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

FORESTS AND CITIES: ESSAYS ON URBAN GROWTH AND DEVELOPMENT IN THE BRAZILIAN AMAZON

Sergio André Castelani

Advisor: Danilo Camargo Igliori

SÃO PAULO
2013

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

Prof. Dr. Reinaldo Guerreiro
Dean of the School of Economics, Business and Accountability

Prof. Dr. Joaquim José Martins Guilhoto
Head of the Departament of Economics

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

SERGIO ANDRÉ CASTELANI

FORESTS AND CITIES: ESSAYS ON URBAN GROWTH AND DEVELOPMENT IN THE BRAZILIAN AMAZON

Doctoral dissertation submitted to the Department of Economics of the School of Economics, Business and Accounting at the University of São Paulo in partial fulfillment of the requirements for the degree of Doctor of Sciences

Advisor: Danilo Camargo Igliori

Rectified Version

Original Version available at Faculdade de Economia, Administração e Contabilidade - USP

SÃO PAULO

2014

FICHA CATALOGRÁFICA

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

Castelani, Sergio André
Forests and cities: essays on urban growth and development in the Brazilian Amazon / Sergio André Castelani. – São Paulo, 2013.
231 p.

Tese (Doutorado) – Universidade de São Paulo, 2013.
Orientador: Danilo Camargo Igliori.

1. Econometria 2. Urbanização 3. Insumo-produto 4. Amazônia - Brasil 5. Migração I. Universidade de São Paulo. Faculdade de Economia, Administração e Contabilidade. II. Título.

CDD – 330.015195

AGRADECIMENTOS

À minha família por me dar todo o apoio hoje e sempre, em todas as minhas decisões. Em especial aos meus pais Leonice Castelani e Sergio Castelani por todo o suporte e amor durante toda minha vida. Às minhas irmãs Priscila e Cristiane Castelani, por serem pessoas incríveis que eu adoro.

À minha mulher Sara Santos, pelo amor, carinho e paciência nos momentos mais difíceis.

A meu orientador e amigo Danilo Igliori, pela ajuda e paciência, e pelos anos de confiança em mim depositados.

Ao professor e coorientador Joaquim Guilhoto, por toda a ajuda, paciência e suporte prestados. À minha amiga e parceira Ana Barufi, pela parceria importantíssima e ajuda fundamental.

Ao meu grande amigo Thiago "Jesus" Fonseca, por ser amigo, e ter me ajudado muito nas horas difíceis. E pelas infinitas conversas e parcerias musicais. Ao amigo e pesquisador Rafael Barbieri, pelo companheirismo, parcerias e sugestões. À Denise Imori, também pela amizade e cooperação.

À FAPESP, pelo suporte financeiro fundamental para que esta tese fosse possível.

Ao professor André Chagas, pelas sugestões e discussões a respeito deste trabalho.

Ao Eustáquio Reis, pelas idéias e suporte nos congressos. Aos professores Ricardo Abramovay, Carlos Roberto Azzoni e Gabriel Madeira, por toda a ajuda e comentários.

Ao pessoal do projeto RAS (Rede Amazônia Sustentável), em especial ao professor Toby Gardner, pelas sugestões e pela motivação sugerida.

Aos meus amigos e colegas de curso Thiago Angelis, Henrique Vinhais, Paula Pereda.

Aos meus parceiros de música e amigos Pedro Silva, Guilherme Momesso, Fernanda Peviani, Patrick Maia, Jara Arraiz, Thiago Colli, Rodrigo Mozambique, Claudião, Pedro Bezerra, Gustavo Lamounier e Sansão, por me permitirem fazer parte deste mundo incrível da música.

Aos meus grandes amigos pessoais e sempre companheiros Guilherme Raffaelli, Raone Costa, Rodrigo Campos, Ari Pinheiro, Chicão Poltronieri e Glauco Eiji, pela amizade e companheirismo de sempre.

“Think global, act local”
John Lennon

ABSTRACT

The Brazilian Amazon has been undergoing a process of population growth and urbanization in recent decades. Its urban population increased from 42% to 71% between 1960 and 2010, and in the last decade, its overall population grew around 21%. Such rises bring important consequences not only economically, but also in environmental terms, especially considering that the largest remaining rainforest in the world is located within this region. Nevertheless, this scenario is still poorly addressed by literature. Bearing this in mind, this thesis aims to examine some economic and environmental aspects related to this context, and is specifically divided into three essays.

Firstly, a spatial econometric approach is implemented, based on the framework of spatial economics models, in order to investigate whether this process of urbanization has been generating local economic growth and development.

In the second essay, aiming to measure the environmental impacts of such population growth and urbanization, an Interregional Input-Output model is built, for the year of 2004, merging data regarding the productive structure and land use transition in the Brazilian Amazon. Specifically, this method allows the measurement of how much local deforestation may be attributed to the consumption of goods and services demanded by households living within the region, considering all direct and indirect production of inputs and outputs in every region of Brazil. Moreover, in order to capture the effects of local urbanization, special focus is given to the demands of the families living within the five Brazilian Amazon metropolitan regions.

