets can be approximated by the economy-wide values, and they are weighted by an average migration cost between region r and other regions r' , $\gamma(c_{r'r})$, where $\gamma' < 0$ and $\gamma'' > 0$.

$$q_r = \left(\frac{w_r}{(1-u_r)\bar{w}_r} \right)^{-\eta} \left(\frac{w_r}{\gamma(c_{r'r})(1-u_{r'})\bar{w}_{r'}} \right)^{-\mu} \quad (4.2)$$

Firm's profits are given by $\pi_r = R_r(L_r) - C_r L_r - w_r L_r$. In this equation, $R_r(L_r)$ is the revenue received by firms, C_r is the replacement cost defined in Equation 4.1. Workers can go on a strike in a bargaining process. In this case, firms receive 0 payoff, while workers may receive a fraction $\varphi \in [0,1]$ of the expected local alternative wage or a fraction of the alternative wage in neighbouring areas ($\theta \in [0,1]$), with $\theta + \varphi < 1$. Then, the equilibrium is achieved through a Nash bargaining framework, which involves the optimisation of the weighted combination of gains of workers and firms:

$$\max N = \max(w_r - \varphi(1-u_r)\bar{w}_r - \theta\gamma(c_{r'r})(1-u_{r'})\bar{w}_{r'})^{\beta_r} \pi_r^{1-\beta_r} \quad (4.3)$$

In this setting, β is the bargaining power of workers. In equilibrium, $w_r = \bar{w}_r = \bar{w}_{r'}$, and all firms are assumed to be identical. Then, following Manning (2011), the equilibrium wage will be given by:

$$w_r = b_r \left(\frac{R_r}{L_r} - C_r \right) + (1-b_r)(\eta + \mu)C_r \quad (4.4)$$

Where:

$$0 \leq b_r = \frac{\beta_r}{1 - (1-\beta_r)(\varphi(1-u_r) + \theta\gamma(c_{r'r})(1-u_{r'}))} \quad (4.5)$$

When $\beta = 0$, the model becomes solely based on the efficiency wage mechanism, with $w_r^* = (\eta + \mu)C_r$, and in this case the elasticity of wages to the unemployment rate is given by:

$$\lambda_r^* = \frac{\partial w_r^*}{\partial u_r} \frac{u_r}{w_r^*} = -\eta \frac{u_r}{1-u_r} < 0 \quad (4.6)$$

On the other hand, if $\eta = \mu = 0$ and turnover costs are taken as exogenous, $w_r^B = b_r \left(\frac{R_r}{L_r} - H_r \right)$, the elasticity of wage to unemployment will be:

$$\lambda_r^B = \frac{\partial w_r^B}{\partial u_r} \frac{u_r}{w_r} = - \frac{(1 - \beta_r)\varphi u_r}{1 - (1 - \beta_r)\varphi(1 - u_r)} \leq 0 \quad (4.7)$$

However, the combination of efficiency wage and wage bargaining mechanisms generate a rent effect of the unemployment rate on bargained wages, whose elasticity is given by:

$$\lambda_r^R = \frac{\partial \left(\frac{R_r}{L_r} - C_r \right)}{\partial u_r} \frac{u_r}{\left(\frac{R_r}{L_r} - C_r \right)} = \eta \frac{u_r}{1 - u_r} \left(\frac{C_r}{\frac{R_r}{L_r} - C_r} \right) > 0 \quad (4.8)$$

Then, the total elasticity of wages in relation to unemployment will be:

$$\lambda_r = \frac{w_r^*}{w_r} \lambda_r^* + \left(1 - \frac{w_r^*}{w_r} \right) \lambda_r^B + \frac{b_r(1 + \eta) \left(\frac{R_r}{L_r} - C_r \right)}{w_r} \lambda_r^R \quad (4.9)$$

where $\frac{w_r^*}{w_r} \leq 1$. Taking the derivative of Equation 4.9 with respect to β_r generates the following relationship:

$$\frac{\partial \lambda_r}{\partial \beta_r} = \underbrace{-(\lambda_r^B - \lambda_r^*)}_{(-)} \frac{\partial \left(\frac{w_r^*}{w_r} \right)}{\partial \beta_r} + \underbrace{\left(1 - \frac{w_r^*}{w_r} \right)}_{(+)} \frac{\partial \lambda_r^B}{\partial \beta_r} + \underbrace{(1 + \eta) \left(\frac{R_r}{L_r} - C_r \right)}_{(+)} \lambda_r^R \underbrace{\frac{\partial b}{\partial \beta_r}}_{(+)} \quad (4.10)$$

According to Barth et al. (2002), this result can be interpreted in the following way. Under a pure efficiency wage setting, the introduction of a bargaining mechanism may lead to a higher sensitivity in relation to unemployment, if the efficiency wage effect is small and φ is high. However, as β_r increases, the alternative wage effect embedded in the turnover costs term diminishes, and the rent-sharing goes up, leading λ_r , which is negative, to values closer to zero. Then, if two different groups of workers with distinct levels of bargaining power are compared, those with a higher level of β_r will face less wage flexibility in the labour market in comparison with the other group.

In addition, the unemployment rate in neighbouring areas is inversely related to turnover costs (a higher unemployment rate in neighbouring areas reduces the expected wage associated with a migration decision of the worker, decreasing turnover costs). This term is included in the rent effect, and a higher u_r , will decrease wage flexibility to the local unemployment rate. On the other hand, the larger the commuting distance to other regions, the lower the turnover costs, indicating that the unemployment elasticity of pay will be greater in regions that are more isolated (LONGHI; NIJKAMP; POOT, 2006).

Agglomeration economies are also relevant in the context of the wage curve, as they are related to higher productivity and higher wages, accompanied by lower unemployment rates (SATO, 2000). This is an additional mechanism to generate the wage curve. Therefore, regions with higher urban concentration are expected to be in the left-upper side of the wage curve, while less-agglomerated areas are on the lower-right side. On the other hand, the efficiency wage model presupposes that regional amenities generate spatial differences in the equilibrium of wages and unemployment (CARD, 1995; ELHORST; BLIEN; WOLF, 2007). Then, high-amenity regions are supposed to present lower wages and higher unemployment rates, while low-amenity regions will be in the opposite location of the wage curve (high wages, low unemployment). This issue requires further investigation, but the agglomeration degree is an essential element to generate the wage curve.

In summary, the model developed here indicates that if the main differences between the formal and the informal sectors can be summarised by distinct degrees of bargaining power, where the former is supposed to present a lower wage flexibility to the business cycle than the latter. Accessibility to jobs can also affect the local wage, meaning that local density and the proximity to job opportunities in neighbouring areas may affect the observed wage curve. The next sections will investigate how these elements interact in the Brazilian labour market.

4.5 Empirical strategy

The model discussed in the previous section indicates that there is a negative relationship between wages and the local level of the unemployment rate. A reduced form of this model can be represented by:

$$w_{r,t} = \gamma_1 u_{r,t} + \sum_k X_{r,t,k} \beta_k + \theta_r + \delta_t + \epsilon_{r,t} \quad (4.11)$$

$$r = \{1, \dots, R\}, t = \{1, \dots, T\}$$

However, as discussed in Section 4.3, the empirical literature has started to control for individual heterogeneity in order to obtain a coefficient that actually captures this relationship. Therefore, individual wages at the individual level are related to the unemployment rate at the local level, controlling for observed and unobserved characteristics at both levels whenever possible.

$$w_{i,r,t} = \gamma_1 u_{r,t} + \sum_k X_{i,r,t,k} \beta_k + \theta_r + \delta_t + \varepsilon_{i,r,t} \quad (4.12)$$

$$i = \{1, \dots, N\}, r = \{1, \dots, R\}, t = \{1, \dots, T\}$$

In this initial formulation, $w_{i,r,t}$ is the logarithm of the individual hourly wage and $u_{r,t}$ is the logarithm of the unemployment rate in region r . Moreover, θ_r is the region effect, δ_t is a time effect and $X_{i,r,t}$ is a set of observed characteristics of individual i (age, age², education, sector, occupation, etc.). Whenever there is longitudinal data available at the individual level, an additional term α_i representing the individual fixed effect can be included to control for unobserved individual characteristics fixed in time. The parameter of interest in the equation above will be γ_1 , representing the elasticity of wages against the level of the unemployment rate.

According to the main elements of the theoretical framework discussed before, these models will be estimated for different population groups. Namely, whenever possible, the estimation of the wage curve will be conducted for the formal and informal sectors separately and for different degrees of urban agglomeration. Furthermore, following the argument of Meghir and Whitehouse (1996), the analysis is restricted only to men, as they will most likely stay in the labour market even when facing adverse conditions (low wages, high unemployment, inadequate work conditions).

4.5.1 Extensions

A few other issues must be investigated in detail to assess the consistency and robustness of the main results. The next subsections will briefly discuss these problems and the strategies developed to avoid them.

4.5.1.1 Multi-level analysis

The wage curve with controls for individual heterogeneity is characterised by variables in different aggregation levels. If this multilevel nature of the data is not taken into account, there may be a miscalculation of the significance of the effect of aggregated variables over micro-units. Considering the discussion developed by Moulton (1990), individuals in the same labour market may share some components of the variance that are not entirely explained by their observed characteristics, unobserved heterogeneity (constant in time) or by more aggregated

variables, such as the unemployment rate (at the area level). In this case, the error term in the equation associated with individual wages, X_{irt} and u_{rt} will be positively correlated across individuals in the same local market and as a consequence the standard error of the unemployment effect will be downwardly biased, generating higher t-statistics (CARD, 1995).

A solution proposed in some studies involves calculating averaged values of all individual-level variables at the area level. Then, the wage curve would be estimated directly at this aggregated level (BALTAGI; BLIEN; WOLF, 2000; BLANCHFLOWER; OSWALD, 1994; SANTOLIN; ANTIGO, 2009). Another possibility is to estimate the wage curve in two stages (BELL; NICKELL; QUINTINI, 2002). In the first stage, the logarithms of individual hourly wages are regressed against individual characteristics and a region-year dummy, which will synthesize the regional characteristics that explain wage differences over time. This first stage can be estimated either with repeated cross-sections or with a panel of individuals (with the control of the unobserved individual heterogeneity constant in time), depending on data availability.

$$w_{i,r,t} = \lambda_{r,t} + \delta_t + \sum_k X_{i,r,t,k} \beta_k + \eta_{i,r,t} \quad (4.13)$$

Then, the estimated region-year dummy $\hat{\lambda}_{r,t}$ is used in the second stage as a dependent variable, which will be explained by regional characteristics, including the unemployment rate.

$$\hat{\lambda}_{r,t} = \mu_r + \delta_t + \gamma_2 u_{r,t} + v_{r,t} \quad (4.14)$$

In this case, the parameter of interest in the estimation will be γ_2 , while μ_r and δ_t represent region fixed effects and time dummies.

4.5.1.2 Endogeneity of the unemployment rate

One of the main concerns of the empirical literature examining the wage curve regards the potential endogeneity of the unemployment rate due to simultaneity. The most common strategy is to consider time lags of this variable as instruments, with the underlying hypothesis that wages in the present do not affect the unemployment rate in the past nor do they share a common causing factor.

4.5.1.3 Selection bias

Most labour market analyses may be subject to the problem of selection bias, especially when the idea is to obtain estimates that are representative for the whole population and not only for a specific group with available information. In the case of the wage curve, the equation is supposed to represent the whole population of a working age, while data is observed only for employed individuals (a case of incidental truncation – WOOLDRIDGE, 2002, p. 560).

Within this context, the sample of employed individuals for whom wages are observed may not be randomly selected from the whole population. Card (1995) pointed out that Blanchflower and Oswald (1994) failed to discuss any potential effect of selection bias in the case of the wage curve, but composition bias can affect the elasticity of wages with respect to local unemployment rates. The selection model developed by Heckman (1979) aims to deal with this issue.

It is possible to revise Equation 4.12, rewriting it in a simplified version:

$$w = \gamma_1 u + X\beta + \varepsilon \quad (4.15)$$

But the wage is observed only if the person works ($L = 1$). The estimation of the parameters of interest must take into account this additional information in order to correct the limiting distribution of the estimator. The probability of being an employee is given by:

$$L = 1[Z\delta + \xi > 0] \quad (4.16)$$

in which L represents the condition of being an employee, while Z is a vector of variables associated with this probability. The main assumptions underlying this method are the following: (i) L and Z are always observed for all individuals; (ii) (ε, ξ) is independent of Z with 0 mean; (iii) $\xi \sim N(0,1)$; (iv) $E(\varepsilon|\xi) = \theta\xi$ (linear relationship between the error terms).

Then, the estimating equation can be derived by considering that the element to be estimated is actually $E(w|u, X, L = 1)$. Combining the assumptions above with this statement generates the final model:

$$E(w|u, X, L = 1) = \gamma_1 u + X\beta + \zeta\lambda(Z\delta) \quad (4.17)$$

where $\lambda(\cdot) \equiv \frac{\phi(\cdot)}{\Phi(\cdot)}$ is the inverse Mills ratio, which represents $E(L = 1|Z) = E(\xi|\xi > -Z\delta) = \lambda(Z\delta)$. A consistent estimator of δ is obtained with a first-stage probit estimation of the selection equation for the whole population. After that, it is possible to

consistently estimate γ_1 and β by regressing w against u , X and $\lambda(Z\hat{\delta})$ using the selected sample.

4.5.1.4 Spatial dependence and neighbouring effects

The model presented in Section 4.4 highlighted the possibility that the unemployment rate in neighbouring areas could affect wages directly (with a negative relationship) or indirectly, by increasing local wage flexibility. The subject of spatial effects in the wage curve has been significantly explored in the literature, sometimes including the spatial lag of the unemployment rate (LONGHI; NIJKAMP; POOT, 2006) or spatial autocorrelation in the error and in the dependent variable (BALTAGI; BLIEN; WOLF, 2012; FINGLETON; PALOMBI, 2013; RAMOS; NICODEMO; SANROMÁ, 2015; ELHORST; BLIEN; WOLF, 2007; BUETTNER, 1999).

This issue can be investigated in the following way. Departing from Equation 4.14, which is already at the region level, the error term can present spatial autocorrelation, which is tested using LM test statistics discussed by Elhorst (2014). Then, according to these tests results, two main options are available to account for spatial dependence. The spatial lag model can be specified as:

$$\hat{\lambda}_{r,t} = \mu_r + \delta_t + \rho W \hat{\lambda}_{r,t} + \gamma_2 u_{r,t} + v_{r,t} \quad (4.18)$$

While the spatial error model is written as:

$$\hat{\lambda}_{r,t} = \mu_r + \delta_t + \gamma_2 u_{r,t} + \epsilon_{r,t} \quad (4.19)$$

$$\epsilon_{r,t} = \varphi W \epsilon_{r,t} + v_{r,t} \quad (4.20)$$

where W is the spatial weights matrix based on inverse distances and common local labour market areas (see subsection 6.3.1 for further details), constant in time. The error terms $v_{r,t}$ are supposed to be independently and identically distributed for all r , with zero mean and variance σ^2 . The estimation of these different models follow Elhorst (2014), based on MATLAB routines available online³².

³² Available at <http://www.regroningen.nl/elhorst/software.shtml> and accessed on September 10, 2015.

4.5.1.5 Dynamic wage curve

Another possible extension to the basic wage curve model consists of including time-lagged wages in the original regression, obtaining a dynamic wage curve (BALTAGI; BLIEN; WOLF, 2009, 2012). The Phillips curve can be an alternative explanation to the relationship between wages and the unemployment rate (BLANCHARD; KATZ, 1997). In the literature, the estimation of a dynamic wage setting equation aims to test which relationship is more prone to be happening in a certain country. For instance, considering data aggregated at the regional level over time, the wage curve to be estimated can be the following (BELL; NICKELL; QUINTINI, 2002):

$$w_{r,t} = a_r + (1 - \lambda)w_{r,t-1} + \beta u_{r,t} + d_t + \varepsilon_{r,t} \quad (4.21)$$

In this case, there are three possible outcomes:

- If $(1 - \lambda) = 0$, then $\lambda = 1$, and the wage curve explanation dominates;
- If $(1 - \lambda) = 1$, $\lambda = 0$, and the Phillips curve seems to be more adequate;
- If $0 < \lambda < 1$, there is a dynamic wage curve, in which the impact of past regional changes is prolonged over more than one time period, and the long run elasticity of unemployment can be represented by β/λ .

4.6 Empirical analysis

The empirical analysis of the wage curve in Brazil is still limited by data availability. Differently from other case studies mentioned in Section 4.3, in Brazil there is still no longitudinal database with national coverage applied over a long period comparing the formal and the informal sectors. The main databases associated with labour market outcomes are provided by the Brazilian Institute of Geography and Statistics (Instituto Brasileiro de Geografia e Estatística – IBGE) with publicly available microdata:

- Monthly Employment Survey (Pesquisa Mensal de Emprego – PME) – with a moving panel of individuals (who are interviewed over sixteen months), for the period of March 2002 to June 2015, covering only six metropolitan regions (Recife, Salvador, Belo Horizonte, Rio de Janeiro, São Paulo, and Porto Alegre);
- National Household Sample Survey (Pesquisa Nacional por Amostra de Domicílios – PNAD), with repeated cross-sections for the period 1995-2013, with national coverage and state-level regional disaggregation;

- Demographic Census, a decennial survey consisting of repeated cross sections, for the whole country, allowing for municipal disaggregation.

This basic description shows that each of these databases covers a particular aspect of the labour market. Therefore, the next subsections will develop separate analyses of the wage curve with each of them, namely:

- Subsection 4.6.1 presents the results of the wage curve in a more aggregated perspective, with the PME database, in which individual heterogeneity is controlled for with the introduction of individual fixed effects and it is also possible to understand the relationship between wage flexibility and the business cycle; furthermore, results regarding the dynamic wage curve are provided;
- In Subsection 4.6.2, conclusions are drawn from the PNAD database, covering the whole country for a long period, controlling for state-level unobserved heterogeneity. It also provides a comparison between wage flexibility for different degrees of urbanisation (rural, urban, and metropolitan areas);
- Finally, Subsection 4.6.3 explores Census data, including Minimum Comparable Areas (MCAs) controls³³, neighbourhood effects and the effect of different urban scales over wage flexibility.

All these elements, even if investigated separately, provide a deep understanding of the wage curve in Brazil, with an important contribution to the literature given by the application of different methods to one specific case study. Table A.10 in the Appendix provides a detailed description of all the variables considered, including methods of calculation and data sources.

4.6.1 The wage curve in a more aggregated context: a comparison between the formal and the informal sectors

The first set of results presented here aim to measure wage flexibility in a more aggregated context, with monthly data, controlling for individual heterogeneity constant over time. Due to data availability, the wage curve is measured only for six metropolitan regions, which account for almost 25% of the population of the country. The next subsection will discuss the main characteristics of this database, while subsection 4.6.1.2 presents descriptive statistics. Then, subsection 4.6.1.3 discusses the main results and conclusions of this group of models.

³³ See Subsection 4.3 for a longer explanation.

4.6.1.1 Description of the database – Monthly Employment Survey

The Monthly Employment Survey (Pesquisa Mensal de Emprego – PME) is conducted by the Brazilian Institute of Geography and Statistics (Instituto Brasileiro de Geografia e Estatística – IBGE) providing monthly indicators about demographic, social and economic characteristics of the labour market in six metropolitan regions (Recife, Salvador, Belo Horizonte, Rio de Janeiro, São Paulo, and Porto Alegre). The survey underwent a deep methodology review in 2001, adopting the international patterns recommended by the International Labour Organization (ILO). This restriction implies that the period covered here will range from March 2002 to June 2015. However, all regressions will consider information from March 2003 on, in order to allow one-year lag of the unemployment rate to be considered as an instrumental variable.

The surveyed information is restricted to data on the labour market, and it can be analysed with its cross-section dimension or with a moving panel structure (this is the only IBGE survey in longitudinal format). Households are visited for two periods of four consecutive months, eight months apart from each other. Ribas and Soares (2010) provided a methodology and a Stata routine to reduce attrition in the PME panel, identifying individuals within the household and excluding those who have moved or who present inconsistent information over time³⁴.

A few steps were conducted to obtain the final database. Only male individuals aged 15 to 59 years old, working as employees in the formal or the informal sector were kept (excluding unemployed and out-of-the economically active population individuals, self-employed individuals, public sector employees, military, employers, and workers in the production for self-consumption). Formality status was defined by the existence or not of a contract under the CLT regulation (see Section 4.3). Furthermore, individuals employed in the public administration sector were also excluded, as well as those with wages equal to zero or with missing information, or with less than 20 or more than 60 working hours per week. Only individuals observed in the second and in the sixth interviews (twelve months apart from each other) were selected to be in the database.

The final database comprised 193,934 observations (96,967 surveyed individuals, observed in two moments in time, representing more than 49 million individuals over the period of more

34 The Data Zoom website (<http://www.econ.puc-rio.br/datazoom/english/>) provides packages including these algorithms to assess the longitudinal nature of the survey. We are thankful to Data Zoom, developed by the Department of Economics at PUC-Rio, for providing the codes for accessing IBGE microdata.

than 12 years when the sampling scheme was taken into account). The distribution of these observations over the years and months is presented in Table A.11 in the Appendix.

The aggregate unemployment rate was calculated for each metropolitan region for each month. In addition, wages were deflated by local inflation indexes (National Consumer Price Index – Índice Nacional de Preços ao Consumidor – INPC – in each metropolitan region). This correction with local inflation indexes aimed to at least partially control for the variation in local living costs over time. The initial differences between each metropolitan region were taken into account by the inclusion of local dummies.

4.6.1.2 Descriptive statistics

The average unemployment rate for the six metropolitan regions surveyed by PME decreased steadily from 2003 to 2014 and increased in the last few months of the period analysed. On the other hand, wages for the formal and the informal sector presented an upward trend in the period, with the important remark that the formal sector paid higher salaries on average than the informal sector.