Finally, in the third essay, given that population growth and urbanization processes are directly related to migration flows, an econometric model was implemented in order to investigate the determinants of immigration and emigration flows between the Brazilian Amazon and the rest of Brazil. This estimation allowed the comparison of the reasons that have been driving the exit and the entry of individuals in the region. Such methodology makes use of estimators which take into account econometric problems commonly attributed by literature to the modeling of migration flows, such as the sample selection issue regarding the potential differences in skills between migrants and non-migrants.

As main results, we find evidence that these local processes of urbanization and population growth have been causing a "trade -off" in the region: on the one hand, such processes seem to be driving local economic growth and development, but on the other, they also tend to increase regional deforestation. Furthermore, we find evidence that although the immigration and emigration flows of the Brazilian Amazon have been currently well-balanced, local vegetative growth still has been fueling such population growth and urbanization. Moreover, we find that the motivations which lead individuals to immigrate to the Brazilian Amazon are quite distinct from those that encourage people to leave it: whereas the former seek immediate higher levels of real income, the latter seem to move to cities with higher levels of education.

RESUMO

A Amazônia Brasileira vem passando por um processo de crescimento populacional e urbanização nas últimas décadas. Sua população urbana passou de 42% para 71% entre 1960 e 2010, e na última década o crescimento populacional da região foi de 21%. Tais processos trazem consigo consequências importantes não apenas em termos econômicos, mas também em termos ambientais, especialmente se considerando que a maior floresta tropical do mundo está localizada nesta região. No entanto, este aspecto é ainda pouco estudado pela literatura econômica.

Tendo isto em vista, este trabalho se propõe a estudar alguns aspectos econômicos e ambientais relacionados este quadro, divididos em três ensaios.

No primeiro, são utilizados métodos de econometria espacial, baseados em modelos de economia espacial, para investigar se este processo de urbanização tem causado crescimento e desenvolvimento econômico local.

No segundo ensaio, a fim de medir os impactos ambientais do crescimento populacional e da urbanização locais, informações a respeito da estrutura produtiva e do uso do solo na Amazônia Brasileira são cruzadas em um modelo Inter-regional de Insumo-Produto, que mede o quanto do desmatamento anual da floresta Amazônica é devido ao consumo de bens e serviços por parte das famílias que vivem na região, considerando toda a cadeia produtiva brasileira. A fim de capturar os efeitos da urbanização, foco especial é dado às demandas das famílias que vivem nas cinco regiões metropolitanas da Amazônia Brasileira.

Por fim, no terceiro ensaio, devido ao fato de que urbanização e crescimento populacional são processos diretamente relacionados à migração de indivíduos, é desenvolvida uma metodologia econométrica que investiga os determinantes dos fluxos imigratórios e emigratórios entre a Amazônia Brasileira e o restante do Brasil, no intuito de comparar os motivos que causam a entrada e saída de pessoas na região. Tal metodologia faz uso de estimadores que levam em conta problemas comumente atribuídos pela literatura na modelagem de fluxos migratórios, como a questão da seleção amostral relativa a potenciais diferenças de habilidade entre populações de migrantes e de não migrantes.

Como principais resultados, encontramos evidências de que tais processos de urbanização e crescimento populacional têm causado um "trade-off" econômico-ambiental na região: por um lado tais processos têm promovido o desenvolvimento e crescimento econômico local, mas por outro lado, eles também vêm causando aumento dos níveis do desmatamento regional. Além disso, encontramos evidência de que embora os fluxos imigratórios e emigratórios relativos à Amazônia encontrem-se equilibrados atualmente, o crescimento vegetativo local ainda alimenta os processo de crescimento populacional e urbanização. Finalmente, nossos resultados apontam que as motivações que levam indivíduos a se mudarem para a Amazônia são distintas daquelas que levam pessoas a se deixarem a região: enquanto os primeiros buscam maiores níveis imediatos de renda real, os últimos buscam mais claramente se mudarem para municípios com melhores níveis educacionais.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF MAPS

LIST OF TABLES

1. INTRODUCTION

Brazilian Amazon holds the largest remaining tropical forest in the world, which renders the region a global importance to current and future generations, and its conservation is undoubtedly fundamental to global ecological equilibrium. It is widespread knowledge that the region's biodiversity is immense. Also, recent literature found evidence supporting that deforestation in Brazilian Amazon is highly correlated to global warming and climate change. Imori et al. (2011), for example, using an Input-Output model, calculate that deforestation in Brazilian Amazon was responsible for about 58% of total greenhouse gases emissions in Brazil, and about 2% of total global emissions, in the year of 2004. Economically, the region hosts the current agricultural and cattle raising frontiers in Brazil, which holds capital-intensive cultivation of soybeans, whose exports have been boosting the recent surpluses in the Brazilian trade balance (Morton et al. 2006; Vera-Diaz et al. 2009). Politically, it is considered an area of strategic importance, due to its natural resources availability.

In parallel to such environmental and political-economic importance, a less known aspect about the region is that within the last two decades, it has been going through a considerable process of urbanization and population growth. Census data shows that its population grew by more than 4 million inhabitants throughout the last decade, representing an increase rate of 21.08%, which stands for twice the rate for the rest of Brazil. Such expressive growth is flourishing over a process of intense regional urbanization, as urban population grew by 3.7 million inhabitants from 2000 to 2010, representing an increase of 27.35% of this share of the population in the decade. Still, it is reasonable to believe that such urbanization has not yet reached its peak, as the share of urban population within Brazilian Amazon is about 73%, whereas it reaches the rate of 86.11% in the rest of the country, where occupation and urbanization are more consolidated historically (see Becker, 2013).

Furthermore, such urbanization process is not happening only among the region's biggest cities. São Felix do Xingú, which is a medium-sized municipality located at southern Pará, faced a population increase of 264% in the last decade, jumping from 34,628 to 91,340 inhabitants. This growth was mostly based on urbanization, as its urban population went from 36% of the city's population in 2000, to almost 50% in 2010. This is only one example among several other medium and small municipalities within the region which are going through the same process.

Despite these recent urban and population growth, however, the region is still among the less socio-economically developed areas of Brazil, only behind the Northeast macro-region. The recently calculated HDI-M (Human Development Index per municipality) for the year of 2010, calculated by PNUD (the United Nations Development Program), presented an average of 0.660 for Brazilian Amazon as a whole, whereas the average index for Brazil is 0.727. For the Southeast Brazilian macro-region, which is the most developed of the country, this index assumes the value of 0.764. Considering that even the overall Brazilian HDI level is still relatively low when compared to developed countries, this evidences that the Brazilian Amazon region still has a gap of social development and economic growth to fulfill.

Given this picture concerning the region's environmental importance, its strategic political relevance, and its transforming socio-economic structure, these processes of population growth and urbanization bring out important and interrelated questions: 1) What are the environmental impacts related to such urbanization ad population growth, especially in terms of deforestation? 2) Is this urbanization causing economic growth and development in Brazilian Amazon? 3) How much local urbanization and population growth are being determined by immigration and emigration flows recently, and how differences between determinants of these two kinds of flows may be affecting the urban and demographic structure of Brazilian Amazon?

These are relevant questions in regional, national and global scale: in socio-economic terms, their answers may contribute to bring economic development and growth to of one of the poorest regions in Brazil. Environmentally they are fundamental in worldwide terms, as deforestation impacts of this process sprawl throughout the globe, due to its collective nature.

Given such relevance, these are the questions that this work proposes to address and bring its contribution. Specifically, this thesis tries to address these processes of urbanization and population growth taking place in Brazilian Amazon from a perspective which tries to embrace several interrelated aspects.

In order to do so, this thesis is structured in four chapters, with the first being this introduction, which also sums up the main results found throughout the study. In chapter 2, a spatial econometric analysis is implemented in order to identify recent relationships between urbanization and economic development and growth within region. In Chapter 3, we implement an inter-regional Input-Output model in order to measure the relative demand driven impacts of local population and urbanization over the region's deforestation, mainly focusing on the impacts caused by the consumption of goods and services from families living within the Brazilian Amazon metropolitan areas. In chapter 4, we try to investigate how these urban changes may be related to migration flows regarding Brazilian Amazon, by trying to identify and compare what has been causing immigration flows from other parts of Brazil towards the region, and what has been driving emigration flows from Brazilian Amazon to other parts of Brazil. Also, we try to measure how much of the region's recent population increase may be attributed to migration. In order to do so, we firstly measure the relative size and the recent evolution of these flows. Secondly, we implement econometric methods which allow the comparison of the possible determinants of emigration and immigration. Such econometric methodology tries to account for several estimation problems regarding the use of migration flows as the dependent variable. Finally, in chapter 5, we conclude by summarizing and discussing the main implications of the results from all essays, relating them to literature on the subject. In this final chapter, we also propose some major policy guidelines, also briefly discussing their possible implications and impacts.

Our main results show that even though such urbanization and population growth are still ongoing processes within the region, they already exert relevant impacts in the region's economy and environment. Specifically, we find evidence that in the year of 2004, even though Brazilian Amazon's population represented only 13% of the national total, the consumption vectors from local families could be held responsible for around 30% of the total yearly deforestation taking place within the region, when we account for both direct and indirect production designed to attend such demand. The specific demand vector from families who live within the Amazonian Metropolitan Regions is responsible for more than a half of this 30%, even though only 25% of the local population actually live in these areas. This result may be attributed to an increase in local families' individual consumption level, generally caused by urbanization (especially in metropolitan regions). Moreover, considering that since 2004 both urbanization and population continued to grow, it seems likely that such deforestation impacts may have become even higher at the present date.

Furthermore, we have found that such growth of local urbanization also seem to be driving economic development and growth throughout the region. Specifically, our empirical results suggest that higher initial levels of urbanization of Amazon municipalities in the year 2000 have caused employment, per capita GDP and municipal HDI to grow between 2000 and 2010. Also, we found evidence that cities which presented the highest levels of development in Brazilian Amazon in 2010 are also the ones with the higher urbanization rates. Moreover, we find evidence that the growth of local population and urban centers are contributing to the emergence of an endogenous growth process of local markets, or put in other words, they are increasing the relative importance of internal markets as foundations for the region's economic development dynamic, even though we also find evidence that external markets still remain as important drivers of growth and development of Brazilian Amazon.