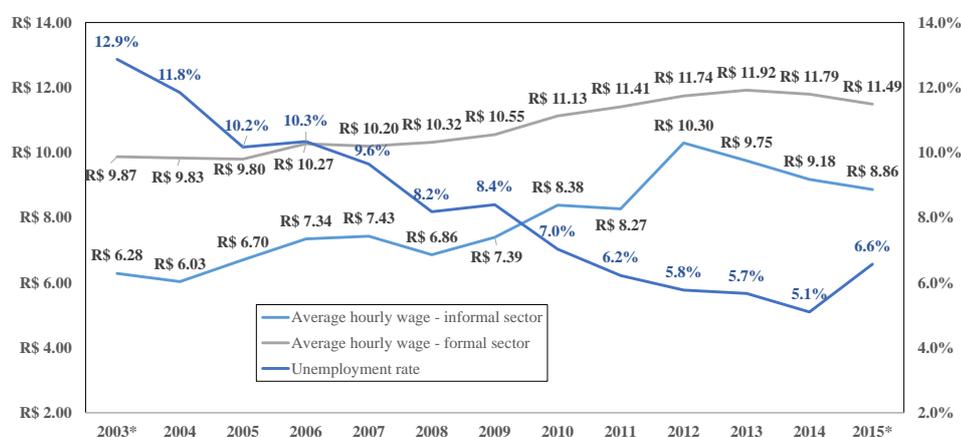


Figure 4.1 – Average hourly wage in the formal and the informal sectors and unemployment rate, 2003-2015.

* For 2015, the average was calculated only until June.

Obs.: Unemployment rate was calculated for the whole population, while average hourly wages were calculated only for the sample analysed.

Source: PME-IBGE.

Table 4.1 shows that this general trend for wages and unemployment can also be found when each metropolitan region is analysed separately. However, the timing of adjustment may vary between them. Furthermore, in 2015 wages were higher in Rio de Janeiro, São Paulo, Porto

Alegre and Belo Horizonte, while higher unemployment rates could be found in Salvador, Recife and São Paulo.

Table 4.1 – Average unemployment rates and hourly wages (both sectors) by metropolitan region, 2003-2015.

	Unemployment rate					Average hourly wage						
	Recife	Salvador	Belo Horizonte	Rio de Janeiro	São Paulo	Porto Alegre	Recife	Salvador	Belo Horizonte	Rio de Janeiro	São Paulo	Porto Alegre
2003*	14.3%	17.0%	11.0%	9.6%	14.4%	9.7%	R\$ 5.32	R\$ 6.99	R\$ 7.98	R\$ 8.74	R\$ 10.74	R\$ 7.95
2004	13.2%	16.4%	10.9%	9.5%	12.9%	8.9%	R\$ 5.41	R\$ 7.26	R\$ 7.68	R\$ 8.13	R\$ 11.04	R\$ 7.86
2005	13.7%	15.9%	9.0%	8.0%	10.5%	7.6%	R\$ 5.36	R\$ 7.35	R\$ 8.04	R\$ 7.87	R\$ 11.31	R\$ 8.17
2006	15.1%	14.0%	8.7%	8.3%	10.8%	8.3%	R\$ 5.50	R\$ 7.55	R\$ 8.28	R\$ 8.59	R\$ 11.80	R\$ 8.49
2007	12.4%	14.1%	7.8%	7.5%	10.4%	7.5%	R\$ 5.77	R\$ 7.22	R\$ 8.31	R\$ 8.90	R\$ 11.60	R\$ 8.48
2008	9.6%	11.8%	6.7%	7.1%	8.7%	6.1%	R\$ 5.67	R\$ 7.73	R\$ 8.90	R\$ 8.91	R\$ 11.44	R\$ 8.55
2009	10.2%	11.7%	6.6%	6.4%	9.5%	5.8%	R\$ 6.51	R\$ 8.19	R\$ 9.12	R\$ 9.29	R\$ 11.58	R\$ 8.78
2010	9.0%	11.3%	5.7%	5.9%	7.3%	4.7%	R\$ 6.48	R\$ 8.00	R\$ 9.68	R\$ 9.80	R\$ 12.47	R\$ 9.42
2011	6.8%	10.0%	5.1%	5.5%	6.4%	4.7%	R\$ 6.95	R\$ 8.32	R\$ 10.11	R\$ 10.26	R\$ 12.48	R\$ 9.89
2012	6.3%	7.5%	4.6%	5.3%	6.2%	4.2%	R\$ 7.64	R\$ 8.67	R\$ 10.82	R\$ 10.93	R\$ 13.05	R\$ 10.05
2013	6.7%	8.3%	4.3%	4.8%	6.2%	3.7%	R\$ 7.37	R\$ 8.45	R\$ 10.62	R\$ 11.41	R\$ 13.03	R\$ 10.54
2014	6.8%	9.4%	3.9%	3.8%	5.2%	3.9%	R\$ 7.77	R\$ 8.87	R\$ 10.64	R\$ 11.95	R\$ 12.39	R\$ 10.65
2015*	8.3%	11.6%	5.3%	5.0%	6.7%	5.3%	R\$ 8.09	R\$ 8.12	R\$ 10.50	R\$ 12.63	R\$ 11.56	R\$ 10.55

Obs.: Unemployment rate was calculated for the whole population, while average hourly wages were calculated only for the sample analysed.

* For 2003, the average was calculated only from March on, while in 2015, it was calculated only until June.

Source: PME-IBGE.

Table 4.2 – Descriptive statistics of the PME database.

	2003*	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015*
Age	34.3	34.6	34.6	34.6	34.8	35.1	35.5	35.7	35.7	35.9	36.2	36.5	36.8
Metropolitan region													
Recife	4.4%	4.6%	4.5%	4.0%	3.8%	3.4%	3.2%	3.4%	3.8%	3.9%	3.9%	3.8%	4.0%
Salvador	5.3%	5.4%	5.1%	5.1%	5.1%	4.8%	5.2%	4.7%	4.1%	4.5%	4.9%	4.8%	4.4%
Belo Horizonte	10.0%	9.9%	9.2%	9.2%	9.6%	10.1%	10.4%	10.9%	11.0%	10.9%	10.7%	10.1%	9.9%
Rio de Janeiro	24.8%	28.3%	30.0%	30.2%	30.8%	30.2%	28.5%	26.8%	26.3%	26.6%	25.6%	23.2%	21.8%
São Paulo	47.3%	44.0%	43.8%	44.5%	43.5%	43.6%	44.7%	46.2%	46.8%	46.0%	47.4%	50.9%	52.9%
Porto Alegre	8.2%	7.8%	7.4%	7.1%	7.3%	7.8%	7.9%	8.0%	7.9%	8.1%	7.5%	7.3%	7.0%
Student	10.8%	9.9%	9.9%	10.0%	9.9%	9.4%	9.1%	8.4%	8.3%	8.8%	8.4%	7.4%	6.1%
Education level													
Less than 1 year of schooling	2.1%	1.9%	1.7%	1.5%	1.2%	1.2%	1.2%	1.0%	1.2%	1.3%	1.1%	0.8%	0.6%
1 to 3 years of schooling	5.3%	5.3%	4.9%	4.7%	4.6%	4.3%	3.8%	3.2%	2.9%	2.6%	2.2%	2.0%	2.1%
4 to 7 years of schooling	25.1%	25.3%	23.6%	21.8%	21.2%	20.3%	19.9%	18.6%	17.0%	16.4%	15.3%	14.2%	13.3%
8 to 10 years of schooling	22.5%	21.4%	21.4%	21.2%	21.3%	21.3%	20.0%	20.4%	19.9%	18.9%	18.2%	18.7%	16.8%
11 years of schooling or more	45.0%	46.1%	48.3%	50.8%	51.7%	52.8%	55.1%	56.8%	59.0%	60.8%	63.2%	64.3%	67.2%
% of self-reported black or brown	40.3%	43.2%	44.9%	44.1%	45.7%	45.7%	46.2%	46.2%	47.5%	47.7%	47.6%	47.2%	47.2%
% of employees in the formal sector	84.1%	82.6%	82.9%	84.1%	85.0%	85.4%	86.0%	88.0%	88.6%	89.4%	91.1%	92.4%	93.7%
Sector of activity													
Industry	33.5%	32.2%	32.1%	31.5%	30.3%	30.3%	30.6%	30.6%	29.4%	29.0%	28.1%	28.1%	28.6%
Construction	7.0%	7.1%	7.3%	7.4%	7.7%	8.4%	8.6%	8.7%	9.6%	9.4%	9.5%	8.7%	8.3%
Commerce and repair	20.8%	20.9%	20.5%	20.2%	20.7%	20.6%	20.7%	19.7%	19.6%	19.9%	19.7%	19.8%	19.8%
Finance, real estate, serv. to companies	19.6%	19.9%	20.5%	21.2%	21.3%	20.7%	20.0%	20.8%	21.4%	21.4%	21.8%	21.7%	22.1%
Other services	18.6%	19.2%	18.9%	19.1%	19.4%	19.6%	19.6%	19.8%	19.5%	19.9%	20.4%	21.3%	20.7%
Other activities	0.7%	0.6%	0.6%	0.6%	0.6%	0.4%	0.5%	0.4%	0.5%	0.4%	0.5%	0.3%	0.4%

* For 2003, the average was calculated only from March on, while in 2015, it was calculated only until June.

Source: PME-IBGE.

Considering the composition of the database over time (Table 4.2), it is possible to note an increase in the average age from 2003 to 2015. São Paulo gained importance in the database, as well as higher education levels (especially 11 years of schooling or more, which passed from 45.0% to 67.2% of the population in the period). More importantly, the share of workers in the formal sector³⁵ rose from 84.1% to 93.7%. Jobs in the industrial sector lost importance in the total, while the share of workers in the construction and the financial sectors increased over time.

This preliminary analysis indicated that, at least in aggregated terms, there seemed to be a relationship between wages and the local unemployment rate, even if there were only six areas in the database. The following subsection aim to assess whether this relationship still remains after including controls at the individual, time and regional levels.

4.6.1.3 Results and preliminary conclusions

The first set of results attempted to control for the issues discussed in subsections 5.1.2 and 5.1.5 (namely, endogeneity of the unemployment rate and the estimation of a dynamic wage curve). This was possible because of the longitudinal nature of the database, with high-frequency information. Other issues such as the selection bias were not treated here because PME focuses mostly on labour market outcomes, and does not detail other dimensions of the individual's life. Spatial dependence was not investigated because the regions available here are disconnected and cover only a very small part of the country. The low number of regions was also the reason why the estimation here was conducted in only one stage (otherwise, the second stage would be based on a very small number of units).

Some controls were included in the estimated models, following the empirical literature. In the case of the models with individual fixed effects (two time periods for each individual separated by one year), controls that did not vary over time for the same individual or at least over a year were dropped (education level, age, month, among others).

The results presented in Table 4.3 indicated that for the basic model (OLS estimation), a wage curve could be observed only for the informal sector, while the coefficient for the formal sector was not significant. When these two sectors were aggregated, once again it was not possible to obtain a significant effect. However, Model 1 controlled only for observed individual

³⁵ Defined as employees in the private sector with a labour contract following the CLT, in opposition to informal workers who do not have this type of contract in their jobs.

characteristics. When individual fixed effects were included (Model 2), all effects disappeared. Model 3 considered the same estimation as Model 1, but now unemployment rates were instrumented by their time lag. Then, once again, there was an indication of a wage curve relationship only for the informal sector. In Model 4, individual fixed effects were added to the regression, and this generated non-significant coefficients. All these results indicated that the wage curve could be observed only in the case of the informal sector in Brazil, but this result may be generated by the fact that observed individual characteristics were not enough to control for all relevant variability.

Table 4.3 – Elasticity of wages in relation to the unemployment rate for different estimation strategies and sectors.

Models	Total	Informal	Formal
1) OLS	-0.012 (0.016)	-0.102** (0.045)	-0.007 (0.017)
2) FE	-0.001 (0.010)	0.055 (0.039)	-0.001 (0.010)
3) 2SLS	-0.012 (0.021)	-0.161*** (0.062)	-0.003 (0.023)
4) FE 2SLS	-0.004 (0.013)	0.037 (0.053)	0.003 (0.013)
5) Dynamic OLS	-0.008 (0.007)	-0.058** (0.023)	-0.003 (0.007)
6) Dynamic FE	-0.002 (0.009)	0.021 (0.035)	-0.003 (0.009)
7) Dynamic 2SLS	-0.007 (0.010)	-0.078** (0.032)	0.000 (0.010)
8) Dynamic FE 2SLS	-0.001 (0.012)	-0.005 (0.047)	0.005 (0.012)

Controls: (i) Cross-sectional models: age, age², education attainment, metropolitan region, sector of activity, self-reported black or brown, student, year, month; (ii) Longitudinal models: sector of activity, student, year. Instrument: unemployment rate one year before in the metropolitan region. Dynamic models: the wage of the individual in the previous month was included as an explanatory variable. Numbers in parentheses represent standard errors. All models were estimated with *pweight* in Stata to account for the proportional weight of each observation in the whole population. Standard errors are presented in parentheses. Significance levels: * p<0.10, ** p<0.05, *** p<0.01. Complete estimation results are presented in Tables A.12 and A.13 in the Appendix.

Source: Author's own calculations.

Models 5 to 8 allowed for the possibility of the existence of a dynamic wage curve. When individual fixed effects were included (Models 6 and 8), it was not possible to find any significant results. On the other hand, when only the cross-sectional dimension was taken into account (Models 5 and 7), including the time lag of wages generated coefficients that were half of the size of Models 1 and 3 (in absolute terms) for the informal sector, while there was no significant relationship for the formal sector.

The estimation of dynamic models also allowed for testing whether the relationship between wages and the unemployment rate could be better explained by a Phillips curve. Following the strategy mentioned in subsection 4.5.1.5, the results presented in Table 4.4 indicated that when only observed characteristics of the individuals were taken into account, the coefficient of the lagged wage was close to 1 (meaning that the Phillips curve could be a better analytical framework). However, the inclusion of individual fixed effects led to a division by half of these coefficients, signalling that a dynamic wage curve was the more appropriate explanation. The complete regression results can be found in the Appendix (Tables A.12 and A.13).

Table 4.4 – Elasticity of wages in relation to their time lags for different estimation strategies and sectors.

Models	Total	Informal	Formal
5) Dynamic OLS	0.875*** (0.002)	0.840*** (0.006)	0.877*** (0.002)
6) Dynamic FE	0.422*** (0.003)	0.447*** (0.010)	0.399*** (0.003)
7) Dynamic 2SLS	0.875*** (0.002)	0.840*** (0.006)	0.877*** (0.002)
8) Dynamic FE 2SLS	0.422*** (0.003)	0.447*** (0.010)	0.399*** (0.003)

Standard errors are presented in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Source: Author's own calculations.

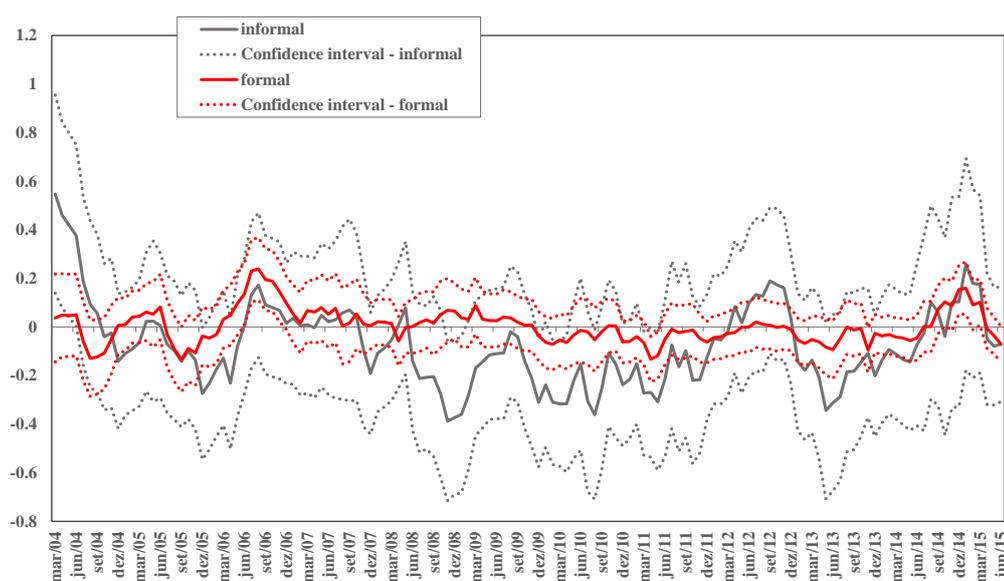


Figure 4.2 – Coefficients and confidence intervals for OLS regressions, with the sample comprising individuals in 12-month periods ending in each month.

Source: Author's own calculations.

One interesting extension for the models presented in Table 4.3 is to check how estimations may vary over time. Figure 4.2 shows the evolution of the coefficients for the formal and the informal sectors (and their confidence intervals) based on the estimation of Model 1. Each observation on the x-axis was obtained by estimating the model for the previous twelve months (moving sample). The confidence intervals showed that elasticities were significantly different from zero only in specific moments for the informal sector. In the case of the formal sector, the coefficient was not different from zero in a consistent way. These results indicated that even when a wage curve was found in aggregated terms, it will not necessarily be constant over time.

In summary, the coefficients for the informal sector seem to match the main results found in the literature for developed economies. A few aspects of this estimation must be highlighted. Firstly, metropolitan regions in a developing country such as Brazil do not necessarily represent the whole country – in general, they are much more diversified than other less dense areas, with better jobs and a labour market with particular characteristics. Moreover, all models including individual fixed effects will compare individuals in only two points in time, reducing within variation, which may explain why no effect was found in these models. These individuals were observed more than once in the database only if they stayed in the same metropolitan region (there were no individuals in the panel who moved during year they are interviewed). Finally, unemployment rates were observed only for six regions, meaning that this variable does not present very much variation between individuals.

4.6.2 Rural and urban areas and the wage curve

The second set of results discussed here involve the estimation of the wage curve with repeated cross-sections that allow for the control of individual observed characteristics and state-level data. This database presents the advantages of having national coverage, characterising rural and urban areas, and a long time series.

4.6.2.1 Description of the database – National Household Sample Survey

The National Household Sample Survey (Pesquisa Nacional por Amostra de Domicílios – PNAD), conducted by IBGE, investigates, on a yearly basis, demographic and socioeconomic characteristics of the household. Among these characteristics, it provides details about labour supply decisions, education level and income of all sources for the sample of individuals. PNAD is a cross-section survey with state-level representativeness, and has been conducted with a

relatively fixed structure ever since the beginning of the 1990s. Every ten years the sampling design is adjusted to the reality presented by the most recent demographic census. In this respect, the survey is not applied during Census years.

PNAD data was analysed for the period from 1996 to 2013 (leaving 1995 as a base year to provide the one-year lag of the unemployment rate as an instrument), after the macroeconomic normalisation and the implementation of the Real as the currency of the country. There are two years in this period in which the survey was not conducted, 2000 and 2010 (Census years). However, the fact that the time lag of the unemployment rate was considered as an instrumental variable requires the calculation of this rate for these missing years. This was done by obtaining the average of the rates observed in the previous and posterior years. Due to this restriction, two different periods of analysis were compared, 1996 to 2013 and 2002 to 2009. The former did not require this imputed unemployment rate, and its results served as a baseline for the main results. Furthermore, rural areas of the Northern states of Rondônia, Acre, Amazonas, Roraima, Pará and Amapá were excluded from the analysis, as they were not surveyed before 2004.

Following the same procedures of subsection 4.6.1, only male individuals aged 15 to 59 years old were kept in the database. The population of interest comprised employees with positive wages working in the private sector with or without a contract under the CLT regulation. The unemployment rate was calculated for each state considering all economically active individuals (unemployed and employed), and wages were deflated by the average inflation index of the whole country (National Consumer Price Index – Índice Nacional de Preços ao Consumidor – INPC). It was not possible to correct wages for local components of this index because they were available only for some states.

The results in this subsection sought to understand how the urbanisation level and different sectors of the labour market affect the level of wage flexibility. Furthermore, there was an effort to assess the impact of a possible selection bias in the final results by estimating a Heckman selection model. These selection models were estimated in a specific way for each case. When formal and informal employees in the private sector were considered, the Heckman selection model aimed to control for the probability of being an employee among all male individuals aged 15 to 59 years old. On the other hand, when the model considered only formal or informal workers, the probit model calculated the probability of working in this specific sector.

4.6.2.2 Descriptive statistics

The analysis must consider not only the characteristics of individuals in the formal and the informal sectors, but also how they compare with individuals outside of these groups (workers in other occupations – domestic workers, self-employed, public employees, employers – unemployed individuals, people out of the labour force). The reason to follow this procedure is that these other groups were part of the comparison group in the Heckman selection model.

Table 4.5 presents the main characteristics of the database. The first aspect to be highlighted is the fact that the formal sector presented a much higher wage than the informal sector. For the labour force ex-formal and informal, this average was calculated only for those who work. The average age in the informal sector was lower, as well as their tenure at the same company. In the formal sector it was possible to observe higher values of per capital income ex-wages, meaning that these families had a more stable financial condition. In terms of their regional distribution, workers in the informal sector were relatively more concentrated in the Northeast and in the North, while workers in the formal sector were more present in the Southeast and the South.

The education level of workers in the formal sector was higher (2013, 9% of them had 15 years of schooling or more, while only 3.8% of workers in the informal sector had achieved that level). It is noteworthy that between 1996 and 2013, real wages and the educational level of the population increased significantly, while the participation of informal workers in the labour force went down.

Informal jobs were relatively more concentrated in agriculture, construction and services in 2013. At the same time, the formal sector was relatively more present in manufacture, transport and communication, and commerce. A similar pattern was observed for occupational groups: those requiring administrative, technical, artistic, scientific and transportation skills were relatively more frequent amongst jobs in the formal sector. In the informal sector, agriculture, livestock and extractive occupations showed a much higher participation in 2013.

Table 4.5 – Descriptive statistics of the PNAD database.

	1996				2013			
	Labour force ex-formal and informal	Formal or informal	Formal	Informal	Labour force ex-formal and informal	Formal or informal	Formal	Informal
ln(hourly wage)	1.74	1.40	1.68	0.95	2.05	1.79	1.92	1.39
Age	37.27	30.39	31.85	28.11	39.79	33.57	34.25	31.60
Tenure	9.44	4.40	4.78	3.81	10.52	4.83	4.89	4.63
ln(per capita income ex-wage)	4.43	4.42	4.56	4.21	5.87	5.84	5.94	5.56
Region								
North	5.2%	3.7%	2.7%	5.3%	7.3%	5.6%	4.9%	7.8%
Northeast	30.9%	21.4%	14.6%	32.6%	31.1%	22.4%	17.7%	36.6%
Southeast	41.1%	52.4%	59.2%	41.1%	39.9%	47.7%	51.6%	35.8%
South	15.6%	15.8%	18.2%	11.9%	14.0%	15.9%	17.4%	11.6%
Centre-West	7.1%	6.8%	5.3%	9.2%	7.6%	8.3%	8.4%	8.2%
Education level								
Less than 1 year of schooling	14.1%	12.5%	7.0%	21.7%	8.3%	5.7%	4.0%	11.1%
1 to 3 years of schooling	16.1%	15.6%	12.7%	20.5%	7.2%	5.6%	4.0%	10.3%
4 to 7 years of schooling	33.4%	36.5%	36.3%	36.9%	23.2%	20.1%	17.2%	29.2%
8 to 10 years of schooling	16.6%	17.8%	20.9%	12.6%	22.0%	21.3%	20.5%	23.8%
11 to 14 years of schooling	14.2%	14.0%	18.3%	6.9%	29.2%	39.5%	45.3%	21.9%
15 years of schooling or more	5.6%	3.6%	4.9%	1.4%	10.1%	7.7%	9.0%	3.8%
% of self-reported black or brown	44.1%	44.4%	37.8%	55.3%	54.1%	54.5%	51.5%	63.6%
Student	18.0%	11.6%	10.0%	14.2%	19.6%	9.6%	8.4%	13.2%
Sector of activity								
Agriculture	20.3%	18.5%	9.0%	34.1%	11.9%	10.8%	6.2%	24.9%
Manufacture	4.0%	26.5%	34.3%	13.8%	3.4%	22.2%	26.1%	10.3%
Construction	7.3%	11.2%	8.3%	16.1%	10.9%	15.1%	12.8%	22.5%
Other industrial activities	1.4%	1.3%	1.5%	0.8%	0.7%	1.8%	2.2%	0.7%
Commerce	9.1%	14.7%	15.8%	12.8%	7.4%	16.4%	17.5%	13.0%
Services	9.1%	11.9%	12.1%	11.6%	7.9%	9.9%	8.5%	14.2%
Auxiliary services	2.7%	4.0%	4.5%	3.2%	2.7%	7.6%	8.9%	4.0%
Transport and communication	4.0%	6.4%	8.0%	3.7%	4.9%	9.9%	11.0%	6.5%
Social	3.4%	3.2%	3.7%	2.4%	4.2%	3.0%	3.2%	2.4%
Public administration	7.2%	0.0%	0.0%	0.0%	8.5%	0.0%	0.0%	0.0%
Other activities	1.4%	2.2%	2.7%	1.4%	1.1%	3.2%	3.7%	1.6%
Unemployed or out of the labour force	30.1%	0.0%	0.0%	0.0%	36.3%	0.0%	0.0%	0.0%
Occupation group								
Technical, scientific, artistic or similar	4.4%	3.8%	4.7%	2.3%	5.8%	5.1%	5.5%	3.8%
Administrative	8.7%	11.6%	16.0%	4.5%	8.6%	13.8%	16.3%	6.1%
Agriculture, livestock, extractive activities	19.3%	17.7%	8.3%	33.1%	12.0%	10.4%	5.7%	24.8%
Manufacture	13.6%	32.8%	33.3%	32.0%	14.6%	32.7%	32.9%	32.1%
Commerce and auxiliary activities	9.3%	9.5%	9.0%	10.2%	7.2%	13.6%	13.6%	13.3%
Transport, communication, services and other	14.5%	24.6%	28.7%	17.9%	15.5%	24.5%	26.0%	19.8%
Head of the household	56.2%	58.8%	65.0%	48.6%	47.7%	51.4%	53.9%	43.9%
% of people who have children under 15	45.4%	48.5%	53.6%	40.1%	27.9%	35.3%	36.8%	30.6%
Number of observations (unweighted)	59,606	36,862	22,873	13,989	39,752	47,038	35,305	11,733
Number of observations (weighted)	27,544,714	17,499,508	10,895,624	6,603,884	34,974,593	26,908,249	20,270,411	6,637,838

* All calculations were based on a weighted average of the observations (using *pweight* in Stata).

Source: PNAD-IBGE.

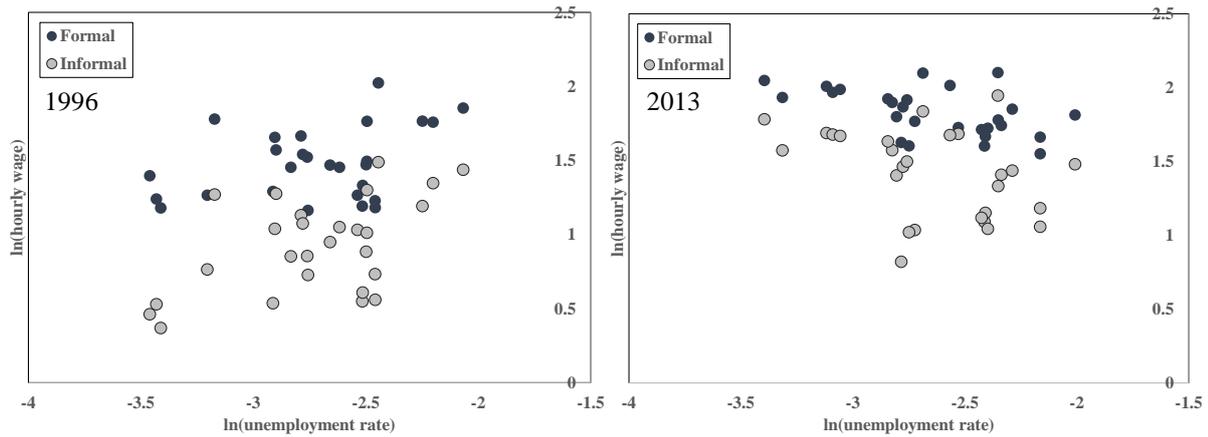


Figure 4.3 – Logarithm of the average hourly wage for the sample and logarithm of the unemployment rate at the state level, formal and informal sectors, 1996 and 2013.

Source: Author's own calculations.

Finally, salaries and the unemployment rate should be compared in each state over time. It is possible to note from Figure 4.3 that in 1996 the aggregated relationship between wages and the local unemployment rate for the formal and the informal sectors seemed to have a positive sign, while in 2013 their correlation seemed negative (more negative for the informal sector). In fact, Morrison and Poot (1998) identified that aggregated data may not provide any convincing evidence of the wage curve. However, when individual and local characteristics were controlled for, this pattern may change. The next subsection will discuss how wage flexibility varies according to the sector and the local characteristics of the labour market.

4.6.2.3 Results

The first set of results of the wage curve estimated with PNAD data was based on Equation 4.12, (one-stage regression), with individual and regional-level data. This strategy was adopted here due to the limited number of regions (27 states), which could limit large-sample assumptions for a possible two-stage estimation as proposed in subsection 5.1.1. Furthermore, as discussed in subsection 6.2.1, two different time periods were analysed (1996 to 2013 and 2002 to 2009), to assess the possible varying effects over time, as well as how the results may be possibly affected by the instruments based on an imputation for 2000 and 2010 unemployment rates.

In Table 4.6, the first three columns present the results for pooled cross-sections OLS estimates of individual wages for the period from 1996 to 2013, controlling for state fixed effects. For the whole country (rural and urban), all coefficients were positive, with a lower value for the informal sector (Models 1a, 2a and 3a). However, then the model was estimated specifically

for rural areas (Models 1b, 2b and 3b), only the elasticity of unemployment of the informal sector was significant, and with the expected negative sign. On the other side, for urban areas (Models 1c, 2c and 3c), all elasticities were positive and significant, counteracting the expected result of the wage curve.

Table 4.6 – Elasticity of individual wages in relation to the state-level unemployment rate, for the whole country, urban and rural areas, and for the whole labour force, formal and informal sectors, 1996 to 2013.

1996 to 2013						
Models	OLS			IV		
	Total	Informal	Formal	Total	Informal	Formal
Total						
	Model 1a	Model 2a	Model 3a	Model 4a	Model 5a	Model 6a
ln(unemployment rate)	0.198*** (0.006)	0.052*** (0.010)	0.174*** (0.007)	0.266*** (0.010)	0.025 (0.017)	0.249*** (0.011)
F test instrument				164,515	50,967	106,030
Adjusted R ²	0.564	0.470	0.526	0.564	0.470	0.526
Rural						
	Model 1b	Model 2b	Model 3b	Model 4b	Model 5b	Model 6b
ln(unemployment rate)	-0.005 (0.015)	-0.049** (0.021)	-0.013 (0.019)	-0.051** (0.025)	-0.134*** (0.037)	-0.054* (0.030)
F test instrument				23,161	9,690	11,538
Adjusted R ²	0.443	0.298	0.432	0.442	0.297	0.432
Urban						
	Model 1c	Model 2c	Model 3c	Model 4c	Model 5c	Model 6c
ln(unemployment rate)	0.200*** (0.006)	0.062*** (0.012)	0.182*** (0.007)	0.274*** (0.011)	0.034* (0.020)	0.266*** (0.012)
F test instrument				136,989	38,618	94,736
Adjusted R ²	0.545	0.460	0.523	0.545	0.460	0.522

Controls at the individual level: age, age², tenure, tenure², education level, sector of activity, occupational group, head of the household, has children under 15, self-reported black or brown. Complete tables for the period from 1996 to 2013 are presented in the Appendix. All models were estimated with *pweight* in Stata to account for the proportional weight of each observation in the whole population. Standard errors are presented in parentheses. Significance levels: * p<0.10, ** p<0.05, *** p<0.01. Complete estimation results are presented in Table A.14 of the Appendix.

Source: Author's own calculations.

Finally, the last three columns in the right hand side considered an IV estimation, with a one-year lag of the unemployment rate as an instrument for the actual unemployment rate in each state. Now, the coefficient for the unemployment rate over wages in the informal sector in the whole country was no longer significant, even though the coefficients for the formal sector and the whole economy were positive (and larger than the ones observed in the OLS estimations).

When only individuals living in rural areas were taken into account, the elasticity of the unemployment rate with both sectors analysed simultaneously was negative and slightly significant, while the elasticity of the unemployment rate for wages in the informal sector increased in comparison to the OLS estimation and reached -0.134. Finally, the IV estimation for the urban sector indicated that the elasticity for the informal sector was not significant, while it was positive and significant for the formal sector and for the whole urban labour market. For all models, the F statistic indicated that the lag of the unemployment rate was a relevant instrument for the unemployment rate in t .

Table 4.7 – Elasticity of individual wages in relation to the state-level unemployment rate, for the whole country, urban and rural areas, and for the whole labour force, formal and informal sectors, 2002 to 2009.

Models	2002 to 2009					
	Total	OLS		Total	IV	
		Informal	Formal		Informal	Formal
Rural and Urban						
	Model 1a	Model 2a	Model 3a	Model 4a	Model 5a	Model 6a
ln(unemployment rate)	0.106*** (0.011)	0.040** (0.020)	0.098*** (0.013)	0.106 (0.077)	-0.680*** (0.182)	0.232*** (0.079)
F test instrument				4,594	1,022	3,271
R ² 2nd stage	0.558	0.456	0.529	0.558	0.450	0.528
Rural						
	Model 1b	Model 2b	Model 3b	Model 4b	Model 5b	Model 6b
ln(unemployment rate)	0.021 (0.030)	0.038 (0.040)	-0.021 (0.036)	-1.500*** (0.473)	-4.661*** (1.781)	-0.499* (0.296)
F test instrument				135	16	163
R ² 2nd stage	0.403	0.267	0.397	0.365	.	0.390
Urban						
	Model 1c	Model 2c	Model 3c	Model 4c	Model 5c	Model 6c
ln(unemployment rate)	0.103*** (0.012)	0.039* (0.022)	0.095*** (0.013)	0.185** (0.079)	-0.447** (0.178)	0.253*** (0.083)
F test instrument				4,302	1,055	3,008
R ² 2nd stage	0.541	0.450	0.526	0.541	0.447	0.526

Controls at the individual level: age, age², tenure, tenure², education level, sector of activity, occupational group, head of the household, has children under 15, self-reported black or brown. Instrument for the unemployment rate at the state level was the local unemployment rate one year before. All models were estimated with *pweight* in Stata to account for the proportional weight of each observation in the whole population. Standard errors are presented in parentheses. Significance levels: * p<0.10, ** p<0.05, *** p<0.01. Complete estimation results can be provided by the author under demand.

Source: Author's own calculations.

When the period of analysis was restricted to 2002 to 2009, there seemed to be more indication of the presence of a wage curve in the relationships of the labour market. In general, elasticities

were lower or even became negative, a sign that the wage curve seemed to better express this specific period. This regression covered the same period studied by Baltagi, Rokicki and Souza (2014a). Even though their results seemed to better confirm the expected shape of the wage curve, the main conclusions were very similar to the ones found here. Wage flexibility seemed to be higher in rural areas and in the informal sector. This result was in accordance with the model discussed in Section 4.4 and added some conclusions to the conclusions obtained in subsection 6.1.

The last set of results in this block refer to the estimation of the wage curve accounting for a possible selection bias. As mentioned before, each model had a different first stage estimation. All of them were based on probit models, but the dependent variable in the selection regression was specific for each case.

Table 4.8 – Second-stage results of the IV estimation of the elasticity of individual wages in relation to the state-level unemployment rate with a Heckman selection model in the first stage.

Models	1996 to 2013			2002 to 2009		
	Total	Informal	Formal	Total	Informal	Formal
Total						
	Model 1a	Model 2a	Model 3a	Model 4a	Model 5a	Model 6a
ln(unemployment rate)	0.296*** (0.011)	-0.010 (0.019)	0.390*** (0.013)	0.227*** (0.087)	-0.763*** (0.205)	0.547*** (0.090)
F test instrument	128,480	138,479	86,812	3,472	1,978	2,682
Adjusted R ²	0.586	0.500	0.533	0.580	0.479	0.537
N	560543	198023	362520	306648	111572	195076
Rural						
	Model 1b	Model 2b	Model 3b	Model 4b	Model 5b	Model 6b
ln(unemployment rate)	-0.076*** (0.029)	-0.147*** (0.041)	-0.098*** (0.035)	-1.738*** (0.660)	-5.258* (2.869)	-0.419 (0.349)
F test instrument	17,728	31,384	9,856	75	0	125
Adjusted R ²	0.451	0.321	0.443	0.353	.	0.405
N	58774	37626	21148	32069	21039	11030
Urban						
	Model 1c	Model 2c	Model 3c	Model 4c	Model 5c	Model 6c
ln(unemployment rate)	0.295*** (0.012)	0.010 (0.022)	0.362*** (0.013)	0.241*** (0.087)	-0.567*** (0.196)	0.466*** (0.092)
F test instrument	106,826	100,940	78,588	3,305	2,185	2,511
Adjusted R ²	0.572	0.493	0.541	0.568	0.479	0.545
N	501769	160397	341372	274579	90533	184046

Controls at the individual level: age, age², tenure, tenure², education level, sector of activity, occupational group, head of the household, has children under 15, self-reported black or brown. Instrument for the unemployment rate at the state level was the local unemployment rate one year before. All models were estimated with *pweight* in Stata to account for the proportional weight of each observation in the whole population. Standard errors are presented in parentheses. Significance levels: * p<0.10, ** p<0.05, *** p<0.01. The complete estimation for 1996-2013 is provided in Table A.15 in the Appendix.

Source: Author's own calculations.

For instance, in the case of the model for the formal sector in rural areas, the selection equation compared individuals in the formal sector against all other workers (including those in the informal sector) and people out of the labour force. Furthermore, in the second stage, all models were estimated with IV (and the instrument was the logarithm of the time lag of the unemployment rate).

The results of Table 4.8 indicated that wage flexibility in the rural sector seemed to be even stronger when selection bias over observable characteristics was taken into account. In the case of Model 5b, the instrument did not explain much of the unemployment rate, generating problems for the estimation of the second stage of the IV model.

In summary, the analysis of the wage curve with state-level data provided some insights into the differences of wage flexibility between urban and rural areas. In fact, more accessibility to alternative jobs in agglomerated urban areas seemed to reduce the bargaining power of firms, decreasing their capacity to change offered wages according to the business cycle. There was also indication of labour market dualism and a much higher wage flexibility could be found in the informal sector, in accordance with the literature and with previous results discussed here (Section 4.3).

4.6.3 City size, spatial dependence, and the wage curve

There is a strong concern of whether state-level regional aggregation is satisfactory for the estimation of the wage curve. It is not clear if the unemployment rate in such large areas can substantially affect the decisions of economic agents at a more disaggregated local level. Furthermore, the state-level unemployment rate does not vary much between individuals, decreasing its explanatory power over the variation of wages. These issues may affect the estimation results. This subsection will be based on a multi-level analysis (see subsection 4.5.1.1) and will explore the wage curve in a more disaggregated context (municipalities), controlling for urban size, proximity, spatial dependence, and a possible endogeneity of the unemployment rate.

4.6.3.1 Description of the database – Demographic Census

Every 10 years the Brazilian Institute of Geography and Statistics (Instituto Brasileiro de Geografia e Estatística – IBGE) conducts the Demographic Census, which covers the whole

country with regional disaggregation at the municipal level (or at the neighbourhood level for bigger municipalities). The Demographic Census investigates the main characteristics of individuals and households (education, migration, labour market supply aspects, family composition, living conditions, among others), providing details on the life conditions of the population in each municipality, and serving as a very relevant policy instrument in a country with continental measures such as Brazil.

There is a shorter questionnaire applied to the whole population, while specific individual characteristics are investigated in a longer set of questions applied to a sample. Microdata at the individual level is available for this sample, which can range from 5 to 50 percent of the population depending on the size of each municipality (for instance, the sample size corresponds to 5% of the total population of municipalities with more than 500,000 inhabitants). Even if there are some changes over the years in this survey, the main structure and most methodological aspects have been kept constant over time, mainly in the last three censuses.

The analysis that follows was based on information obtained from the demographic censuses of 1991, 2000 and 2010. Over this period, there has been a significant change in the number of municipalities in Brazil, with the emancipation and consequent foundation of new local administrations. In 1991, there were 4,491 municipalities; in 2000, 5,507; and in 2010, 5,565 municipalities. Any analysis involving data over this time period should take into consideration the fact that the borders of municipalities have been changing, due to merging or splitting processes to create new municipalities (Reis et al., 2011). Therefore, 4,258 Minimum Comparable Areas (MCAs) for the period 1991-2010 were defined keeping constant the borders and areas of each regional unit of analysis. An important detail is that the municipality of reference was based on the place the individual worked (in order to avoid wage sub-estimation in municipalities that aggregate jobs and wage super-estimation in dormitory towns).

A second and bigger level of regional aggregation based on Census data is the Immediate Region of Urban Articulation (Regiões de Influência de Cidades – REGIC areas henceforth), a compound of municipalities that constitutes a labour market area. The criteria to set these local regions were defined by IBGE based on the flow of goods and services (IBGE, 2007) for the 2010 municipal configuration. Their configuration was harmonised with the 1991-2010 MCAs previously defined, obtaining 478 REGIC areas (from originally 482).³⁶ Here, they were

³⁶ The definition of the 1991-2010 MCAs and respective REGIC areas can be made available by the author upon request.

included in the concept of neighbourhood. Therefore, MCAs were considered neighbours if they belonged to the same REGIC area, and the strength of this neighbourhood was based on the inverse distance between the centroids of each MCA). For the cases in which only one MCA formed a REGIC area, the neighbourhood comprised the closest MCA, even if it belonged to another REGIC area.

Similarly to the previous procedures to generate the final database, only male individuals aged 15 to 59 years old were kept, employed with or without a formal contract (excluding self-employment, domestic work, workers in the public administration sector, employers, unemployed and individuals out of the economically active population), working from 20 to 60 hours per week. Furthermore, individuals working in the public sector or public administration were also excluded, in addition to those without information about the location of their jobs. Wages in 2000 were deflated to real values of August 2010 by the national INPC index of the period. In addition, a sampling procedure was adopted to generate a computationally manageable database. The original sample in 2000 comprised 2,237,102 interviews, representing 19,381,332 individuals in the population. In 2010, these numbers were 2,706,474 and 26,290,915, respectively.

Finally, the results that follow were based in two-stage regressions (see subsection 4.5.1.1), with the spatial wage obtained in the first stage being considered as the dependent variable in the second stage. However, there is not enough information to obtain estimates for all MCAs, because the sample does not cover the whole country. Then, when spatial dependence was explicitly considered in the second stage, it was necessary to input values for the missing data, and this was done by calculating the average of the spatial wage among the neighbours³⁷.

4.6.3.2 Descriptive statistics

The main descriptive statistics of the Census database are presented in two parts. Table 4.9 depicts the distribution of some variables over the sample of individuals for the estimation of the first stage of the wage curve. In general, real wages increased between 2000 and 2010. Over this period, education attainment of the population also went up, and workers were dislocated

³⁷ This procedure was adopted for 18 MCAs in 2000 and 3 MCAs in 2010 for the whole labour market (formal and informal), 454 MCAs in 2000 and 170 MCAs in 2010 for the formal sector, and 63 in 2000 and 29 in 2010 for the informal sector.

from agriculture to commerce and services. Manufacture had a relevant participation in the total employment analysed in this database, followed by occupations in commerce and service.

Table 4.9 – Descriptive statistics of the variables in the first stage (individual level).