Therefore, gathering results concerning environmental and economic impacts of local urbanization brings evidence regarding the existence of an undesirable trade-off: on the one hand, urbanization and population growth seem to be promoting economic progress in one of the poorest regions of Brazil, but on the other, they are also driving higher levels of local deforestation. In terms of policy implications, this means that local sustainable growth will present itself as an environmental dilemma to policy makers in the years to come, especially considering that urbanization is still an undergoing process within the region. Thus, this suggests that new solutions for forest conservation will have to be created and implemented, along with the strengthening of conservation policies such as the increase of legal reserves or the creation of national parks and other conservation areas.

To which concerns the results regarding the role of migration flows on local population growth and urbanization, we firstly have found evidence that although the number of migrants of Brazilian Amazon (both immigrants and emigrants) have been slightly declining since 1995, and even though Amazon immigration and emigration are practically in balance nowadays, population and urbanization continued to grow during the 2000 decade. This suggests that such the recent population growth must be mostly based on vegetative growth, that is, on the reproduction of the population who have already been living within Brazilian Amazon borders throughout the last decades. Nevertheless, we have estimated and compared the drivers of Amazon immigration and emigration flows, in order to be able to trace possible future population trends, and results showed some relevant differences between the determinants of these two kinds of flows.

Specifically, our results suggest that Amazon immigration flows are mostly driven by young population coming from both more and less urbanized cities of Brazil, in majority located at the poorest Northeast macro-region of the country, and moving towards the more highly urbanized centers of Amazon. Moreover, according to our findings, short run real income differentials between destinations and origins act as one of the main drivers of such immigration, as most of the economic theories would predict. On the other hand, Amazon emigration flows are also mostly composed by young individuals moving to highly urbanized areas of Brazil. However, they also tend to leave from more urbanized cities of Brazilian Amazon. Furthermore, in opposition to Amazon immigration flows results, Amazon emigration does not seem to be driven by short run real income differentials between origins and destinations, but instead, our results point that such emigration flows are mostly driven by the differentials in basic and superior education levels of the migrants' origins and destinations. Such result does not necessarily mean that these emigrants move seeking to increase their own educational levels. In fact, it is possible to interpret this evidence in two different ways: a) it is possible that these migrants are in fact seeking to increase their education level, in order to elevate their future (or long run) real income earnings; b) it is possible that these origin and destination municipalities' education level might be capturing the effects of other unobserved (and positively correlated to education) explanatory variables which were not included[1] in our estimations, such as culture or infra-structure.

Considering that the region's ongoing urbanization process will probably continue through the next decades, since its urbanization rate has not yet caught up to the average national rate, and as such urbanization seem to be bringing local economic growth and development, then this evidence concerning migration determinants can be interpreted as a sign that the current drivers of immigration flows towards Brazilian Amazon might increase in the years to come.

Finally, to sum up, we conclude that Brazilian Amazon currently faces an increasing trade-off between deforestation and economic growth caused by urbanization and population growth. Apparently, there are no signs that such trade-off will decrease in the short run, as individuals who are moving to the region are being driven by its increasing real income level, which in turn is naturally associated to the economic growth brought by the undergoing urbanization process which the region has been going through recently. Given this scenario, future research is needed to in order to draw specific policies which take this trade-off into account, aiming to avoid the increase of local deforestation, but also without decreasing the pace of its recent economic development.

[1] Due to data unavailability.