	2000	2010
ln(hourly wage)	2.78	2.98
Age	32.01	33.71
Student	14.0%	12.0%
Self-reported black or brown	44.4%	51.2%
Head of the household	59.6%	50.6%
Education level		
Less than 7 years of schooling	53.6%	37.5%
8 to 10 years of schooling	19.7%	20.7%
11 to 14 years of schooling	21.1%	33.0%
15 years of schooling or more	5.6%	8.7%
Sector of activity		
Agriculture	14.9%	10.9%
Manufacture and construction	34.7%	33.2%
Other industrial activities	2.8%	2.7%
Commerce	18.3%	19.3%
Services	5.4%	6.1%
Auxiliary services	5.9%	7.4%
Transport and communication	7.9%	7.9%
Social services	2.0%	1.9%
Education	3.5%	2.5%
Other activities	4.7%	8.0%
Occupation group		
Leaders	3.8%	3.5%
Scientific, artistic or similar	4.6%	6.3%
Technical level	7.4%	7.2%
Administrative service	8.9%	9.1%
Commerce and service	20.1%	21.7%
Agriculture, livestock, extractive activit	13.6%	9.5%
Manufacture	41.6%	42.7%

Source: Demographic Census.

The first stage estimation generates a spatial wage at the MCA-year level, which then is used in the second stage as the dependent variable.

The upper part of Table 4.10 shows the mean spatial wages in 2000 for each group of MCAs, classified according to their density. It is noticeable that the spatial wage is higher for denser MCAs, in which it is also possible to find higher unemployment rates. However, the comparison of the estimated spatial wages for different sectors because they result from different regressions.

The section in the middle of Table 4.10 presents the same variables for 2010. Comparing these two years, the spatial wage increased for the formal sector and decreased for the informal sector, while the unemployment rate went down. Finally, the lower part of Table 4.10 presents the same variables averaged over all MCAs, corroborating the main findings of the two other parts.

Table 4.10. Descriptive statistics of the second stage – spatial wage.

2000				
	Spatial wage - total	Spatial wage - formal	Spatial wage - informal	ln(unemployment rate)
Density of the MCA				
Less than 1	-0.191	-0.219	-0.017	-2.249
1 to less than 2	-0.303	-0.289	-0.144	-2.360
2 to less than 5	-0.376	-0.351	-0.215	-2.274
5 to less than 10	-0.397	-0.376	-0.238	-2.415
10 to less than 20	-0.331	-0.342	-0.188	-2.482
20 to less than 50	-0.271	-0.303	-0.152	-2.228
50 or more	-0.009	-0.107	0.103	-1.860
2010				
	Spatial wage - total	Spatial wage - formal	Spatial wage - informal	ln(unemployment rate)
Density of the MCA				
Less than 1	0.052	0.052	0.172	-2.666
1 to less than 2	-0.070	-0.049	0.055	-2.731
2 to less than 5	-0.119	-0.110	0.005	-2.738
5 to less than 10	-0.145	-0.132	-0.006	-2.859
10 to less than 20	-0.120	-0.134	0.034	-2.942
20 to less than 50	-0.078	-0.126	0.060	-2.811
50 or more	0.094	-0.023	0.246	-2.617
	Spatial wage - total	Spatial wage - formal	Spatial wage - informal	ln(unemployment rate)
2000	-0.302	-0.310	-0.157	-2.310
2010	-0.082	-0.104	0.058	-2.809

Source: Author's elaboration.

4.6.3.3 Results

The main focus of this subsection is in the relationship between the spatial wage and the local unemployment rate. This spatial wage is obtained as the MCA-year effect net of observed individual characteristics. Complete tables with first-stage results are provided in Table A.16 in the Appendix.

Table 4.11 – Regression of the spatial wage against local characteristics, OLS and IV, 2000 and 2010.

	OLS						IV					
	Total	Informal	Formal	Total	Informal	Formal	Total	Informal	Formal	Total	Informal	Formal
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11	Model 12
ln(unemployment rate)	0.007 (0.005)	-0.028*** (0.006)	-0.006 (0.006)				0.002 (0.020)	-0.021 (0.023)	-0.047* (0.027)			
ln(unemployment rate) * density < 1				-0.052*** (0.008)	-0.084*** (0.009)	-0.060*** (0.010)				-0.098*** (0.023)	-0.122*** (0.027)	-0.141*** (0.031)
ln(unemployment rate) * 1 <= density < 2				-0.011* (0.007)	-0.041*** (0.008)	-0.029*** (0.009)				-0.052** (0.022)	-0.080*** (0.026)	-0.095*** (0.030)
ln(unemployment rate) * 2 <= density < 5				0.010* (0.006)	-0.023*** (0.007)	-0.011 (0.007)				-0.021 (0.022)	-0.047* (0.025)	-0.083*** (0.030)
ln(unemployment rate) * 5 <= density < 10				0.018*** (0.005)	-0.018*** (0.006)	-0.001 (0.007)				-0.011 (0.021)	-0.037 (0.024)	-0.065** (0.028)
ln(unemployment rate) * 10 <= density < 20				0.011** (0.005)	-0.026*** (0.006)	-0.002 (0.006)				-0.027 (0.020)	-0.049** (0.023)	-0.071*** (0.027)
ln(unemployment rate) * 20 <= density < 50				-0.009* (0.005)	-0.043*** (0.007)	-0.011* (0.007)				-0.061*** (0.021)	-0.074*** (0.024)	-0.095*** (0.028)
ln(unemployment rate) * density >= 50				-0.036*** (0.007)	-0.081*** (0.008)	-0.037*** (0.008)				-0.133*** (0.024)	-0.154*** (0.027)	-0.167*** (0.032)
Region	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R ²	0.515	0.440	0.217	0.531	0.455	0.228	0.465	0.381	0.147	0.435	0.354	0.135
F test - instrument: ln(unemployment rate) _{t-1}							700.4	699.6	682.1			
Minimum eigenvalue										91.1	90.1	88.7
N	8,487	8,416	7,885	8,487	8,416	7,885	4,148	4,109	3,726	4,148	4,109	3,726

*Additional controls: constant, percentage of individuals aged 25 or more who completed at least secondary school. Dependent variable is the spatial wage that results from the first stage regression of the logarithm of the individual hourly wage controlled by education attainment, sector of activity, occupation, age, age squared, if the individual is a student, the head of the household, or self-reported black or brown. The local unemployment rate is instrumented by its 10-year time lag. First-stage estimation results are presented in Table A.16 in the Appendix.

Source: Author's elaboration.

The initial results of the second stage are presented in Table 4.11. First of all, the spatial wage of the informal sector is negatively correlated with the local unemployment rate, while for the total and the formal sectors the results are not significant (Models 1 to 3). When these results are disaggregated by different density levels, the wage elasticity to the local unemployment rate is negative for all density classes in the informal sector, while for the formal sector, whenever this elasticity is negative, it is significant. There is a U shape relationship between wages and the unemployment rate according to density levels. This elasticity is higher in absolute terms for low and high-density MCAs, while it is lower for medium-density MCAs.

As mentioned in Section 4.5, the local unemployment rate may be endogenous to the wage level. Therefore, Models 7 to 12 provide a tentative instrumental variables estimation, with the 10-year lag of the unemployment rate as an instrument (and the lagged unemployment rate iterated with density groups). Now, all significant coefficients are negative, corroborating the expected shape of the wage curve, but the formal sector is the only one that presents a significant elasticity for all MCAs (Models 7 to 9).

Table 4.12 – Regressions for the spatial wage with MCAs fixed effects.

	FE					
	Total	Informal	Formal	Total	Informal	Formal
ln(unemployment rate)	-0.022*** (0.008)	-0.047*** (0.011)	-0.002 (0.012)			
ln(unemployment rate) * density < 1				-0.029* (0.017)	-0.056** (0.022)	-0.001 (0.025)
ln(unemployment rate) * 1 <= density < 2				-0.032** (0.013)	-0.038** (0.017)	-0.019 (0.020)
ln(unemployment rate) * 2 <= density < 5				-0.036*** (0.011)	-0.048*** (0.014)	-0.040*** (0.015)
ln(unemployment rate) * 5 <= density < 10				-0.039*** (0.009)	-0.061*** (0.012)	-0.014 (0.014)
ln(unemployment rate) * 10 <= density < 20				-0.022** (0.009)	-0.049*** (0.012)	-0.008 (0.013)
ln(unemployment rate) * 20 <= density < 50				-0.008 (0.010)	-0.033** (0.013)	0.015 (0.014)
ln(unemployment rate) * density >= 50				0.037*** (0.013)	-0.023 (0.017)	0.063*** (0.018)
MCA fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes
R ² between	0.300	0.058	0.007	0.244	0.059	0.022
R ² within	0.319	0.214	0.303	0.389	0.219	0.307
N	8487	8416	7885	8487	8416	7885

*Additional controls: constant, percentage of individuals aged 25 or more who completed at least secondary school. Dependent variable is the spatial wage that results from the first stage regression of the logarithm of the individual hourly wage controlled by education attainment, sector of activity, occupation, age, age squared, if the individual is a student, the head of the household, or self-reported black or brown. The local unemployment rate is instrumented by its 10-year time lag. First-stage estimation results are presented in Table A.16 in the Appendix. Source: Author's elaboration.

However, there is a potential problem here. Local unobserved characteristics of the labour markets may be affecting the results. Therefore, in Table 4.12, fixed effects for MCAs are included. These results are very different from the main conclusions obtained from Table 4.11. In fact, areas with higher population density will show a lower level of wage flexibility, as expected from the model formulated in Section 4.4. Therefore, the main conclusions obtained from Table 4.11 were being driven by unobserved local effects.

Finally, the last set of results allows for the existence of neighbourhood effects. As mentioned previously, the empirical literature has started to include not only the spatial lag of the unemployment rate as an explanatory variable, but also to control for spatial dependence. Models 1, 6 and 11 in Table 4.13 present the basic specification: the spatial wage is explained by local characteristics and the local unemployment rate. In this first estimation, LM tests indicated that the spatial autoregressive model (SAR) seemed to be the most appropriate spatial specification.

Then, the spatial lag of the unemployment rate was included in Models 2, 7 and 12. These results show that the informal sector not only had a higher elasticity for the local unemployment rate, but also for the unemployment rate in the neighbouring areas. Nonetheless, this formulation does not tackle the spatial dependence, which is once again found by LM tests. One last try before estimating the SAR model is to control for unobserved local characteristics, including MCA fixed effects. As expected, the elasticity of wages in relation to local unemployment rates has decreased, but the informal sector still showed the highest coefficients (Models 3, 8 and 13).

Moving to Models 4, 9 and 14, the estimation of a SAR model without local fixed effects indicates that the lag of the dependent variable is significant, and that the unemployment elasticity becomes non-significant (at least for the total labour market). However, the coefficient for the spatial lag of the unemployment rate is still negative and significant. Finally, the last three models (5, 10 and 15) indicate that whenever a SAR model with spatial fixed effects is estimated, the elasticity of the spatial wage to the local unemployment rate is very similar in the local level for the informal and the formal sectors (-0.023 and -0.024, respectively). However, the elasticity for the unemployment rate in the neighbours is much higher for the informal sector (-0.117, against -0.038 in the formal sector).

In sum, the estimation of the wage curve with disaggregated data shows that it is important to include spatial effects in the estimated model. In fact, the theoretical framework discussed in

Section 4.4 had already indicated that whenever another local labour market is sufficiently close to the region analysed, it is possible that workers look for opportunities in these neighbouring areas. This effect is supposed to be even more important here because REGIC areas (local labour markets) are composed by MCAs (the spatial unit of analysis). Therefore, a common shock inside the REGIC area is supposed to affect all MCAs that compose it, with spillover effects between them. This result indicates that there are relevant spatial effects inside a common labour market area. In addition, workers in the informal sector seem to be more affected by these neighbourhood effects.

This analysis is complemented by the results presented in Table 4.14. In that case, all models include the iteration between the local unemployment rate and the density group of the MCA. Furthermore, the inclusion of spatial effects is done through the iteration of the spatial lag of the unemployment rate and density groups.

The comparison of Models 1, 6 and 11 of Table 4.14 indicate that there is a higher wage flexibility in the informal sector. The elasticity of the spatial wage in relation to the local unemployment rate follows a U-shaped curve according to the local density. These results corroborate the initial conclusions obtained in Tables 4.12 and 4.13. Furthermore, spillover effects seemed to be more relevant for the informal sector, especially when fixed effects at the MCA level are taken into account (Models 3, 8 and 13). A great share of the difference in the wage flexibility between the formal and the informal sectors is captured by the spatial lag of the unemployment rate. This is an indication that the whole labour market area (REGIC area) seems to be more relevant for workers in the informal sector, while workers in the formal sector are usually more influenced by the unemployment rate at the MCA in which they are currently working.

Once again, LM tests indicated that a spatial autoregressive model would be more adequate to deal with spatial dependence. Combining the spatial lag of the dependent variable (spatial wage) with density groups and controlling for local fixed effects, Models 5, 10 and 15 showed that the wage flexibility in relation to the business cycle is much higher in absolute values in the informal sector, for low-density MCAs. In fact, these conclusions meet the predictions of the model discussed in Section 4.4, in which a worker employed in a less dense labour market has a lower bargaining power. Being surrounded by fewer job opportunities (higher unemployment rates in the neighbouring areas) reduces once again workers' bargaining power, and allow firms to adjust real wages according to the moment of the economy.

Table 4.13 – Regression of the spatial wage against local characteristics, OLS, FE and SAR, 2000 and 2010.

	Total					Formal					Informal				
	OLS	OLS	FE	SAR	SAR - FE	OLS	OLS	FE	SAR	SAR - FE	OLS	OLS	FE	SAR	SAR - FE
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11	Model 12	Model 13	Model 14	Model 15
ln(unemployment rate)	-0.035***	-0.024***	-0.018***	0.005	-0.016*	-0.051***	-0.034***	-0.025***	-0.018***	-0.024**	-0.068***	-0.044***	-0.026***	-0.021***	-0.023**
W*ln(unemployment rate)		-0.026***	-0.102***	-0.033***	-0.074***		-0.039***	-0.059***	-0.024***	-0.038**		-0.052***	-0.149***	-0.032***	-0.117***
Density >= 1 and density < 2	-0.096***	-0.096***				-0.063***	-0.063***				-0.102***	-0.102***			
Density >= 2 and density < 5	-0.154***	-0.154***				-0.123***	-0.122***				-0.155***	-0.155***			
Density >= 5 and density < 10	-0.178***	-0.179***				-0.150***	-0.150***				-0.172***	-0.173***			
Density >= 10 and density < 20	-0.166***	-0.167***				-0.156***	-0.157***				-0.160***	-0.162***			
Density >= 20 and density < 50	-0.111***	-0.111***				-0.142***	-0.143***				-0.115***	-0.116***			
Density >= 50	-0.062***	-0.059***				-0.096***	-0.090***				-0.046**	-0.039*			
W*Dependent variable				0.617***	0.303***				0.431***	0.359***				0.549***	0.193***
Constant	Yes	Yes		Yes		Yes	Yes		Yes		Yes	Yes		Yes	
% people completed at least middle school	Yes	Yes	Yes	Yes	Yes	Yes	Yes								
Region	Yes	Yes				Yes	Yes				Yes	Yes			
MCA fixed effects			Yes		Yes			Yes		Yes			Yes		Yes
LM lag	1918.3***	1930.9***	481.9***			1040.5***	1033.9***	746.6***			1093.3***	1079.6***	159.9***		
LM error	494.6***	489.7***	325.0***			317.9***	304.5***	242.2***			338.2***	312.1***	112.8***		
LM lag robust	1426.0***	1443.6***	336.8***			775.7***	783.1***	622.8***			785.0***	795.0***	120.3***		
LM error robust	2.3	2.3	179.9***			53.1***	53.7***	118.4***			29.9***	27.5***	73.2***		
R ²	0.523	0.523	0.318	0.592	0.836	0.221	0.222	0.200	0.300	0.682	0.451	0.453	0.247	0.487	0.784
N	8,516	8,516	8,516	8,516	8,516	8,516	8,516	8,516	8,516	8,516	8,516	8,516	8,516	8,516	8,516

*Dependent variable is the spatial wage that results from the first stage regression of the logarithm of the individual hourly wage controlled by education attainment, sector of activity, occupation, age, age squared, if the individual is a student, the head of the household, or self-reported black or brown. First-stage estimation results are presented in Table A.16 in the Appendix.

Source: Author's elaboration.

Table 4.14 – Regression of the spatial wage against local characteristics and the iteration of density groups and the local unemployment rate, OLS, FE and SAR, 2000 and 2010.

	Total					Formal					Informal				
	OLS	OLS	FE	SAR	SAR - FE	OLS	OLS	FE	SAR	SAR - FE	OLS	OLS	FE	SAR	SAR - FE
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11	Model 12	Model 13	Model 14	Model 15
ln(unemployment)*Density < 1	-0.092***	-0.085***	-0.058**	-0.093***	-0.060*	-0.103***	-0.041	-0.027	-0.045	-0.032	-0.121***	-0.113***	-0.079***	-0.129***	-0.081*
ln(unemployment)*Density >= 1 and density < 2	-0.051***	-0.013	0.033	-0.012	0.033	-0.076***	-0.007	-0.030	-0.016	-0.035	-0.079***	-0.026	0.046*	-0.020	0.047
ln(unemployment)*Density >= 2 and density < 5	-0.028***	-0.017	-0.031**	-0.001	-0.027	-0.055***	-0.057***	-0.094***	-0.050***	-0.089***	-0.059***	-0.045***	-0.050***	-0.034**	-0.047**
ln(unemployment)*Density >= 5 and density < 10	-0.018***	-0.007	-0.038***	0.002	-0.039**	-0.043***	-0.013	-0.014	-0.010	-0.020	-0.052***	-0.020	-0.038***	-0.011	-0.039**
ln(unemployment)*Density >= 10 and density < 20	-0.023***	-0.004	-0.012	0.017*	-0.005	-0.042***	-0.025**	-0.027	-0.045	-0.032	-0.058***	-0.023*	-0.024*	-0.010	-0.018
ln(unemployment)*Density >= 20 and density < 50	-0.041***	-0.049***	-0.046***	-0.001	-0.038*	-0.047***	-0.057***	-0.030	-0.016	-0.035	-0.073***	-0.070***	-0.035*	-0.031**	-0.026
ln(unemployment)*Density >= 50	-0.057***	-0.070***	0.110***	0.011	0.099**	-0.059***	-0.042	-0.094***	-0.050***	-0.089***	-0.102***	-0.124***	0.088**	-0.051*	0.090*
W*ln(unemployment)*Density < 1		-0.020	-0.093***	0.014	-0.057		-0.087***	-0.014	-0.010	-0.020		-0.035	-0.123***	0.015	-0.083
W*ln(unemployment)*Density >= 1 and density < 2		-0.052**	-0.170***	-0.038*	-0.139***		-0.094***	0.000	-0.011	0.009		-0.082***	-0.218***	-0.061**	-0.185***
W*ln(unemployment)*Density >= 2 and density < 5		-0.024	-0.113***	-0.037***	-0.083***		-0.018	-0.060***	-0.030**	-0.056**		-0.041**	-0.130***	-0.033*	-0.097***
W*ln(unemployment)*Density >= 5 and density < 10		-0.024*	-0.106***	-0.031***	-0.073***		-0.051***	0.094**	0.007	0.083		-0.059***	-0.155***	-0.047***	-0.118***
W*ln(unemployment)*Density >= 10 and density < 20		-0.031**	-0.113***	-0.050***	-0.091***		-0.036**	-0.032	-0.053*	-0.006		-0.062***	-0.158***	-0.052***	-0.129***
W*ln(unemployment)*Density >= 20 and density < 50		-0.004	-0.057***	-0.041***	-0.040*		-0.009	-0.066**	-0.057**	-0.034		-0.028*	-0.126***	-0.036**	-0.102***
W*ln(unemployment)*Density >= 50		-0.001	-0.130***	-0.063***	-0.111***		-0.039	-0.020	-0.005	-0.001		-0.008	-0.209***	-0.037	-0.185***
W*Dependent variable				0.610***	0.280***				-0.036***	-0.058**				0.528***	0.187***
Constant	Yes	Yes		Yes		Yes	Yes		Yes		Yes	Yes		Yes	
% people completed at least middle school	Yes														
Region	Yes	Yes				Yes	Yes				Yes	Yes			
MCA fixed effects			Yes		Yes			Yes		Yes			Yes		Yes
LM lag	1,936.1	1,948.5	389.0			1,045.2	1,025.5	667.4			800.1	1,086.9	145.9		
LM error	474.3	470.5	396.2			305.6	294.0	255.2			1,100.9	298.2	134.0		
LM lag robust	1,462.8	1,479.7	224.9			786.9	782.6	525.6			25.8	812.6	101.7		
LM error robust	1.0	1.7	232.1			47.3	51.1	113.4			326.6	24.0	89.8		
R ²	0.521	0.522	0.335	0.598	0.838	1045.227	1025.513	667.443	0.000	0.000	0.450	0.453	0.253	0.492	0.785

*Dependent variable is the spatial wage that results from the first stage regression of the logarithm of the individual hourly wage controlled by education attainment, sector of activity, occupation, age, age squared, if the individual is a student, the head of the household, or self-reported black or brown. First-stage estimation results are presented in Table A.16 in the Appendix.

Source: Author's elaboration.

The preferred estimates are Models 5, 10 and 15 of Tables 4.13 and 4.14. They provide comparisons between the formal and the informal sectors and the whole labour market, control for unobserved local characteristics and try to deal with spatial dependence. From Table 4.13, the estimated wage elasticity in relation to the unemployment rate is very similar for the formal and the informal sectors (-0.024 and -0.023, respectively), but its spatial lag has a much larger coefficient in the case of the informal sector (-0.117, against -0.038 for the formal sector).

The main conclusions of these results are the following: (i) wage flexibility is higher in less dense local labour markets; (ii) wages in the informal sector are more flexible than in the formal sector; (iii) it is essential to control for unobserved local characteristics in order to obtain the true elasticity of wages to local unemployment rates; (iv) it is important to control for spatial effects when the unit of analysis is sufficiently small; (v) a significant part of the difference between the formal and the informal sectors rely on spatial effects.

4.7 Conclusion

This chapter aimed to provide a comprehensive analysis of the wage curve in Brazil, exploring different databases in order to obtain a deep understanding of the mechanisms that are behind the process of wage bargaining. In order to do that, a theoretical framework combining wage bargaining, institutional costs and efficiency wages is provided, distinguishing the formal and the informal sectors by the relative level of bargaining power of workers employed in each of them.

Many empirical issues are investigated, aiming to understand their influence over the relationship of wages and local unemployment rates. Among them, it is possible to highlight multilevel analysis, endogeneity of the unemployment rate, selection bias, spatial dependence, and the degree to which the wage curve is dynamic or not.

The first set of results examined the wage curve in a more aggregated context, for six metropolitan regions. The informal sector seems be more flexible to economic downturns. This result seems to corroborate other findings in the literature for the wage curve in developing economies. However, the analysis of just a few metropolitan regions of the country is not enough to get a deep understanding of the whole underlying processes. Another conclusion of this set of results is that individual fixed effects should be controlled for whenever possible, as they explain a large share

of wage differentials. However, in the particular case of PME, their inclusion do not leave enough variation to capture the wage curve relationship.

Then, the analysis of the wage curve with state-level data from PNAD provided some insights into the differences of wage flexibility between urban and rural areas. In fact, higher accessibility to alternative jobs in agglomerated urban areas seemed to reduce the bargaining power of firms, decreasing their capacity to change offered wages according to the business cycle. There was also an indication of labour market dualism and a much higher wage flexibility could be found in the informal sector, following the literature (Section 4.3).

Finally, Census data seems to provide the best regional disaggregation to study the wage curve in Brazil. In fact, when local characteristics were adequately controlled for, the findings suggested that wage flexibility was higher in less dense local labour markets and wages in the informal sector were much more adjustable than those in the formal sector. Furthermore, spatial dependence is a relevant issue to account for all possible effects related to the accessibility of jobs to neighbouring areas and for common shocks at the local labour market level. In this sense, a large part of the difference between the formal and the informal sector is due to the spatial lag of the unemployment rate.

All these results corroborate the main predictions from the model discussed in Section 4.4. In fact, labour market dualism is an essential ingredient in the evaluation of the wage curve in developing economies. However, this type of analysis should be conducted in the adequate regional level (labour market areas). When this step is followed, it is possible to find a relevant impact of city size on the relative bargaining power of workers and firms.

5. CONCLUSION

This final chapter summarizes the main findings of this dissertation and discusses possible opportunities for future research on agglomeration economies in Brazil.

Cities seem to have triumphed, despite all possible negative consequences of urban concentration. The increasing agglomeration of workers and firms may generate congestion costs, and even though large cities present higher prices (transportation, land, work, life cost), there is still an ongoing trend of economic agents being attracted towards urban centres.

In Brazil, the urbanisation rate went from 77.6% to 85.7% between 1995 and 2015. During the past decades, these changes were fuelled by migration flows from rural to urban areas and from lagging to leading regions. Despite this trend towards urban concentration, there is still a pronounced disparity in wage and productivity levels according to the city size. For instance, GDP per capita in the sixteen largest urban centres in 2012 (with one million inhabitants or more) was more than twice the GDP per capita observed in municipalities with population up to 50,000. Such resilient differential indicates that productivity seems to be higher in large urban centres, but congestion and life costs can also present this pattern. Saturated cities in terms of access to public services may be attractive for individuals who can profit the most from their advantages. At the same time, medium-sized cities become the destination for less-skilled individuals, looking for a lower life-cost. Thus, understanding the main consequences of this process is of keen importance for Brazil.

Agglomeration economies constitute one of the possible explanations for the success of cities, being associated to an increasing productivity in large urban areas. They can be investigated in the context of local labour markets, whose outcomes result from a combination of spatial and labour market equilibriums. This dissertation aims to bring new elements to the discussion of the relationship between urban size and local labour markets outcomes in a developing economy.

Each empirical exercise developed here presents a different perspective over the location decisions of firms and workers, the resulting mix of sectors and skills, and the relative bargaining power that sets wages at the local level. According to the results obtained in Chapter 2, the location of firms in large cities is more profitable for technology and knowledge-intensive sectors. At the same time,

highly educated individuals are more likely to benefit from working in large cities. They are also more prone to migrate to denser areas and to stay there afterwards, as discussed in Chapter 3.

In a sense, these two main results complement each other: large cities become topmost centres of high-complexity economic sectors, with self-selected high-skilled individuals. Firms in these sectors are more likely to demand labour with higher qualification, and high-qualified workers are able to find better jobs. Matchings in the labour market will be more frequent and will have a higher quality. This type of environment stimulates knowledge exchange and innovation, increasing local productivity.

Economic interactions in large cities have additional effects over bargaining schemes and equilibrium outcomes of local labour markets. In this context, wage flexibility estimated by the wage curve is lower in large urban areas, indicating that a high matching probability in the labour market decreases firms' capacity to adjust wages (Chapter 4). The higher sector diversity usually found in bigger cities is one additional factor to reduce local wage sensitivity to economic downturns, as this industrial composition is able to accommodate uncorrelated random sector-specific shocks and reduce uncertainty.

Therefore, agglomeration economies seem to reinforce the main mechanisms that generate spatial concentration and regional inequality. Large urban centres become even more productive, with a larger share of high-skilled workers and an economic composition concentrated in high-complexity sectors. Furthermore, firms in these centres will have a lower capacity of adjusting wages according to economic fluctuations, resulting in a higher wage level.

Contributions of this work to the literature are manifold. Firstly, it provides variations of the estimation of agglomeration economies in Brazil, controlling for individual unobserved characteristics and exploring sector specificities. Secondly, it aims to understand the importance of city size to migration decisions, providing a deep understanding of the sorting process of workers. A special focus is given to the relevance of city size in migration decision. Furthermore, workers who migrate are more or less benefited from agglomeration economies depending on their qualification and their previous work experience. This is the first study aiming to measure dynamic agglomeration economies and the learning process that takes place in cities of a developing country, according to the knowledge of the author.

The final contribution relates to the estimation of the wage curve. A theoretical framework encompassing the main specificities of the labour market in a developing economy is proposed, with a large informal sector and potential spatial spillovers. This model adapts union bargaining and efficiency wage elements into a unique model, with an explicit regional dimension. The main predictions of this model are tested with three databases, covering different issues that are relevant in this literature. This investigation indicates that formal jobs in large urban areas will be less flexible to economic downturns. Such study is an effort to understand the implications of agglomeration economies in a more practical way.

5.1 Specific results

The main specific results found in each chapter can be summarised in the following way. Chapter 1 has presented a brief literature review on static and dynamic agglomeration economies, with a discussion of the main mechanisms that generate them. This discussion indicates that local characteristics may have a once-and-for-all positive effect over productivity (static advantages). Moreover, interactions between economic agents may imply in a specific path for local productivity growth, generating dynamic agglomeration externalities. In this sense, agglomerations may have long-lasting benefits, affecting the productivity growth trend for each city and economic agent over time.

Then, Chapter 2 provides the estimation of agglomeration economies with a clear focus on their industrial scope. Five broad sectors are defined (S1 – Manufacturing low-tech; S2 – Manufacturing medium-tech; S3 – Manufacturing high-tech; S4 – Services less-knowledge; and S5 – Services high-knowledge), and the main conclusion is that the best possible industrial mix to foster productivity may vary according to the sector technological intensity. In the case of S5, it seems that only diversity has a positive and significant relationship with local wages, suggesting that the Jacobian perspective is more appropriate. For S1, S2, and S4, even if diversity is positive and significantly related to wages, the combination of positive specialisation and negative competition coefficients indicates that the MAR framework could be more adequate to explain these patterns. For S3, both Marshall's and Jacobs's perspectives seem to be relevant. These results are robust to different specifications and estimation strategies.

Chapter 3 presents two main sets of results. The first of them aims to understand selection in initial and return migration of workers. Among the main conclusions, highly-educated individuals are more likely to migrate to work in a different place. In addition, the way incentives are laid down in space makes it more likely for regional inequalities to increase, as cities with a higher percentage of skilled workers attract more highly-qualified individuals.

Furthermore, second migration seems to reinforce the effects of first-time migration, because workers who decide to return to their city of origin are usually less qualified than those who stay in the first destination, and individuals who decide to move to a third place are positively selected once again. Initial productivity of workers who later decided to migrate to a dense city is higher than that of those who move to less dense cities. However, the lowest initial wage level is found for workers who have not migrated in any time. These results are in accordance to previous studies estimating migration models for workers in the formal sector in Brazil.

The second set of results of Chapter 3 concerns the estimation of the urban wage premium highlighting the differences between static and dynamic agglomeration externalities. Whenever workers' fixed effects are taken into account, they capture most of the effect of local density over wages. The inclusion of previous work experiences in different cities aims to capture dynamic agglomeration advantages. In this case, static agglomeration advantages become insignificant and whenever years of previous experience are combined with the current place of work, individuals working in less dense cities who had previous experience in denser areas will benefit the most from these gains. Finally, the inclusion of worker characteristics in these dynamic advantages settings shows that previous experience has a relevant impact on wage growth only in cities with at least the same density level of the current place of work. Therefore, dynamic advantages are expected to be more relevant for migrants who move to denser cities.

Finally, Chapter 4 provides different estimations of the wage curve in Brazil, based on a model that predicts that the informal sector in rural areas will be the most flexible segment in terms of wage adjustment to economic downturns. The empirical analysis corroborates these predictions, especially when the most appropriated regional disaggregation is considered (MCA level). The inclusion of spatial spillovers indicates that the accessibility to jobs in neighbouring areas and common shocks at the local labour market level (REGIC area) are both relevant. According to these results, a large share of the difference between the formal and the informal sectors is explained by

the spatial lag of the unemployment rate. Therefore, workers in the informal sector seem to be more mobile as they pursue better opportunities in neighbouring labour markets. As a result, the unemployment rate in the neighbouring areas will be more relevant for workers in the informal sector.

5.2 Future research

All these results indicate that agglomeration economies in Brazil are likely to stimulate spatial concentration and increase regional inequalities. This concentration process still seems to be in motion, indicating that congestion costs are not high enough to supersede agglomeration advantages. Workers and firms self-select themselves into agglomerated urban areas, in which they find a more diversified environment and a larger share of high-skilled individuals. Bigger centres also provide the conditions for workers to bargain for higher wages. This is the reason why certain firms prefer to locate in medium-sized cities, where they may have more bargaining power in the labour market. Furthermore, this effect increases hiring costs for firms in big cities.

Large cities are successful because they generate opportunities for more productive and risk-taking individuals to look for a job and interact with other firms and workers. In this environment, workers are more likely to find a better matching for their qualifications. However, whenever the trade-off defined by higher wages and high living costs is not perceived as positive, some individuals may decide to move to less dense urban centres.

The analysis of the existing literature and the contributions provided by this dissertation indicate a non-exhaustive list of topics that future research on agglomeration economies should focus:

- Generate complete theoretical models covering all relevant dimensions to workers and firms' location decisions and their consequences for local labour markets;
- Include the housing market in the model to account for local costs and to provide a complete estimation of workers' decisions;
- Provide a unified framework to understand dynamic agglomeration economies related to the learning process in cities;
- Bring agglomeration economies to a closer relation with economic theory;

- Include other types of proximity to improve the analysis of the interactions between workers and firms;
- Find alternatives to data limitations in the estimation of the wage curve in developing countries; and
- Further develop the analytical framework to investigate the wage curve, allowing for additional channels of interaction in space.

REFERENCES

- AGÉNOR, P-R. The analytics of segmented labor markets. In: AGÉNOR, P-R.; IZQUIERDO, A.; JENSEN, H.T. (Eds.) **Adjustment Policies, Poverty, and Unemployment: the IMMPA framework**. Malden: Blackwell Publishing, 1st Edition, p. 8-66, 2007.
- AMADEO, E.J.; CAMARGO, J.M. Chapter 7: Brazil: regulation and flexibility in the labor market. In: EDWARD, S.; LUSTIG, N.C. (Eds.) **Labor Markets in Latin America: Combining social protection with market flexibility**. Washington: Brookings Institution Press, p. 201-234, 1997.
- AMARAL, P.V., LEMOS, M., SIMÕES, R., CHEIN, F. Regional imbalances and market potential in Brazil. **Spatial Economic Analysis**, Vol. 5, N. 4, p. 463-482, 2010.
- ANDERSSON, M.; KLAESSON, J.; LARSSON, J.P. The sources of the urban wage premium by worker skills: spatial sorting or agglomeration economies? **Papers in Regional Science**, Vol. 93, N. 4, p. 727-747, 2014.
- ARROW, K.; DEBREU, G.; Existence of an equilibrium for a competitive economy. **Econometrica**, Vol. 22, p. 265-290, 1954.
- BALTAGI, B.H.; BASKAYA, Y.S.; HULAGU, T. How different are the wage curves for formal and informal workers? Evidence from Turkey. **Papers in Regional Science**, Vol. 92, N. 2, p. 271-284, 2013.
- BALTAGI, B.H.; BLIEN, U.; WOLF, K. The East Germany wage curve 1993-1998, **Economic Letters**, Vol. 69, p. 25-31, 2000.
- BALTAGI, B.H.; BLIEN, U.; WOLF, K. New evidence on the dynamic wage curve for Western Germany: 1980-2004. **Labour Economics**, Vol. 16, p. 47-51, 2009.
- BALTAGI, B.H.; BLIEN, U.; WOLF, K. A dynamic spatial panel data approach to the German wage curve. **Economic Modelling**. Vol. 29, p. 12-21, 2012.
- BALTAGI, B.H.; ROKICKI, B. The spatial Polish wage curve with gender effects: evidence from the Polish Labor Survey. **Regional Science and Urban Economics**, Vol. 49, p. 39-57, 2014b.
- BALTAGI, B.H.; ROKICKI, B.; SOUZA, K.B. The Brazilian wage curve: new evidence from the National Household Survey. **IZA Discussion Paper** N. 8468, 2014a.
- BARROS, R.P.; MENDONÇA, R.S.P. Flexibilidade do mercado de trabalho brasileiro: uma avaliação empírica. **Instituto de Pesquisa Econômica e Aplicada**, Texto para Discussão 452, 1997.

BARTH, E.; BRATSBERG, B.; NAYLOR, R.A.; RAAUM, O. Explaining variations in the wage curve. Memorandum 2002/03, **University of Oslo**, 2002.

BEAUDRY, C., SCHIFFAUEROVA, A. Who's Right, Marshall or Jacobs? The Localization versus Urbanization Debate, **Research Policy**, Vol. 38, 318-33, 2009.

BEHRENS, K.; DURANTON, G.; NICOUD, F.R. Productive cities: sorting, selection and agglomeration. **Journal of Political Economy**, Vol. 122, N. 3, p. 507-553, 2014.

BELL, B., NICKELL, S., QUINTINI, G. Wage equations, wage curves and all that. **Labour Economics**, Vol. 9, p. 341-360, 2002.

BERRY, C.R.; GLAESER, E.L. The divergence of human capital levels across cities. **Papers in Regional Science**, Vol. 84, N. 3, p. 407-444, 2005.

BLANCHARD, O.; KATZ, L. What do we know and do not know about the natural rate of unemployment. **Journal of Economic Perspectives**, Vol. 11, N. 1, 1997.

BLANCHFLOWER, D.G.; OSWALD, A.J. **The Wage Curve**. MIT Press, Cambridge, 1994

BLANCHFLOWER, D.G.; OSWALD, A.J. An introduction to the wage curve. **The Journal of Economic Perspectives**, Vol. 9, N. 3, p. 153-167, 1995

BLANCHFLOWER, D.G.; OSWALD, A.J. The wage curve reloaded. **NBER Working Paper Series**, Working Paper 11338, 2005.

BLIEN, U.; DAUTH, W.; SCHANK, T.; SCHNABEL, C. The institutional context of an 'Empirical Law': the wage curve under different regimes of collective bargaining. **British Journal of Industrial Relations**, Vol. 51, N. 1, p. 59-79, 2013.

BOOTH, A. Wage determination and imperfect competition. **IZA Discussion Paper**, N. 8034, 2014.

BOSCHMA, R.A. (2005), Proximity and innovation: a critical assessment, **Regional Studies**, Vol. 39, N. 1, p. 61-74, 2005.

BOTELHO, F.; PONCZEK, V. Segmentation in the Brazilian labor market. **Economic Development & Cultural Changes**, Vol. 59, N. 2, p. 438-463, 2011

BRAKMAN, S., GARRETSEN, H., VAN MARREWIJK, C. **The New Introduction to Geographical Economics**, Cambridge, UK: Cambridge University Press, 2009.

BUETTNER, T. The Effect of Unemployment, Aggregate Wages, and Spatial Contiguity on Local Wages: An Investigation with German District Level Data, **Papers in Regional Science**, Vol. 78, p. 47-67, 1999.

CAIRNCROSS, F. **The Death of Distance 2.0**, London: Texere Publishing Limited, 2001.

CAMAGNI, R., CAPELLO, R. Knowledge-based economy and knowledge creation: the role of space. In FRATESI, U., SENN, L. (Eds.), **Growth and Innovation of Competitive Regions: The role of internal and external connections**, Berlin Heidelberg: Springer- Verlag, p. 145–165, 2009.

CAMAGNI, R.; CAPELLO, R. CARAGLIU, A. Agglomeration economies and urban growth. **AISRe – ABC Workshop, The Frontier of the debate in Regional and Urban Economics**, available in http://www.aisre.it/images/call_for_paper/Workshop_dottorandi_ABC/03_Capello.pdf and accessed in 15/07/2015, 2015.

CAMPBELL, C.; ORSZAG, J. M. A model of the wage curve, **Economics Letters**, Vol. 59, p. 119-125, 1998.

CAPELLO, R. Proximity and regional innovation processes: is there space for new reflections? In: TORRE, A., WALLET, F. (Eds.) **Regional Development and Proximity Relations**. Cheltenham, UK: Edward Elgar, p. 163-194, 2014.

CAPELLO, R.; NIJKAMP, P. Introduction: regional growth and development theories in the twenty-first century – recent theoretical advances and future challenges. In: CAPELLO, R.; NIJKAMP, P. (Eds.) **Handbook of Regional Growth and Development Theories**. Cheltenham: Edward Elgar, p. 1-16, 2009.

CARD, D. The wage curve: a review. **Journal of Economic Literature**, Vol. 33, N. 2, p. 785-799, 1995.

CARLSEN, F.; RATTSO, J.; STOKKKE, H. Education, experience and dynamic urban wage premium. **Norwegian University of Science and Technology Working Paper**, 2013.

CARNEIRO, F.G.; HENLEY, A. Wage determination in Brazil: the growth of union bargaining power and informal employment. **The Journal of Development Studies**, Vol. 34, N. 4, p. 117-138, 1998.

CICCONE, A., HALL, R.E. Productivity and the density of economic activity. **The American Economic Review**, Vol. 86, N. 1, p. 54-70, 1996.

COMBES, P., DURANTON, G., GOBILLON, L., PUGA, D., ROUX, S. The Productivity Advantages of Large Cities, **Econometrica**, Vol. 80, N. 9, p. 2543–2594, 2012.

COMBES, P.P., DURANTON, G., GOBILLON, L. Spatial wage disparities: sorting matters! **Journal of Urban Economics**, Vol. 63, p. 723-742, 2008a.

COMBES, P.P., DURANTON, G., GOBILLON, L. The identification of agglomeration economies. **Journal of Economic Geography**. Vol. 11, p. 253-266, 2011.

COMBES, P.P., DURANTON, G., GOBILLON, L., ROUX, S. Chapter 1. Estimating agglomeration economies with history, geology, and worker effects. In: GLAESER, E.L. **Agglomeration Economics**. Chicago: The University of Chicago Press, p. 15-66, 2010.

COMBES, P.P., MAYER, T., THISSE, J.F. **Economic Geography**. Princeton: Princeton University Press, 1st Edition, 2008b.

COMBES, P.P.; GOBILLON, L. The empirics of agglomeration economies. In: DURANTON, G.; HENDERSON, J.V.; STRANGE, W.C. (Eds.) **Handbook of Regional and Urban Economics Vol 5A**, Amsterdam: North-Holland, 1st Edition, p. 247-348, 2015.

CORSEUIL, C.H.L.; FOGUEL, M.N. Economic expansion and an increase in labour market formality: a poaching approach. **Revista Brasileira de Economia**, Vol. 66, N. 2, p. 207-224, 2012.

D' COSTA, S.; OVERMAN, H.G. The urban wage growth premium: sorting or learning? **Regional Science and Urban Economics**. Vol. 48, p. 168-179, 2014.

DE GROOT, H.L.F, POOT, J., SMIT, M.J. Agglomeration externalities, innovation and regional growth: theoretical perspectives and meta-analysis. In: CAPELLO, R., NIJKAMP, P. (Eds.) **Handbook of Regional Growth and Development Theories**. Cheltenham, UK: Edward Elgar, p. 256-281, 2014.

DE LA ROCA, J. Selection in initial and return migration: evidence from moves across Spanish cities. **NYU Working Paper**, N. 23, April 2015.

DE LA ROCA, J.; PUGA, D. Learning by working in cities. **Working Paper** Available in <http://diegopuga.org/papers/esurban.pdf> and accessed in 01/07/2015, 2014.

DURANTON, G. Agglomeration and jobs. In: FISCHER, M.M.; NIJKAMP, P. (eds.) **Handbook of Regional Science**. Springer-Verlag Berlin Heidelberg, p. 631-648, 2014a.

DURANTON, G. Growing through cities in developing countries. **The World Bank Research Observer Advance Access**, p. 1-35, April, 2014b.

DURANTON, G., PUGA, D. Micro-foundations of urban agglomeration economies. In: Henderson, J.V., Thisse, J.F. (Eds.) **Handbook of Regional and Urban Economics. Vol. 4**. Amsterdam: North-Holland, p. 2063-2117, 2004.

DURANTON, G., PUGA, D. The growth of cities. In: Aghion, P., Durlauf, S.N. (Eds.) **Handbook of Economic Growth**, Vol. 2, Amsterdam: North Holland, 1st edition, p. 781-85, 2014.

EL BADAoui, E.; STROBL, E.; WALSH, F. The formal sector wage premium and firm size. **Journal of Development Economics**. Vo. 91, p. 37-47, 2010.

ELHORST, J.P. **Spatial Econometrics: From cross-sectional data to spatial panels**. Springer Briefs in Regional Science, New York: Springer, 1st Edition, 2014.

ELHORST, J.P.; BLIEN, U.; WOLF, K. New evidence on the wage curve: a spatial panel approach. **International Regional Science Review**. Vol. 30, N. 2, p. 173-191, 2007.

ELVERY, J.A. City size and skill intensity. **Regional Science and Urban Economics**. Vol. 40, p. 367-379, 2010.