2. SPATIAL ECONOMICS OF GROWTH AND DEVELOPMENT IN THE BRAZILIAN AMAZON

2.1. Introduction, Motivation and Objective

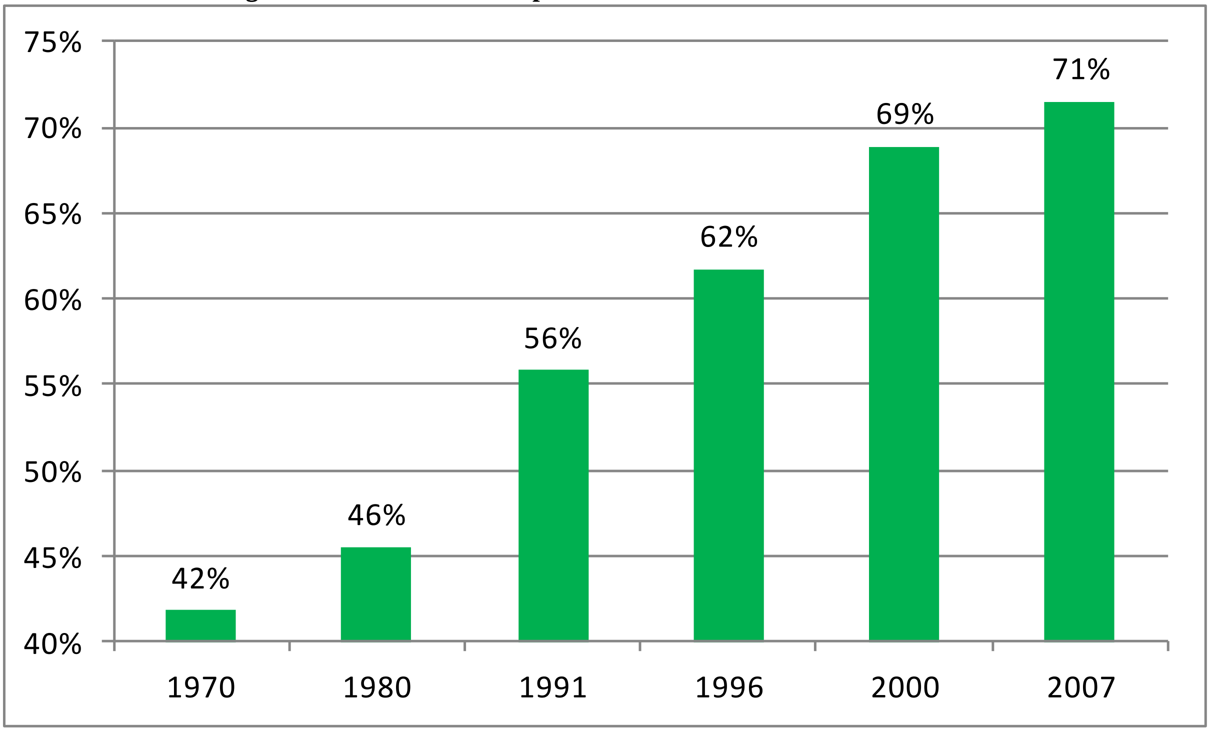
Brazilian Amazon is undoubtedly one of the most important areas of the globe, environmentally. It hosts the largest part of the major remaining rainforest in the world. Its mining resources and biodiversity are almost incommensurable. These are well-known facts concerning the region's current reality. Recently, on the other hand, one less known contextual aspect has been emerging: the region has been undergoing a process of intense population growth and urbanization, which is evidenced by the Brazilian demographic Census databases: from 2000 to 2010, its population grew by more than 4 million inhabitants, representing an increase rate of 21.08%, which equals almost twice the rate for the rest of Brazil (approximately 11%). Such growth has been flourishing over a process of intense urbanization: urban population[2] grew by 3.7 million inhabitants from 2000 to 2010, representing an increase rate of 27.35% throughout the decade, which once again was two times larger than the urban population growth in the rest of Brazil. Figure 2.1 shows the evolution of urban population in the Brazilian Amazon Legal Area over the recently: its urban population share rose from 42% in 1970, to 71% in 2007.

Still, one might argue that such urbanization has not yet reached its peak, since urbanization rates in Brazil as a whole were at the rate of 86.11% in 2010, i.e., 15 p.p. higher than the rate of Brazilian Amazon. In a recent publication, Becker (2013) present an argument defending this idea: according to the author, urbanization within the Amazon region has not yet reached the same consolidation level from the rest of Brazil, neither in terms occupation, nor in terms of promoting economic development within the region.

Table 2.1, and figures 2.2 and 2.3 sum up this situation, and bring more detailed information regarding this referred population growth and urbanization process which Brazilian Amazon has been going through recently, also comparing it to data regarding the Rest of Brazil.

[2] All data referring to Urban population in this paper is calculated following the urban population IBGE criteria, which is sometimes contested by literature (see Veiga 2002). Nevertheless, as we use these data, in general, to compare different time periods, one could say that if there are miscalculations on this data, these miscalculations would tend to incorporate the same bias, which would attenuate literature criticism.