ESTEVIÃO, M.; CARVALHO-FILHO, I. Institutions, informality, and wage flexibility: evidence from Brazil. **IMF Working Paper**, WP 12/84, 2012.

FALLY, T., PAILLACAR, R., TERRA, C. Economic geography and wages in Brazil: evidence from micro-data. **Journal of Development Economics**, Vol. 91, p. 155-168, 2010.

FIELDS, G.S. Labour market modelling and the urban informal sector: Theory and evidence [Electronic version]. In: TURNHAM, D.; SALOMÉ, B.; SCHWARZ, A. (Eds.), **The informal sector revisited**. Paris: Organisation for Economic Co-operation and Development, p. 49-69, 1990.

FIELDS, G.S. A guide to multisector labor market models. **Social Protection Discussion Paper Series**, N. 505, 2005.

FINGLETON, B.; PALOMBI, S. The wage curve reconsidered: is it truly an 'empirical law of economics'? **Région et Développement**, N. 38, p. 49-92, 2013.

FONTES, G.G., SIMÕES, R.F., OLIVEIRA, A.M.H.C. Urban attributes and wage disparities in Brazil: a multilevel hierarchical model. **Regional Studies**, Vol. 44, N. 5, p. 595-607, 2010.

FREGUGLIA, R.S.; GONÇALVES, E.; SILVA, E.R. Composition and determinants of the skilled out-migration in the Brazilian formal labour market: a panel data analysis from 1995 to 2006. **Economia**, Vol. 15, p. 100-117, 2014.

FREGUGLIA, R.S.; MENEZES-FILHO, N.A. Inter-regional wage differentials with wage heterogeneity: evidence from Brazil. **The Annals of Regional Science**, Vol. 49, Issue 1, p. 17-34, 2012.

FREGUGLIA, R.S.; PROCÓPIO, T.S. Efeitos da mudança de emprego e da migração interestadual sobre os salários no Brasil informal: evidências a partir de dados em painel. **Pesquisa e Planejamento Econômico**, Vol. 43, N. 2, p. 255-278, 2013.

FUJITA, M., THISSE, J.F. **Economics of Agglomeration: Cities, industrial location and regional growth**. Cambridge: Cambridge University Press, 2nd edition, 2012.

GLAESER, E.L. Are cities dying? **The Journal of Economic Perspectives**, Vol. 12, N. 2, p. 139-160, 1998.

GLAESER, E.L. Learning in cities. **Journal of Urban Economics**, Vol. 46, p. 254-277, 1999.

GLAESER, E.L. Introduction. In: GLAESER, E.L. (Ed.) **Agglomeration Economics**. National Bureau of Economic Research conference report. Chicago: The University of Chicago Press, Chicago, p. 1-14, 2010.

GLAESER, E.L. **Triumph of the City: How our greatest invention makes us richer, smarter, greener, healthier, and happier**. London, UK: Macmillan, 2011.

GLAESER, E.L. **Cities, Agglomeration and Spatial Equilibrium**. Oxford; New York: Oxford University Press, 2008.

GLAESER, E.L., GOTTLIEB, J.D. The wealth of cities: agglomeration economies and spatial equilibrium in the United States. **NBER Working Paper** 14806, 2009.

GLAESER, E.L., KALLAL, H.D., SCHEINKMAN, J.A., SHLEIFER, A. Growth in cities. **Journal of Political Economy**, Vol. 100, p. 1126-1152, 1992.

GLAESER, E.L., MARÉ, D.C. Cities and skills. **Journal of Labour Economics**, Vol. 19 N. 2, p. 316-342, 2001.

GLAESER, E.L.; RESSEGER, M.G. The complementarity between cities and skills. **Journal of Regional Science**, Vol. 50, N. 1, p. 221-244, 2010.

GORDON, I.R. Ambition, human capital acquisition and the metropolitan escalator. **Regional Studies**, Vol. 49, N. 6, p. 1042-1055, 2015.

GORDON, I.R.; MCCANN, P. Industrial clusters: complexes, agglomeration and/or social networks? **Urban Studies**, Vol. 37, p. 513-532, 2000.

GRAHAM, D.J.; MELO, P.S.; JUWATTANAJULPAISARN, P.; NOLAND, R.B., Testing for causality between productivity and agglomeration economies. **Journal of Regional Science**, Vol. 50, N. 5, p. 935-951, 2010.

GREENWOOD, M.J. Internal migration in developed countries. In: ROSEMBERG, M.R.; STARK, O. (Eds.) **Handbook of Population and Family Economics**, Vol. 1B, Amsterdam: North Holland, p. 647-720, 1997.

GROOT, S.P.T., DE GROOT, H.L.F. Estimating the skill bias in agglomeration economies externalities and social returns to education: evidence from Dutch matched worker-firm micro-data. **Tinbergen Institute Discussion Paper** TI 2014-088/VII, 2014b.

GROOT, S.P.T., DE GROOT, H.L.F., SMIT, M.J. Regional wage differences in the Netherlands: micro evidence on agglomeration externalities. **Journal of Regional Science**, Vol. 54, Issue 3, p. 503-523, 2014a.

HANLON, W.; MISCIO, A. Agglomeration: dynamic approach. **NBER Working Paper** 20728, 2014.

- HECKMAN, J.J. Selection bias as a specification error. **Econometrica**, N. 1, p. 153-161, 1979.
- HENDERSON, J.V. Efficiency of resource usage and city size. **Journal of Urban Economics**, Vol. 19, p. 47-70, 1986.
- HENDERSON, J.V. Cities and development. **Journal of Regional Science**, Vol. 50, N. 1, p. 514-540, 2010.
- HENLEY, A.; REZA ARABSHEIBANI, G.; CARNEIRO, F.G. On defining and measuring the informal sector: evidence from Brazil. **World Development**, Vol. 37, N. 5, p. 992-1003, 2009.
- HEUERMAN, D., HALFDANARSON, B., SUEDEKUM, J. Human capital externalities and the urban wage premium: two literatures and their interrelations. **Urban Studies**, Vol. 47, N. 4, p. 749-767, 2010.
- HOWELLS, J.R.L. Tacit knowledge, innovation and economic geography, **Urban Studies**, Vol. 39, N. 5-6, p. 871-884, 2002.
- IBGE – Instituto Brasileiro de Geografia e Estatística. **Divisão Urbano Regional**, 2013.
- JACOBS, J. **The Economy of Cities**. New York: Vintage, 1969.
- KAHN, M.E. Chapter 11 - New evidence on trends in the cost of urban agglomeration. In: GLAESER, E.L. (Ed.) **Agglomeration Economics**. National Bureau of Economic Research conference report. Chicago: The University of Chicago Press, p. 339-354, 2010.
- KOURTIT, K., NIJKAMP, P. In Praise of Megacities in a Global World. **Regional Science Policy and Practice**, Vol. 5, p. 167–182, 2013.
- KRUGMAN, P. The increasing returns revolution in trade and geography. **The American Economic Review**, Vol. 99, N. 3, p. 561-571, 2009.
- LALL, S.V., TIMMINS, C., YU, S. Connecting lagging and leading regions: the role of labor mobility. **Brookings-Wharton Papers on Urban Affairs**, pp. 151-174, 2009.
- LINDLEY, J., MACHIN, S. Spatial changes in labour market inequality. **Journal of Urban Economics**, Vol. 79, p. 121-138, 2014.
- LONGHI, S.; NIJKAMP, P.; POOT, J. Spatial heterogeneity and the wage curve revisited. **Journal of Regional Science**, Vol. 46, N. 4, p. 707-731, 2006.
- LUCAS, R. E. On the mechanisms of economic development. **Journal of Monetary Economics**, Vol. 22, p. 3-42, 1988.

LUCAS, R. E. B. Internal migration in developing countries. In: ROSEMBERG, M.R.; STARK, O. (Eds.) **Handbook of Population and Family Economics, Vol. 1B**, Amsterdam: North Holland, p. 721-798, 1997.

MACIENTE, A.N. The determinants of agglomeration in Brazil: input-output, labour and knowledge externalities. **PhD Dissertation**, University of Illinois at Urbana-Champaign, 2013.

MALONEY, W.F. Informality revisited. **World Development**, Vol. 32, N. 7, p. 1159-1178, 2004.

MANNING, A. The real thin theory: monopsony in modern labour markets. **Labour Economics**, Vol. 10, p. 105-131, 2003.

MANNING, A. Imperfect competition in the labor market. In: CARD, D. ASHENFELTER, O. **Handbook of Labor Economics Vol. 4B**, Amsterdam: North Holland, 1st Edition, p. 972-1041, 2011.

MARSHALL, A. **The Principles of Economics**. London: Macmillan, 1890.

MEGHIR, C.; WHITEHOUSE, E. The evolution of wages in the United Kingdom: evidence from micro-data. **Journal of Labor Economics**, Vol. 14, N. 1, p. 1-25, 1996.

MELO, P.C., GRAHAM, D.J., NOLAND, R.B. A meta-analysis of estimates of urban agglomeration economies. **Regional Science and Urban Economics**, Vol. 39, p. 332-342, 2009.

MELO, P.C.; GRAHAM, D.J. Agglomeration economies and labour productivity: evidence from longitudinal worker data for GB's travel-to-work areas. **SERC Discussion Paper** 31, October 2009.

MENEZES-FILHO, N.A.; MENDES, M.; ALMEIDA, E.S. O diferencial de salários formal-informal no Brasil: segmentação ou viés de seleção? **Revista Brasileira de Economia**, Vol. 58, N. 2, p. 235-248, 2004.

MION, G., NATICCHIONI, P. The spatial sorting and matching of skills and firms. **Canadian Journal of Economics**, Vol. 42, N. 1, p. 28-55, 2009.

MORETTI, E. Local labor markets. In: CARD, D. ASHENFELTER, O. (Eds.) **Handbook of Labor Economics, Vol. 4B**, Amsterdam: North-Holland, p. 1237-1313, 2011.

MORGAN, K. The exaggerated death of geography: learning, proximity and territorial innovation stems, **Journal of Economic Geography**, Vol. 4, p. 3-21, 2004.

MORRISON, P.S.; PAPPS, K.L.; POOT, J. Wages, employment, labour turnover and the accessibility of local labour markets. **Labour Economics**, Vol. 13, p. 639-663, 2006.

MORRISON, P.S.; POOT, J. The wage curve and local labour markets in New Zealand. In: MORRISON, P.S. (Ed.), **Labour, Employment and Work in New Zealand**. Victoria University of Wellington, Wellington, 1998.

- MOULTON, B.R. An illustration of a pitfall in estimating the effects of aggregate variables on micro units. **The Review of Economics and Statistics**. Vol. 72, N. 2, p. 334-338, 1990.
- NARITOMI, J., SOARES, R.R., ASSUNÇÃO, J.J. Institutional development and colonial heritage within Brazil. **The Journal of Economic History**. Vol. 72, Issue 2, p. 393-422, 2012.
- NIJKAMP, P.; POOT, J. The last word on the wage curve? **Journal of Economic Surveys**, Vol. 19, N. 3, p. 421-450, 2005.
- PAPAGEORGIU, T. Worker sorting and agglomeration economies. **McGill University Working Paper**, Sep. 2014.
- PERO, V.L. A carteira de trabalho no mercado de trabalho metropolitano brasileiro. **Pesquisa e Planejamento Econômico**. Vol. 22, N. 2, p.305-342, 1992.
- PESSOA, A. Agglomeration and regional growth policy: externalities versus comparative advantages. **The Annals of Regional Science**. Vol. 53, N. 1, p. 1-27, 2014.
- PORTER, M.E. **The Competitive Advantage of Nations**. Houndmills: Macmillan, 1990.
- PUGA, D. The magnitude and causes of agglomeration economies. **Journal of Regional Science**, Vol. 50, N. 1, p. 203-219, 2010.
- RAMALHO, H.M.B.; QUEIROZ, V.S. Migração interestadual de retorno e autosseleção: evidências para o Brasil. **Pesquisa e Planejamento Econômico**, Vol. 41, N. 3, p. 369-396, 2011.
- RAMOS, R.; DUQUE, J.C.; SURINACH, J. Is the wage curve formal or informal? Evidence for Colombia. **Economics Letters**, Vol. 109, p. 63-65, 2010.
- RAMOS, R.; NICODEMO, C.; SANROMÁ, E. A spatial panel wage curve for Spain. **Letters in Spatial and Resource Sciences**, Vol. 8, N. 2, p. 125-139, 2015.
- RATTSO, J.; STOKKE, H.E. Migration and dynamic agglomeration economies: regional income growth in Norway. **Norwegian University of Science and Technology Working Paper**, Feb. 2011.
- REIS, E.J.; PIMENTEL, M.; ALVARENGA, A.I.; SANTOS, M.C.H. Áreas mínimas comparáveis para os períodos intercensitários de 1872 a 2000. **1º Simpósio Brasileiro de Cartografia Histórica**, Paraty-RJ, accessed in 20/09/2015, available at https://www.ufmg.br/rededemuseus/crch/simpósio/REIS_EUSTAQUIO_JOSE_R_ET_AL.pdf , 2011.
- REIS, M.C. Os impactos das mudanças na demanda por trabalho qualificado sobre o desemprego por nível de qualificação durante os anos 90 no Brasil. **Revista Brasileira de Economia**, Vol. 60, N. 3, p. 297-319, 2006.

RIBAS, R.P.; SOARES, S.S.D. O atrito nas pesquisas longitudinais: o caso da Pesquisa Mensal de Emprego (PME/IBGE). **Estudos Econômicos**, Vol. 40, N. 1, p. 213-244, 2010.

ROSENTHAL, S.S., STRANGE, W.C. Chapter 49 – Evidence on the nature and sources of agglomeration economies. In HENDERSON, J.V., THISSE, J.F. (Eds.) **Handbook of Regional and Urban Economics. Volume 4**. New York: Elsevier, p. 2119-2171, 2004.

SANTOLIN, R.; ANTIGO, M.F. Curvas de salários dinâmicas: um estudo dos determinantes da histerese do desemprego no Brasil. **Cedeplar / FACE / UFMG**, Texto para Discussão 368, 2009.

SANTOS-JUNIOR, E.R.; MENEZES-FILHO, N.A.; FERREIRA, P.C. Migração, seleção e diferenças regionais de renda no Brasil. **Pesquisa e Planejamento Econômico**, Vol. 35, N. 3, p. 299-331, 2005.

SATO, Y. Search theory and the wage curve. **Economics Letters**, Vol. 66, p. 93-98, 2000.

SCITOVSKY, T. Two concepts of external economies. **Journal of Political Economy**, Vol. 62, N. 2, p. 143-151, Apr. 1954.

SILVA, A.S.P.; MONSUETO, S.E. Flexibilidade do mercado de trabalho: uma análise comparativa entre segmentos socioeconômicos no Brasil (2002-2009). **Proceeding of the 40th Brazilian Economics Meeting, ANPEC – Brazilian Association of Graduate Programs in Economics**, accessed in 15/10/2014, at http://www.anpec.org.br/encontro/2012/inscricao/files_I/i12-351de703e4da2073e562bebdf5fd0981.pdf, 2012.

SILVA, M.V.B., SILVEIRA-NETO, R.M. Dinâmica da concentração da atividade industrial no Brasil entre 1994 e 2004: uma análise a partir de economias de aglomeração e da nova geografia econômica. **Economia Aplicada**, Vol. 13, N. 2, p. 299-331, 2009.

SIMÕES, R.F.; FREITAS, E.E. Urban attributes and regional differences in productivity: evidence from the external economics of Brazilian micro-regions from 2000 – 2010. **International Journal of Economics**, Vol. 1, N. 2, p. 30-44, 2014.

SOARES, F.V. Some stylized facts of the informal sector in Brazil. **IPEA, Texto para Discussão**, N. 1020, 2004.

SOUZA, I. V.; MACHADO, A.F. Curva de rendimentos: uma análise no mercado de trabalho urbano e rural no Brasil (1981/99). **Revista de Economia e Sociologia Rural**, Vol. 42, N. 1, p. 35-54, 2004.

STARRETT, D. Market allocations of location choice in a model with free mobility. **Journal of Economic Theory**, Vol. 17, p. 21-37, 1978.

STIMSON, R.J. Proximity and Endogenous Regional Development. In: TORRE, A., WALLET, F. (eds), **Regional Development and Proximity Relations**, Cheltenham, UK and Northampton, MA, USA: Edward Elgar Publishing, p. 47–93, 2014.

STORPER, M., VENABLES, A.J. Buzz: face- to- face contact and the urban economy, **Journal of Economic Geography**, Vol. 4, N. 4, p. 351–370, 2004.

TAVEIRA, J. G.; ALMEIDA, E. Os determinantes regionais da atração do migrante qualificado. **Análise Econômica**, Ano 32, N. 69, p. 199-224, 2014.

TORRE, A., RALLET, A. Proximity and Localization, **Regional Studies**, Vol. 39, N. 1, p. 47-59, 2005

TURUNEN, J. Disaggregated wage curves in the United States: evidence from panel data of young workers, **Applied Economics**, Vol. 30, N. 12, p. 1665-1677, 1998.

ULYSSEA, G. Informalidade no mercado de trabalho brasileiro: uma resenha da literatura. **Revista de Economia Política**, Vol. 26, N. 4, p. 596-618, 2006.

ULYSSEA, G. Regulation of entry, labor market institutions and the informal sector. **Journal of Development Economics**, Vol. 91, p. 87-99, 2010.

UNFPA **State of World Population 2007: Unleashing the potential of urban growth**, New York: UNFPA, 2007.

UNITED NATIONS, Department of Economic and Social Affairs, Population Division. **World Urbanization Prospects: The 2014 revision**, (ST/ESA/SER.A/366), 2015.

VAN OORT, F.G., LAMBOOY, J.G. Cities, knowledge and innovation. In: Fischer, M.M., NIJKAMP, P. (Eds.) **Handbook of Regional Science**, Berlin Heidelberg: Springer-Verlag, p. 475-488, 2014.

WHEELER, C.H. Cities and the growth of wages among young workers: evidence from the NLSY. **Journal of Urban Economics**. Vol. 60, p. 162-184, 2006.

WOOLDRIDGE, J.M. **Econometric Analysis of Cross Section and Panel Data**. England: MIT Press, 1st Edition, 2002.

YANKOW, J.J. Why do cities pay more? An empirical examination of some competing theories of the urban wage premium. **Journal of Urban Economics**, Vol. 60, p. 139-161, 2006.

APPENDIX A1 – ADDITIONAL TABLES FOR THE INTRODUCTION

Table A.1 – Description of the main databases considered in the dissertation.

Database	Period	Regional coverage	Design	Information	Database structure
RAIS-MIGRA	annual, 1995-2008	Municipalities - whole country	Longitudinal - individual level	Formal sector, employed individuals, registration data	Population
Identified RAIS	annual, 2004, 2008, and 2012	Municipalities - whole country	Longitudinal - individual level	Formal sector, employed individuals, registration data	Population
Demographic Census	every 10 years, 1991-2000-2010	Municipalities, neighborhoods - whole country	Repeated cross sections - individual level	Education, work, migration, income, household characteristics, formal/informal sector, all income sources, unemployment	Sample
PNAD - National Household Sample Survey	annual, 1995-2013 (not conducted in 2000 and 2010)	State level, excluding rural North before 2004	Repeated cross sections - individual level	Education, work, migration, income, household characteristics, formal/informal sector, all income sources, unemployment	Sample
PME - Monthly Employment Survey	monthly - 2002-2014	6 metropolitan regions (São Paulo, Rio de Janeiro, Porto Alegre, Belo Horizonte, Recife and Salvador)	Individual - moving panel (individuals are interviewed for 4 months, stay out of the sample for 8 months and are interviewed for 4 months again, if they don't move)	Education, work, labor income, formal/informal sector, unemployment	Sample

Source: Author's elaboration.

APPENDIX A2 – ADDITIONAL TABLES FOR CHAPTER 2

Table A.2 - Description of the factors defined by Maciente (2012)³⁸ which aggregates the skill requirements of each occupation in Brazil.

Factor	Description
1	Cognitive skills
2	Maintenance and operation skills
3	Assistance skills
4	Management skills
5	Design and engineering skills
6	Transportation skills
7	Artistic skills
8	Accuracy and automation skills
9	Supervised work skills
10	Teaching and social science skills
11	Physical strength
12	Telecommunication skills
13	Independence skills
14	Natural science skills
15	Attention skills
16	On-the-job experience
17	Conflict management skills
18	Team-work skills
19	Sales skills
20	Monitoring and compliance skills
21	Clerical skills

Source: Maciente (2012).

³⁸ Maciente (2012) has developed a matching among the occupational classification in Brazil and the profile of skills based on ONET (Occupational Information Network). Using that, the author finds a weighting system for each ability required for the job, based on the worker qualification and the job complexity. Then, with a factor analysis it is possible to calculate the skill requirement of each occupation.

Table A.3 - Indicators and measures of the industrial scope of agglomeration economies.

Indicator	Measure	Formula	Interpretation	Source
Specialization	Location quotient	$LQ_{industry,region} = \frac{E_{industry,region}/E_{region}}{E_{industry,country}/E_{country}}$	LQ > 1: region is more specialised in industry i.	Glaeser et al. (1992)
	Degree of specialization	$SD_{region} = \sum_{ind} \left(\frac{E_{industry,region}}{E_{region}} - \frac{E_{industry}}{E_{country}} \right)^2$	SD close to 0: industrial composition in the region is similar to the national one; SD close to 2: region is completely specialized.	Henderson (2003)
Competition	Degree of competition	$C_{industry,region} = \frac{F_{industry,region}/E_{industry,region}}{F_{industry,country}/E_{industry,country}}$	C > 1: higher competition in the industry in the region than in the country.	Glaeser et al. (1992)
Diversity	Inverse Herfindhal index	$IHI_{region} = \left(\frac{E_{region}^2}{\sum_{ind} E_{industry,region}^2} \right)$	Higher IHI means more diversity.	Combes et al. (2008a, 2011)
	Non-diversity	% of the largest five industries other than the industry in question in the city's employment.	The higher the %, the lower the diversity.	Glaeser et al. (1992)
	Shannon entropy	$S_{region} = - \sum_{industry} \left(\frac{E_{industry,region}}{E_{region}} \ln \frac{E_{industry,region}}{E_{region}} \right)$	High value means high diversification (accounts for the size distribution of sectors).	Groot et al. (2014)
Urbanization	Density	$Den_{region} = \ln \left(\frac{E_{region}}{A_{region}} \right)$	Higher density is related to greater urbanisation.	Groot et al. (2014), Combes et al. (2008a, 2011)

Note 1: Region is area REGIC in the empirical analysis.