Figure 2. 1: % of Urban Population evolution in Brazilian Amazon



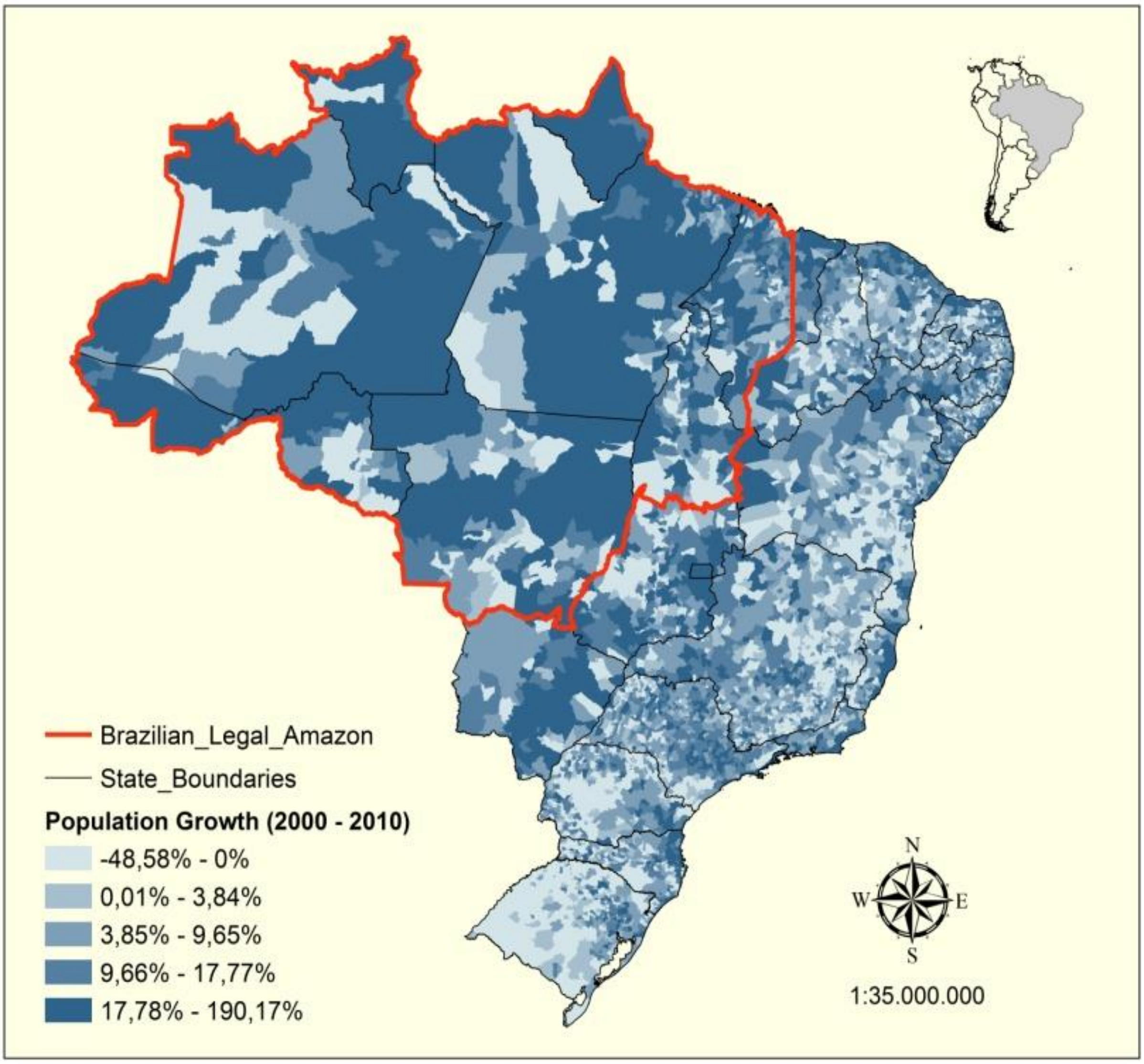
SOURCE: IBGE Census data.

Table 2. 1: Population and Urbanization in Brazilian Amazon and the Rest of Brazil (2000-2010)

	Urban Population				Population	
	Share (2000)	Share (2010)	Abolute Growth (00-10)	% Growth (00-10)	% Growth (00-10)	Absolute Growth (00-10)
Rest of Brazil	82.91%	86.11%	19,179,920	15.46%	11.17%	16,713,195
Brazilian Amazon	68.87%	72.44%	3,791,914	27.35%	21.08%	4,243,434

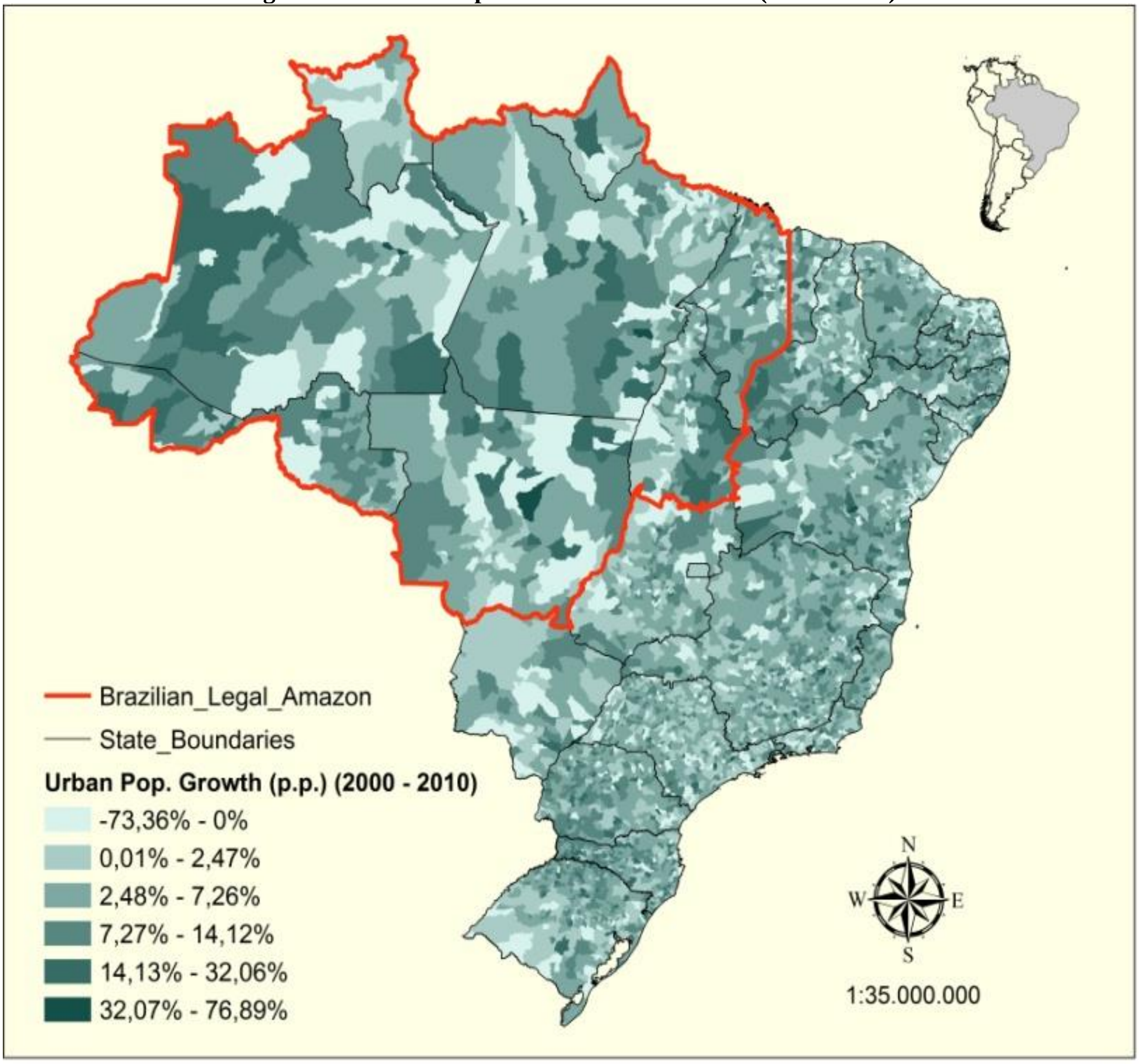
SOURCE: IBGE 2010 Census data, Own Elaboration

Figure 2. 2: Population Growth in Brazil (2000 - 2010)



SOURCE: IBGE 2010 Census data, Own Elaboration

Figure 2. 3: Urban Population Growth in Brazil (2000 - 2010)



SOURCE: IBGE Census data, Own Elaboration

Furthermore, and still according to census data, 28.19% of the total Brazilian Amazon population lived within one of its 5 Metropolitan Regions[3] in 2010. Also, the private Services sector, which is typically associated to urban agglomerations, has increased its participation on the region's GDP composition from 30.79% in 1996 to 35.03% in 2007, according to IBGE's Regional Accounts. Also, this process has been taking place in cities of all sizes within Amazon. Manaus, which is the largest city within the region, and also the main city of a big metropolitan region, has grown from 1.4 million inhabitants in 2000 to 1.8 million in 2008, which represents an increase rate around 28% in only 10 years. Rio Branco, a smaller State capital located in the State of Acre has experienced a population growth from 253,059

[3] The 5 Metropolitan Regions are the following (named usually by the after MR largest city): *Cuiabá-Várzea Grande; Manaus; Macapá; Grande São Luís; Belém*.

inhabitants in 2000 to 336 thousand in 2010, which means a raise of around 32%. As an example of medium-sized cities, the town of São Felix do Xingú, located at southern Pará, grew its population by the outstanding rate of 264%, jumping from 34,628 to 91,340 inhabitants in the last decade, with such growth mostly based on the increase of its urban population, which went from 36% of the city's population in 2000, to almost 50% in 2010.

Bearing all of that in mind, a question arises: are these demographic changes regarding urbanization and population growth correlated to, or even causing, economic growth and development in the Brazilian Amazon?

To try to answer that question is the main goal of this chapter. More specifically, we have two main objectives. Firstly, as already mentioned, we wish to measure how much of this growing urbanization process is contributing to the region's economic growth and development, based on historic background, exploratory analysis and econometric methods structured over neoclassical models. Secondly, we aim to measure how much of the local economic growth and development may be attributed to the region's current internal economic dynamics (in terms of the development of its local markets), comparing it to how much of it still may be attributed to external links and dependence of markets in southern Brazil and foreign countries. Or, put in other terms, our second goal in this chapter is to measure the share of local economic development which may be considered geographically endogenous, as defined by macroeconomic growth theories (see Lucas, 1988; Romer, 1994), and how much still may be considered due to geographically exogenous dependence.

In order to achieve these goals, this chapter is structured in 5 sections, with section 2.1 being this introduction. Section 2.2 presents the literature review of the economic models which will be used as basis to all analysis made throughout the chapter. Section 2.3 presents a historical background of the Brazilian Amazon region, presents the database used in our empirical analysis, and provide an exploratory analysis of this data, bringing the first evidences on the answers to our main questions. Section 2.4 brings a description of the econometric methodology implemented in this chapter, and discusses the results of these regression estimations. Section 2.5 concludes.