Source: Authors' own calculations.

Table A.4 - Definition and description of the main variables considered in the model of Chapter 2.

Variables	Definition	Level	Data source
Hourly wage	Monthly wage received in December divided by 4.3 times the number of weekly hours in the contract.	Individual	RAIS microdata
Age	Age at the end of the year.	Individual	RAIS microdata
Education attainment	Illiterate, incomplete primary school, complete primary school to incomplete high school, complete high school to incomplete college, college degree or more.	Individual	RAIS microdata
Skills - factors	See Annex 4.	Individual	RAIS microdata and Maciente (2012)
Firm size	Size of the firm in which the individual is working: up to 4 employees, 5 to 9, 10 to 19, 20 to 49, 50 to 99, 100 to 249, 250 to 499, 500 to 999, 1,000 employees or more.	Individual	RAIS microdata
Labour density in the formal sector	Total employment divided by the area (in km ²).	REGIC	RAIS - aggregated data
Area	Area in km ² .	REGIC	IPEADATA
Specialisation	See Annex 3.		RAIS - aggregated
Diversity	See Annex 3.		RAIS - aggregated
Competition	See Annex 3.		RAIS - aggregated
Altitude	Average altitude of the REGIC area obtained from the weighted mean of the altitude of each municipality (weight = area).	REGIC	Adapted from Naritomi et al. (2012)
Distance to the Equator line	Distance to the Equator measured as the absolute value of the latitude coordinate - obtained as a weighted average of the distance to the Equator line of the centroids of all municipalities that compose the REGIC area (weight = area).	REGIC	Adapted from Naritomi et al. (2012)
Instruments	Definition	Level	Data source
Population density in 1940	Population in 1940 for minimum comparable areas 1940-2000 is redistributed for minimum comparable areas 2000-2010 based on the share of the population of the latter in the former in 2000, aggregated by REGIC and divided by the area in km ² .	REGIC	IPEADATA
Distance to the coast	Distance to the coast (in 100 km) - obtained as a weighted average of the distance to the sea of the centroids of all municipalities that compose the REGIC (weight = area).	REGIC	Adapted from Naritomi et al. (2012)
Sugar	Proximity to the sugar cane boom (until 1760), calculated as the weighted average of the municipal index for all municipalities belonging to the REGIC area, ranging from 0 (more than 200 km) to 1, according to the proximity to the nearest municipality in the sugar cane areas (Naritomi et al., 2012) - weight = area.	REGIC	Adapted from Naritomi et al. (2012)
Gold	Proximity to the gold boom, calculated as the weighted average of the municipal index for all municipalities belonging to the REGIC area, ranging from 0 (more than 200 km) to 1 according to proximity to the nearest municipality in gold exploration areas (Naritomi et al., 2012) - weight =	REGIC	Adapted from Naritomi et al. (2012)
Coffee	Proximity to the coffee boom until 1886, calculated as the weighted average of the municipal index for all municipalities pertaining to the REGIC area, ranging from 0 (more than 200 km) to 1, according to the proximity to the nearest municipality in coffee areas (Naritomi et al., 2012)	REGIC	Adapted from Naritomi et al. (2012)

Source: Authors' own calculations.

Table A.5 - Aggregations of manufacturing and service industries according to technological knowledge intensity.³⁹

Aggregation	Industry code (CNAE 2 digits)	Industry name
Low technology manufacturing	15	Manufacture of food products and beverages
	16	Manufacture of tobacco products
	17	Manufacture of textiles
	18	Manufacture of wearing apparel
	19	Manufacture of leather and related products
	20	Manufacture of wood products
	22	Printing and reproduction of recorded media
Medium technology manufacturing	36	Manufacture of furniture, other manufacturing
	21	Manufacture of paper, cellulose and paper products
	23	Manufacture of coke, refined petroleum products, nuclear fuel and ethanol
	25	Manufacture of rubber and plastic products
	26	Manufacture of non-metallic mineral products
	27	Manufacture of basic metals
Medium-high and high technology manufacturing	28	Manufacture of fabricated metal products, except machinery and equipment
	24	Manufacture of chemical products
	29	Manufacture of machinery and equipment
	30	Manufacture of office machinery and computer equipment
	31	Manufacture of electrical equipment and materials, and machinery
	32	Manufacture of electronic material and communication equipment and products
	33	Manufacture of medical, precision and optical instrum., industrial automation equip. and chronometers
	34	Manufacture and assembly of motor vehicles, trailers and semi-trailers
Less knowledge-intensive services	35	Manufacture of other transport equipment
	55	Accommodation and food service activities
	60	Land transport
	63	Support activities for transportation and travel agencies
	70	Real estate activities
	71	Rental and leasing activities
	91	Activities of membership organisations
	92	Cultural, recreational and sports activities
	93	Social services
	95	Domestic services
Knowledge-intensive services	99	Activities of extraterritorial organisations and bodies
	61	Water transport
	62	Air transport
	64	Postal and courier activities and telecommunication
	65	Financial services
	66	Insurance and pension plans
	67	Auxiliary activities to financial services, insurance and pension plans
	72	Computer programming, consultancy and related activities
	73	Research and development
	74	Services mainly to companies
	80	Education
	85	Human health and social services
90	Urban cleaning, sewage and related activities	

Source: Authors' own calculations.

³⁹ This classification is based on the one proposed by Eurostat: <http://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:High-tech> (last accessed on 15/01/2015).

Table A.6 - Comparison of selected descriptive statistics of the full sample and the individuals with panel variation, 2012 (for the database 2004-2008-2012).

	S1		S2		S3		S4		S5	
	Manufacturing low-tech		Manufacturing medium-tech		Manufacturing high-tech		Services less-knowledge		Services high-knowledge	
	full sample	panel	full sample	panel	full sample	panel	full sample	panel	full sample	panel
ln(real hourly wage)	2.36	2.34	2.67	2.65	3.08	3.01	2.24	2.20	2.60	2.61
age	40.26	37.98	41.03	39.12	40.86	38.50	43.36	41.43	41.38	39.55
Education attainment										
Illiterate	0.5%	0.3%	0.3%	0.1%	0.1%	0.0%	0.4%	0.2%	0.2%	0.1%
Incomplete primary school	27.8%	20.8%	21.1%	13.4%	10.9%	5.4%	31.3%	19.9%	11.7%	8.1%
Comp. primary school - incompl. middle school	26.0%	23.7%	23.7%	20.6%	18.4%	12.3%	30.7%	28.9%	18.1%	14.5%
Complete middle school - incomplete college	39.2%	45.1%	46.2%	53.8%	50.2%	56.3%	33.0%	45.0%	41.3%	44.0%
College degree or more	6.6%	10.1%	8.8%	12.1%	20.5%	26.0%	4.7%	6.0%	28.8%	33.3%
Firm size										
Up to 4 employees	2.7%	3.1%	2.0%	2.5%	0.8%	1.2%	11.2%	8.6%	2.5%	2.5%
5 to 9	4.0%	4.9%	3.7%	4.2%	1.4%	1.7%	15.8%	10.8%	3.8%	4.2%
10 to 19	6.6%	7.4%	7.6%	7.8%	3.6%	4.2%	10.9%	10.9%	5.7%	6.1%
20 to 49	10.6%	12.0%	12.4%	13.2%	6.8%	8.7%	14.8%	16.7%	9.6%	9.8%
50 to 99	9.4%	10.5%	12.1%	13.9%	8.2%	10.2%	9.6%	12.0%	7.5%	7.9%
100 to 249	14.8%	16.1%	15.7%	19.1%	14.9%	16.2%	11.1%	14.3%	11.3%	12.4%
250 to 499	14.2%	15.2%	15.3%	15.6%	14.9%	15.2%	8.8%	10.4%	11.8%	12.4%
500 to 999	12.7%	13.9%	13.8%	11.7%	13.6%	14.9%	8.4%	8.1%	12.8%	12.7%
1,000 or more employees	25.0%	17.0%	17.3%	12.0%	35.8%	27.6%	9.6%	8.3%	35.1%	32.1%
Macro-region										
North	1.7%	2.5%	1.3%	1.8%	1.6%	2.4%	2.1%	2.7%	2.6%	2.4%
North-east	14.7%	13.7%	7.0%	7.5%	2.8%	2.8%	13.9%	13.9%	13.2%	11.1%
South-east	49.6%	41.0%	70.2%	68.7%	76.0%	74.4%	66.8%	62.5%	67.7%	68.5%
South	29.6%	36.1%	20.0%	19.9%	18.9%	19.4%	12.8%	15.2%	11.1%	13.0%
Centre-West	4.4%	6.7%	1.5%	2.1%	0.7%	1.0%	4.4%	5.7%	5.4%	5.0%
Number of observations	45,902	6,722	34,487	4,645	37,840	6,179	72,014	13,890	58,350	18,032

Source: Authors' own calculations.

Table A.7 - Results for the first-stage regressions, with and without individual fixed effects.

	First stage with OLS					First stage with fixed effects				
	S1.	S2.	S3.	S4. Services	S5. Services	S1.	S2.	S3.	S4. Services	S5. Services
	Manufacturing low-tech	Manufacturing medium-tech	Manufacturing high-tech	less- knowledge	high- knowledge	Manufacturing low-tech	Manufacturing medium-tech	Manufacturing high-tech	less- knowledge	high- knowledge
age	0.0497***	0.058***	0.0704***	0.035***	0.0534***	0.134	0.2843	0.1249	0.1063	0.1224
age2	-0.0005***	-0.0006***	-0.0007***	-0.0004***	-0.0005***	-0.0008***	-0.0009***	-0.0011***	-0.0006***	-0.0009***
illiterate (omitted)										
primary school - incomplete	0.111***	0.1241***	0.0837*	0.0081	-0.0209					
secondary school incomplete	0.1865***	0.2104***	0.1786***	0.0124	0.021					
tertiary education - incomplete	0.2959***	0.3505***	0.2968***	0.0862***	0.15***					
graduate + post-graduate	0.8391***	0.8103***	0.706***	0.6789***	0.6981***					
Factor 1 - Cognitive skills	0.2168***	0.2496***	0.2737***	0.1742***	0.3007***	0.0992***	0.0946***	0.0957***	0.0829***	0.1639***
Factor 2 - Maintenance and operation skills	0.0231***	0.0086***	-0.0058**	-0.0012	0.048***	-0.0041	0.0012	0.0016	-0.0016	-0.008***
Factor 3 - Assistance skills	-0.0249***	0.0048	-0.0211***	0.0031	0.0053*	-0.006	-0.0017	-0.0147***	0.0216***	0.01***
Factor 4 - Management skills	0.0827***	0.0255***	0.0424***	0.0181***	0.0319***	0.0274***	0.003	0.0185***	-0.0028	0.0083***
Factor 5 - Design and engineering skills	0.0118***	0.0041	0.0008	-0.0088***	-0.0381***	-0.0005	-0.016***	-0.0089***	-0.0061*	-0.0152***
Factor 6 - Transportation skills	-0.0174***	-0.0387***	-0.0135***	0.0626***	0.0173***	-0.0088***	-0.0155***	-0.0046	0.0387***	0.0083***
Factor 7 - Artistic skills	0.024***	-0.0068*	0.0127***	0.0459***	0.0112***	-0.0018	-0.0006	-0.0076**	0.0164***	0.0017
Factor 8 - Accuracy and automation skills	0.0204***	0.0122***	0.03***	0.01***	0.0329***	-0.0025	0.0042	0.0007	-0.0049**	0.0156***
Factor 9 - Supervised work skills	0.0383***	0.0447***	-0.001	0.0814***	0.039***	0.0206***	0.0187***	0.0019	0.0464***	0.0227***
Factor 10 - Teaching and social science skills	0.0299***	-0.024***	0.0205***	0.0264***	-0.0018	-0.0034	-0.0221***	-0.0022	0.0086**	-0.0006
Factor 11 - Physical strength	-0.0027	-0.026***	-0.0426***	-0.0385***	-0.0748***	0.0113***	-0.0067**	-0.0142***	-0.0249***	-0.0327***
Factor 12 - Telecommunication skills	0.0475***	0.0207***	0.0042*	0.0584***	0.0255***	0.0169***	0.0053	-0.0024	0.0138***	0.015***
Factor 13 - Independence skills	0.0049**	-0.0008	-0.011***	-0.0187***	0.0056***	-0.0051**	-0.0036	-0.0128***	0.0028*	0.0055***
Factor 14 - Natural science skills	-0.0344***	-0.0527***	-0.032***	-0.0356***	-0.0259***	-0.0231***	-0.0205***	-0.007**	-0.0145***	-0.0061**
Factor 15 - Attention skills	-0.0241***	-0.0275***	-0.0121***	0.0514***	0.0092***	-0.0088***	-0.0076***	-0.0116***	0.0167***	0.0081***
Factor 16 - On-the-job experience	-0.0295***	0.0082***	0.017***	0.031***	-0.0032	-0.0013	0.0062*	0.0047	0.015***	-0.0014
Factor 17 - Conflict management skills	0.0493***	0.0447***	0.0257***	-0.0734***	-0.003	0.0287***	0.0244***	0.0159***	-0.0249***	0.0007
Factor 18 - Team-work skills	-0.0285***	-0.0002	0.0456***	-0.0276***	-0.0279***	-0.0179***	-0.0055	0.0146***	-0.0181***	-0.0068**
Factor 19 - Sales skills	0.0125***	0.0128***	0.0189***	-0.0133***	-0.0056***	0.003	0.0043	-0.0053	0.0011	-0.0044**
Factor 20 - Monitoring and compliance skills	-0.0334***	-0.0195***	-0.0312***	-0.0473***	-0.0543***	-0.0067***	-0.0032	-0.0154***	-0.0171***	-0.018***
Factor 21 - Clerical skills	0.0353***	-0.0164***	0.0207***	0.0083***	0.0455***	0.0011	-0.0123***	0.0071**	-0.005**	0.0088***
firm size (up to 4 employees) - omitted										
firm size 5 to 9	0.0808***	0.0944***	0.1537***	0.0715***	0.1336***	0.0416***	0.0396***	0.0601***	0.0226***	0.1126***
firm size 10 to 19	0.1679***	0.1672***	0.2214***	0.0834***	0.2306***	0.0928***	0.091***	0.1171***	0.0457***	0.1704***
firm size 20 to 49	0.2505***	0.2172***	0.3192***	0.1572***	0.3335***	0.1471***	0.1374***	0.1776***	0.0873***	0.2351***
firm size 50 to 99	0.339***	0.3362***	0.405***	0.2415***	0.4089***	0.1938***	0.1988***	0.2389***	0.1308***	0.2724***
firm size 100 to 249	0.447***	0.4933***	0.5409***	0.3108***	0.48***	0.2403***	0.2772***	0.3192***	0.1816***	0.3146***
firm size 250 to 499	0.5272***	0.6069***	0.6128***	0.3593***	0.4877***	0.2858***	0.3429***	0.3652***	0.2221***	0.3361***
firm size 500 to 999	0.5282***	0.6885***	0.695***	0.346***	0.4889***	0.3113***	0.3808***	0.426***	0.244***	0.3405***
firm size 1000 or more employees	0.587***	0.8077***	0.8325***	0.3989***	0.4366***	0.3613***	0.4729***	0.488***	0.2349***	0.3185***
Observations (individuals * years)	137,706	103,461	113,520	216,042	175,050	137,706	103,461	113,520	216,042	175,050

*: $\alpha = 0.10$; **: $\alpha = 0.05$; ***: $\alpha = 0.01$;

Note 1: Additional controls are dummies for the iteration of REGIC area, year and industry, and a constant term.

Source: Authors' own calculations.

Table A.8 - Second-stage regression results for the spatial wage, with instrumental variables for ln(density) and OLS estimation in the first stage, 2004-2008 and 2008-2012.

		2nd stage with IV (1st stage with OLS) - 2004-2008					2nd stage with IV (1st stage with OLS) - 2008-2012				
		S1.	S2.	S3.	S4. Services	S5. Services	S1.	S2.	S3.	S4. Services	S5. Services
		Manufacturing	Manufacturing	Manufacturing	less-	high-	Manufacturing	Manufacturing	Manufacturing	less-	high-
		low-tech	medium-tech	high-tech	knowledge	knowledge	low-tech	medium-tech	high-tech	knowledge	knowledge
Composition 1		S1	S2	S3	S4	S5	S1	S2	S3	S4	S5
Urbanisation	ln(density)	0.0445***	0.0586***	0.0685***	0.0403***	0.0678***	0.0314***	0.0136	0.0550***	0.0198*	0.0687***
Diversity	ln(herfindhal)	0.0888***	0.0832***	0.105**	0.0613***	0.0169	0.0881***	0.126***	0.152***	0.108***	0.0104
Specialisation	ln(location quotient)	-0.0147*	0.00724	-0.0168	-0.000659	-0.0118	-0.0126*	-0.00663	0.00640	-0.00402	0.00930
Competition	ln(competition degree)	-0.0310***	-0.00330	-0.0457**	-0.0488***	-0.0321*	-0.0283***	-0.0445***	-0.0103	-0.0497***	0.0200
	R ²	0.369	0.360	0.243	0.322	0.500	0.401	0.344	0.244	0.321	0.456
Composition 2		S1	S2	S3	S4	S5	S1	S2	S3	S4	S5
Urbanisation	ln(density)	0.0788***	0.0814***	0.0906***	0.0602***	0.0808***	0.0617***	0.0370***	0.0835***	0.0410***	0.0806***
Diversity	ln(shannon entropy)	0.305***	0.302***	0.337**	0.228***	-0.00537	0.281***	0.248***	0.572***	0.291***	-0.0224
Specialisation	ln(specialisation degree)	0.0853***	0.0797***	0.0647***	0.0594***	0.0205	0.0688***	0.0391***	0.0965***	0.0479***	0.0147
Competition	ln(competition degree)	-0.0107*	-0.0119	-0.0301***	-0.0437***	-0.0208	-0.0119**	-0.0392***	-0.0161*	-0.0416***	0.0152
	R ²	0.383	0.369	0.249	0.330	0.500	0.413	0.343	0.257	0.327	0.454
Composition 3		S1	S2	S3	S4	S5	S1	S2	S3	S4	S5
Urbanisation	ln(density)	0.0736***	0.0680***	0.0891***	0.0423***	0.0701***	0.0501***	0.0203*	0.0824***	0.0226*	0.0665***
Diversity	ln(share 5 biggest sectors)	-0.227***	-0.340***	-0.184*	-0.306***	-0.140*	-0.278***	-0.358***	-0.279***	-0.329***	-0.166**
Specialisation	ln(specialisation degree)	0.0654***	0.0723***	0.0521***	0.0513***	0.0280**	0.0519***	0.0336***	0.0757***	0.0336***	0.0245**
Competition	ln(competition degree)	-0.00452	-0.00736	-0.0295***	-0.0456***	-0.0214	-0.00484	-0.0341***	-0.0159*	-0.0449***	0.0138
	R ²	0.383	0.376	0.248	0.331	0.501	0.415	0.349	0.252	0.324	0.457
Composition 4		S1	S2	S3	S4	S5	S1	S2	S3	S4	S5
Urbanisation	ln(density)	0.0777***	0.0813***	0.0935***	0.0588***	0.0775***	0.0580***	0.0312***	0.0857***	0.0371***	0.0770***
Diversity	ln(herfindhal)	0.158***	0.157***	0.141***	0.112***	0.0386	0.142***	0.164***	0.212***	0.150***	0.0259
Specialisation	ln(specialisation degree)	0.0933***	0.0882***	0.0683***	0.0624***	0.0296**	0.0740***	0.0518***	0.0956***	0.0535***	0.0222*
Competition	ln(competition degree)	-0.00774	-0.00929	-0.0279***	-0.0429***	-0.0220*	-0.0103*	-0.0358***	-0.0139*	-0.0409***	0.0141
	R ²	0.391	0.376	0.253	0.333	0.500	0.418	0.353	0.262	0.331	0.455
Observations	N (REGIC * sector * year)	3930	2620	1878	4673	4308	4195	2834	2234	5184	4910

*: $\alpha = 0.10$; **: $\alpha = 0.05$; ***: $\alpha = 0.01$;

Note 1: Additional controls are ln(area in square kilometres), ln(distance to the equator), ln(altitude), year dummies (2004 omitted), sector dummies, macro-region dummies (North omitted), constant.

Note 2: Instruments for ln(density) are ln(density in 1940), ln(distance to the coast), dummy of sugar cane production in colonial times, dummy of gold prospection in colonial times, dummy of coffee production in colonial and imperial times.

Source: Authors' own calculations.

APPENDIX A3 – ADDITIONAL TABLES FOR CHAPTER 3

Table A.9 - Definition and source of the main variables considered in the model.

Variables	Definition	Level	Data source
Hourly wage	Monthly wage received in December divided by 4.3 times the number of weekly hours in the contract.	Individual	RAISMIGRA microdata
Age	Age at the end of the year.	Individual	RAISMIGRA microdata
Tenure	Number of years working in the same firm	Individual	RAISMIGRA microdata
Education attainment	Less than 8 years of schooling, 8 to 10 years of schooling, 11 to 14 years of schooling, 15 years of schooling or more	Individual	RAISMIGRA microdata
Firm size	Size of the firm in which the individual is working: up to 4 employees, 5 to 9, 10 to 19, 20 to 49, 50 to 99, 100 to 249, 250 to 499, 500 to 999, 1,000 employees or more.	Individual	RAISMIGRA microdata
Labour density in the formal sector	Total employment divided by the area (in km ²).	REGIC	RAIS - aggregated data
Area	Area in km ² .	REGIC	IPEADATA
Sector	manufacture; food and accommodation; transport and communication; finance, insurance, pensions and other services; real estate, rents and services to companies; education; health and social services; personal services and other sectors.	Individual	RAISMIGRA microdata
Macroregion	North, North-east, South-east, South, Centre-West	REGIC	
1st time migration	Dummy that equals one in the year the individual has moved for the first time since 1995	Individual	RAISMIGRA microdata
2nd time migration	Dummy that equals one in the year the individual has moved for the second time since 1995, provided he moved once.	Individual	RAISMIGRA microdata

Source: Authors' own calculations.

APPENDIX A4 – ADDITIONAL TABLES FOR CHAPTER 4

Table A.10 – Definition and source of the main variables considered in the model of the wage curve, PME, PNAD and Census.

PME			
			(continues)
Variables	Definition	Level	Data source
Hourly wage	Monthly wage received in each month divided by 4.3 times the number of weekly hours in the contract, deflated by the local INPC.	Individual	PME panel
Unemployment rate	Individuals without a job who actively looked for a new position in the previous month over the economically active population.	Metropolitan region	PME
Age	Age at the month of the interview.	Individual	PME panel
Student	Dummy variable that equals one if the individual is currently studying and zero otherwise.	Individual	PME panel
Education attainment	Less than 1 year of schooling, 1 to 3 years of schooling, 4 to 7 years of schooling, 8 to 10 years of schooling and 11 years of schooling or more.	Individual	PME panel
Self reported black or brown	Individuals who identified themselves as black or brown.	Individual	PME panel
Sector of activity	Manufacture, construction, commerce and repair, financial services and real estate, other services, other sectors.	Individual	PME panel
Metropolitan region	Place in which the individual lives: Recife, Salvador, Belo Horizonte, Rio de Janeiro, São Paulo and Porto Alegre.	Individual	PME panel
Month	Month in which the interview was done.	Individual	PME panel
Year	Year in which the interview was done.	Individual	PME panel
PNAD			
Variables	Definition	Level	Data source
Hourly wage	Monthly wage received in October divided by 4.3 times the number of weekly hours in the contract, deflated by the national INPC.	Individual	PNAD microdata
Unemployment rate	Individuals without a job who actively looked for a new position in the previous month over the economically active population.	State	PNAD
Age	Age in October of each year.	Individual	PNAD microdata
Tenure	Number of years in the same job.	Individual	PNAD microdata
Per capita income ex-wage	Household per capita income ex-wage obtained by any household member.	Individual	PNAD microdata
Education attainment	Less than 1 year of schooling, 1 to 3 years of schooling, 4 to 7 years of schooling, 8 to 10 years of schooling, 11 to 14 years of schooling, and 15 years of schooling or more.	Individual	PNAD microdata
Student	Dummy variable that equals one if the individual is currently studying and zero otherwise.	Individual	PNAD microdata
Self reported black or brown	Individuals who identified themselves as black or brown.	Individual	PNAD microdata
Sector of activity	Agriculture, manufacture, construction, other industrial activities, commerce, services, auxiliary services, transport and communication, social services, other activities.	Individual	PNAD microdata
Occupation group	Technical, scientific, artistic or similar; administrative; agriculture, livestock, extractive activities; manufacture; commerce and auxiliary activities; transport, communication, services and others.	Individual	PNAD microdata
Head of the household	Dummy variable that equals one if the individual is the head of the household and zero otherwise.	Individual	PNAD microdata
People who have children under 15	Dummy variable that equals one if the individual has children under 15 and zero otherwise.	Individual	PNAD microdata

Demographic Census			(ends)
Variables	Definition	Level	Data source
Hourly wage	Monthly wage received in July divided by 4.3 times the number of weekly hours in the contract, deflated by the national INPC.	Individual	Census microdata
Unemployment rate	Individuals without a job who actively looked for a new position in the previous month over the economically active population.	MCA	Census
Density	Total population divided by the area (in km ²).	MCA	Census
Area	Area in km ² .	MCA	Census
Self reported black or brown	Individuals who identified themselves as black or brown.	Individual	Census microdata
Age	Age in July of each year.	Individual	Census microdata
Education attainment	Less than 7 years of schooling, 8 to 10 years of schooling, 11 to 14 years of schooling and 15 years of schooling or more.	Individual	Census microdata
Student	Dummy variable that equals one if the individual is currently studying and zero otherwise.	Individual	Census microdata
Sector of activity	Agriculture, manufacture, construction, other industrial activities, commerce, services, auxiliary services, transport and communication, social services, other activities.	Individual	Census microdata
Occupation group	Leaders; scientific, artistic or similar; technical level; administrative service; commerce and service; agriculture, livestock, extractive activities; manufacture.	Individual	Census microdata

Source: Author's own calculations.

Table A.11 – Number of observations in each month and year of the PME panel database.

	January	February	March	April	May	June	July	August	September	October	November	December
2003			607	1,038	1,050	1,061	1,146	1,097	1,135	1,086	1,228	1,152
2004	1,185	1,104	1,296	1,228	1,280	1,206	1,309	1,152	1,204	1,175	1,336	1,299
2005	1,286	1,137	1,424	1,269	1,339	1,226	1,349	1,276	1,308	1,272	1,359	1,398
2006	1,339	1,190	1,433	1,322	1,377	1,260	1,442	1,407	1,430	1,360	1,443	1,412
2007	1,322	1,246	1,413	1,370	1,367	1,401	1,476	1,430	1,417	1,384	1,551	1,456
2008	1,348	1,289	1,461	1,468	1,326	1,497	1,479	1,396	1,422	1,377	1,529	1,514
2009	1,454	1,365	1,392	1,513	1,430	1,413	1,552	1,382	1,439	1,407	1,472	1,418
2010	1,399	1,325	1,353	1,474	1,454	1,353	1,494	1,353	1,407	1,406	1,542	1,352
2011	1,315	1,267	1,415	1,424	1,393	1,355	1,462	1,357	1,462	1,368	1,590	1,441
2012	1,293	1,285	1,423	1,381	1,401	1,335	1,422	1,314	1,487	1,337	1,487	1,447
2013	1,330	1,289	1,497	1,474	1,515	1,320	1,497	1,436	1,563	1,402	1,523	1,496
2014	1,454	1,306	1,568	1,553	1,456	1,245	828	807	824	744	801	802
2015	757	662	786	770	655	593						

Source: PME panel.

Table A.12 – Complete estimation results of Models 1, 2, 5 and 6 of Tables 4.3 and 4.4.

	1) OLS			2) FE			5) DYNAMIC OLS			6) DYNAMIC FE		
	Total	Informal	Formal	Total	Informal	Formal	Total	Informal	Formal	Total	Informal	Formal
ln(unemployment rate)	-0.012 (0.016)	-0.102** (0.045)	-0.007 (0.017)	-0.001 (0.010)	0.055 (0.039)	-0.001 (0.010)	-0.008 (0.007)	-0.058** (0.023)	-0.003 (0.007)	-0.002 (0.009)	0.021 (0.035)	-0.003 (0.009)
ln(hourly wage) _{t-1}							0.875*** (0.002)	0.840*** (0.006)	0.877*** (0.002)	0.422*** (0.003)	0.447*** (0.010)	0.399*** (0.003)
Student	0.120*** (0.005)	0.077*** (0.011)	0.134*** (0.005)	0.003 (0.005)	-0.007 (0.016)	0.012** (0.005)	0.017*** (0.003)	0.018*** (0.007)	0.018*** (0.003)	0.006 (0.004)	-0.002 (0.015)	0.011** (0.005)
Age	0.075*** (0.001)	0.063*** (0.002)	0.072*** (0.001)				0.008*** (0.000)	0.009*** (0.001)	0.007*** (0.000)			
Age ²	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)				-0.000*** (0.000)	-0.000*** (0.000)	-0.000*** (0.000)			
Self-reported black or brown	-0.228*** (0.003)	-0.194*** (0.008)	-0.232*** (0.003)				-0.024*** (0.001)	-0.025*** (0.004)	-0.024*** (0.002)			
Education level (less than 1 year of schooling omitted)												
1 to 3 years of schooling	0.061*** (0.011)	0.052* (0.029)	0.063*** (0.012)				0.008 (0.005)	0.021 (0.014)	0.005 (0.006)			
4 to 7 years of schooling	0.181*** (0.010)	0.158*** (0.027)	0.182*** (0.011)				0.021*** (0.005)	0.028** (0.012)	0.019*** (0.005)			
8 to 10 years of schooling	0.327*** (0.010)	0.272*** (0.027)	0.330*** (0.011)				0.043*** (0.005)	0.040*** (0.013)	0.042*** (0.005)			
11 years of schooling or more	0.753*** (0.010)	0.618*** (0.028)	0.759*** (0.011)				0.094*** (0.005)	0.096*** (0.013)	0.092*** (0.005)			
Sector of activity (industry omitted)												
Construction	-0.061*** (0.006)	-0.135*** (0.014)	-0.009 (0.006)	-0.029*** (0.006)	-0.073*** (0.023)	-0.009 (0.007)	-0.011*** (0.002)	-0.028*** (0.007)	-0.003 (0.003)	-0.016*** (0.005)	-0.044** (0.020)	-0.004 (0.006)
Commerce and repair	-0.221*** (0.004)	-0.166*** (0.012)	-0.216*** (0.004)	-0.055*** (0.004)	-0.052*** (0.017)	-0.043*** (0.005)	-0.033*** (0.002)	-0.027*** (0.006)	-0.032*** (0.002)	-0.041*** (0.004)	-0.020 (0.015)	-0.035*** (0.004)
Finance, real estate, serv. to companies	-0.046*** (0.005)	0.130*** (0.016)	-0.059*** (0.005)	-0.022*** (0.005)	0.015 (0.023)	-0.026*** (0.005)	-0.008*** (0.002)	0.017** (0.008)	-0.009*** (0.002)	-0.017*** (0.005)	0.018 (0.020)	-0.022*** (0.005)
Other services	-0.108*** (0.004)	-0.069*** (0.013)	-0.101*** (0.005)	-0.010** (0.005)	0.010 (0.022)	-0.008 (0.006)	-0.015*** (0.002)	-0.014** (0.006)	-0.013*** (0.002)	-0.005 (0.005)	0.016 (0.020)	-0.005 (0.005)
Other activities	-0.274*** (0.018)	-0.285*** (0.034)	-0.232*** (0.021)	-0.067*** (0.017)	-0.088 (0.054)	-0.048** (0.019)	-0.036*** (0.009)	-0.052** (0.022)	-0.026** (0.010)	-0.050*** (0.015)	-0.051 (0.048)	-0.040** (0.018)
Metropolitan region	Yes	Yes	Yes				Yes	Yes	Yes			
Year	Yes	Yes	Yes									
Month	Yes	Yes	Yes				Yes	Yes	Yes			
Individual fixed effects				Yes	Yes	Yes				Yes	Yes	Yes
N	193,934	23,553	170,381	193,934	23,553	170,381	193,934	23,553	170,381	193,934	23,553	170,381
Adjusted R ²	0.355	0.386	0.335				0.858	0.832	0.856			
R ² within				0.0221	0.0255	0.0192				0.2002	0.2347	0.1793
R ² between				0.046	0.0824	0.0357				0.8116	0.7522	0.8017

Source: Author's own elaboration.

Table A.13 – Complete estimation results of Models 3, 4, 7 and 8 of Tables 4.3 and 4.4.

	3) 2SLS			4) FE-2SLS			7) 2SLS DYNAMIC			8) FE-2SLS DYNAMIC		
	Total	Informal	Formal	Total	Informal	Formal	Total	Informal	Formal	Total	Informal	Formal
ln(unemployment rate)	-0.012 (0.021)	-0.161*** (0.062)	-0.003 (0.023)	-0.004 (0.013)	0.037 (0.053)	0.003 (0.013)	-0.007 (0.010)	-0.078** (0.032)	0.000 (0.010)	-0.001 (0.012)	-0.005 (0.047)	0.005 (0.012)
ln(hourly wage) _{t-1}							0.875*** (0.002)	0.840*** (0.006)	0.877*** (0.002)	0.422*** (0.003)	0.447*** (0.010)	0.399*** (0.003)
Student	0.120*** (0.005)	0.077*** (0.011)	0.134*** (0.005)	0.003 (0.005)	-0.007 (0.016)	0.012** (0.005)	0.017*** (0.003)	0.018*** (0.007)	0.018*** (0.003)	0.006 (0.004)	-0.002 (0.015)	0.011** (0.005)
Age	0.075*** (0.001)	0.063*** (0.002)	0.072*** (0.001)				0.008*** (0.000)	0.009*** (0.001)	0.007*** (0.000)			
Age ²	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)				-0.000*** (0.000)	-0.000*** (0.000)	-0.000*** (0.000)			
Self-reported black or brown	-0.228*** (0.003)	-0.194*** (0.008)	-0.232*** (0.003)				-0.024*** (0.001)	-0.025*** (0.004)	-0.024*** (0.002)			
Education level (less than 1 year of schooling omitted)												
1 to 3 years of schooling	0.061*** (0.011)	0.052* (0.029)	0.063*** (0.012)				0.008 (0.005)	0.021 (0.014)	0.005 (0.006)			
4 to 7 years of schooling	0.181*** (0.010)	0.158*** (0.027)	0.182*** (0.011)				0.021*** (0.005)	0.027** (0.012)	0.019*** (0.005)			
8 to 10 years of schooling	0.327*** (0.010)	0.272*** (0.027)	0.330*** (0.011)				0.043*** (0.005)	0.040*** (0.013)	0.042*** (0.005)			
11 years of schooling or more	0.753*** (0.010)	0.617*** (0.028)	0.759*** (0.011)				0.094*** (0.005)	0.096*** (0.013)	0.092*** (0.005)			
Sector of activity	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Metropolitan region	Yes	Yes	Yes				Yes	Yes	Yes			
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month	Yes	Yes	Yes				Yes	Yes	Yes			
Individual fixed effects				Yes	Yes	Yes				Yes	Yes	Yes
F test instrument (ln(unemp. rate) _{t-12})	199,736	23,828	175,604				199,735	23,817	175,601			
R ² within - 1st stage				0.766	0.762	0.765				0.766	0.762	0.765
R ² between - 1st stage				0.951	0.947	0.950				0.951	0.946	0.950
N	193,934	23,553	170,381	193,934	23,553	170,381	193,934	23,553	170,381	193,934	23,553	170,381
Adjusted R ²	0.355	0.386	0.335				0.858	0.832	0.856			
R ² within				0.022	0.026	0.019				0.200	0.235	0.179
R ² between				0.046	0.084	0.036				0.812	0.753	0.802

Source: Author's own elaboration.

Table A.14 – Complete estimation results of Table 4.6.

	Total			Rural			Urban		
	1a) Total	2a)	3a) Informal	1b) Total	2b) Formal	2c) Informal	3a) Total	3b) Formal	3c) Informal
ln(unemployment rate)	0.296***	-0.010	0.390***	-0.076***	-0.147***	-0.098***	0.295***	0.010	0.362***
Student	-0.786***	-0.661***	-0.726***	0.684***	-1.584***	0.369***	-0.613***	-0.583***	-0.598***
Age	0.180***	0.014***	0.312***	0.004	-0.009***	-0.043***	0.167***	0.013***	0.274***
Age ²	-0.003***	-0.001***	-0.004***	0.000***	-0.001***	0.001***	-0.003***	-0.000***	-0.004***
Tenure	0.022***	0.013***	0.020***	-0.001	-0.004***	0.008***	0.024***	0.018***	0.021***
Tenure ²	-0.000***	-0.000***	-0.000***	-0.000***	0.000	-0.000***	-0.000***	-0.001***	-0.000***
Self-reported black or brown	0.063***	0.035***	0.046***	-0.358***	0.577***	-0.188***	0.001	-0.008**	-0.007***
Education level (less than 1 year of schooling omitted)									
1 to 3 years of schooling	0.165***	-0.087***	0.458***	0.261***	-0.605***	-0.007	0.196***	-0.003	0.349***
4 to 7 years of schooling	0.287***	-0.242***	0.905***	0.493***	-1.054***	0.024**	0.333***	-0.114***	0.695***
8 to 10 years of schooling	0.543***	-0.344***	1.473***	0.558***	-1.277***	-0.070***	0.551***	-0.208***	1.140***
11 to 14 years of schooling	0.877***	-0.587***	2.080***	0.819***	-2.574***	-0.125***	0.859***	-0.404***	1.663***
15 years of schooling or more	1.003***	-0.180***	1.993***	1.764***	-2.661***	0.817***	1.063***	0.007	1.697***
Sector of activity (Other activities omitted)									
Agriculture	-0.234***	-0.267***	-0.181***	0.133***	0.096***	0.065***	-0.192***	-0.228***	-0.127***
Manufacture	-0.039***	-0.122***	-0.050***	0.158***	0.216***	0.129***	-0.032***	-0.107***	-0.041***
Construction	-0.108***	-0.156***	-0.035***	0.170***	0.093***	0.189***	-0.107***	-0.159***	-0.027***
Other industrial activities	0.075***	-0.121***	0.081***	0.046***	0.059**	-0.019	0.089***	-0.089***	0.090***
Commerce	-0.176***	-0.201***	-0.184***	0.016	0.033	0.014	-0.170***	-0.193***	-0.178***
Services	-0.187***	-0.249***	-0.129***	0.177***	0.162***	0.070***	-0.181***	-0.242***	-0.125***
Auxiliary services	-0.114***	-0.134***	-0.105***	0.200***	0.176***	0.218***	-0.111***	-0.129***	-0.101***
Transport and communication	-0.005	-0.099***	0.023***	0.100**	0.055	0.064	-0.003	-0.095***	0.023***
Social	-0.188***	-0.202***	-0.179***	0.133***	0.193***	0.054	-0.186***	-0.196***	-0.178***
Occupation group (Transport, communication, services and others omitted)									
Technical, scientific, artistic or similar	0.390***	0.296***	0.455***	-0.121**	-0.104	-0.189***	0.382***	0.287***	0.438***
Administrative	0.182***	0.111***	0.203***	-0.540***	-0.388***	-0.462***	0.176***	0.109***	0.194***
Agriculture, livestock, extractive activities	-0.156***	-0.181***	-0.046***	-0.343***	-0.289***	-0.340***	-0.155***	-0.202***	-0.061***
Manufacture	0.040***	-0.025***	0.081***	-0.367***	-0.309***	-0.344***	0.041***	-0.024***	0.081***
Commerce and auxiliary activities	0.040***	-0.033***	0.097***	-0.349***	-0.233***	-0.393***	0.041***	-0.033***	0.094***
Head of the household	0.286***	-0.025***	0.461***	0.011	-0.321***	-0.208***	0.248***	-0.033***	0.386***
People who have children under 15	0.036***	0.027***	0.023***	-0.152***	0.284***	-0.076***	0.015***	-0.001	0.011***
Lambda	2.619***	2.363***	3.233***	-2.185***	4.823***	-1.408***	2.161***	2.273***	2.721***
State	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R ²	0.586	0.500	0.533	0.451	0.321	0.443	0.572	0.493	0.541
N (unweighted)	560,543	198,023	362,520	58,774	37,626	21,148	501,769	160,397	341,372

Source: Author's own elaboration.

Table A.15 – First stage estimation results of Table 4.7.

	Total (rural and urban)			Rural			Urban		
	Total	Formal	Informal	Total	Formal	Informal	Total	Formal	Informal
Student	-0.435***	-0.303***	-0.337***	-0.500***	-0.415***	-0.361***	-0.421***	-0.273***	-0.337***
Age	0.068***	-0.017***	0.110***	0.022***	-0.011***	0.066***	0.072***	-0.022***	0.112***
Age ²	-0.001***	0.000**	-0.002***	-0.001***	-0.000***	-0.001***	-0.001***	0.000***	-0.002***
Self-reported black or brown	0.075***	0.046***	0.055***	0.213***	0.169***	0.134***	0.049***	0.027***	0.036***
Education level (less than 1 year of schooling omitted)									
1 to 3 years of schooling	0.042***	-0.083***	0.160***	-0.119***	-0.182***	0.091***	0.088***	-0.034***	0.152***
4 to 7 years of schooling	0.045***	-0.215***	0.295***	-0.220***	-0.322***	0.127***	0.102***	-0.151***	0.266***
8 to 10 years of schooling	0.127***	-0.317***	0.491***	-0.205***	-0.399***	0.280***	0.173***	-0.256***	0.441***
11 to 14 years of schooling	0.220***	-0.523***	0.679***	-0.291***	-0.760***	0.427***	0.266***	-0.444***	0.624***
15 years of schooling or more	-0.048***	-0.588***	0.358***	-0.607***	-0.859***	-0.016	0.003	-0.502***	0.317***
Head of the household	0.060***	-0.123***	0.141***	0.095***	-0.127***	0.308***	0.047***	-0.135***	0.125***
People who have children under 15	-0.035***	-0.049***	-0.013***	0.107***	0.058***	0.104***	-0.064***	-0.069***	-0.038***
State	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo R ²	0.062	0.054	0.104	0.072	0.064	0.132	0.063	0.051	0.092
N (unweighted)	1,493,518	1,493,518	1,493,518	172,320	172,320	172,320	1,303,187	1,303,187	1,303,187

Source: Author's own elaboration.

Table A.16 – First stage estimation results of Table 4.11.

	Total	Formal	Informal
Student	0.016*** (0.003)	0.065*** (0.004)	-0.073*** (0.006)
Age	0.055*** (0.001)	0.049*** (0.001)	0.051*** (0.001)
Age ²	-0.001*** (0.000)	-0.000*** (0.000)	-0.001*** (0.000)
Self-reported black or brown	-0.100*** (0.002)	-0.105*** (0.002)	-0.084*** (0.004)
Education level (less than 7 years of schooling omitted)			
8 to 10 years of schooling	0.187*** (0.002)	0.174*** (0.003)	0.162*** (0.005)
11 to 14 years of schooling	0.395*** (0.003)	0.371*** (0.003)	0.370*** (0.006)
15 years of schooling or more	0.990*** (0.006)	0.974*** (0.007)	0.919*** (0.016)
Head of the household	0.167*** (0.002)	0.152*** (0.002)	0.167*** (0.004)
Occupational group	Yes	Yes	Yes
Sector of activity	Yes	Yes	Yes
MCA-year dummy	Yes	Yes	Yes
Adjusted R ²	0.572	0.547	0.490
N (unweighted)	459,441	300,280	159,161

Source: Author's own elaboration.