2.2. Urbanization and Growth: Theory and practice literature

The relationship between urbanization and economic development and growth is a subject which has been widely studied in literature, since economists and geographers recognized that the world is passing through a shift from agriculture-based economies to urban–based economic systems. These new dynamic systems mostly base their development on the creation of new technologies, mass production and consumption patterns and intensive growth of the services sector. According to the UN (United Nations) Global Health Observatory, more than 50% of the world's population lived in urban areas in 2012, a statistic which was around 40% in 1990. Urban residents grew about 60 million people yearly worldwide, on average, and future predictions estimate that urban population share will be around 70% in 2050, around the globe.

Regarding how urban centers may increase in size and dynamics, it is important to understand that urban systems assume different economic structures and shapes, according to their intrinsic characteristics, history, and external economic linkages. One city may grow due to governmental incentives, while other may prosper only due to geographical initial conditions. That being said, and as expected, there is no previously determined recipe on how to promote urbanization and prosperity. Nevertheless, urban nodes development usually share common features which bring some valuable insights on understanding this dynamics: agglomeration externalities and congestion effects, which we discuss in further detail in the following.

Given this picture, economic models regarding the relationship between urbanization and economic development sprawled within the last 60 years. Given the complexity of the subject, we review how economic literature addresses two main aspects of urbanization: 1) What are the causes of urbanization? 2) How urbanization affects economic development and growth? The first question will serve as the basis of this study, while the second will define our methodology, especially concerning the structural form of the econometric analysis we implement in this chapter.

What causes urban growth?

Urban growth is a complex matter, and theories concerning it flourish in several aspects. In this work, we try to encompass the main economic literature on the subject classifying it into three complementary categories of analysis. The first category may be called, as in Henderson (2010), the Urban Models. These models dispose city growth from a Microeconomic perspective, where each city size and structure will depend on the trade-off between agglomeration externalities and congestion effects. Specifically, the foundations of this category dates back to the pioneer study from Marshall (1920), and defends that firms and individuals tend to agglomerate in urban centers depending on the size of the benefits brought by agglomeration externalities, and also on the costs brought by the congestion effects of conglomerating.

As Ioannides & Rossi-Hansberg (2005) point out, these benefits (positive externalities) encompass the main advantages of conglomerating human activity, and might be classified in 4 categories: a) knowledge spillovers, which is given by the natural sharing of information and knowledge caused by face to face interactions between different agents located in the same spatial area; b) thick markets for inputs, in the sense that specialized laborers tend to agglomerate in urban nodes, so that the respective specialized firms can easily match its needed workforce, therefore, diminishing information frictional costs; c) backward and forward linkages, that is, the advantages which agglomeration bring for both firms and consumers, by respectively reducing transportation costs due to the concentration of the demand vectors in few urban conglomerates; and also due to the greater variety of goods which are available for consumers by the fact that firms are also concentrated at these areas; d) Jacobs externalities: as a consequence of the other three agglomeration forces, urban environment tend to favor the development of "new work" (see Jacobs, 1970), which corresponds to the creation of a workforce with higher productivity (due to the more competitive ambient of labor markets in urban areas brought by the higher spatial concentration of workers), which by their turn result in an endogenous reproductive structure of this labor, due to the economic growth caused by the increase of these laborers' consumption standards. Therefore, Jacobs's externalities refer to the creation of endogenous dynamism propitiated by the urban environment.

According to this literature, the higher the size of these agglomeration externalities, the bigger will be the city size. However, as a city grows, congestion effects also increase. These congestion effects are given by several aspects of urban disposal which poses as costs to the individuals and firms within the urban nodes: elevated traffic which raises transportation costs, air and water pollution, high criminality rates, excesses in labor supply, etc. Therefore, city size is determined by the individuals' and firms' decisions whether to move or to leave urban areas, according to their perception of the trade-off between these costs and benefits. In this sense, for example, new technologies that reduce congestion effects or increase the size of agglomeration externalities may contribute positively for a city to grow.

In this sense, urban models explain the internal incentives for a city to grow or decrease in terms of its population. However, in order to comprehend the whole complexity about urbanization, one must include other "external" elements into the analysis. More specifically, in order to an urbanization process to occur, one out of two possibilities must happen: it is necessary that agents migrate from rural to urban areas; or it is necessary that population and production increase in both rural and urban areas, resulting in urbanization accompanied by rural growth. In fact, even if the first possibility occurs, it is also necessary that output in rural areas increase in order to sustain the usually higher consumption standards from individuals who live within urban areas. In order to include this aspect into the urbanization analysis, a group of models from the so-called New Economic Geography (NEG) emerged (see, for example, Krugman 1991, Fujita et al. 1999, Henderson & Thisse 2004), which we consider as being the second category of models in our analysis.

These models became known as Core-Periphery, as they are based on a simple dichotomist division of a municipality into a "core" urbanized (and industrialized) sector, and a "periphery" agricultural region. The urban sector is considered to be the one that pushes forward economic activity, as it is assumed to be more dynamic by presenting increasing returns to scale, whereas the rural (agricultural) sector pursue a production function based on constant returns to scale, and exists mostly to sustain both urban and rural population. These increasing returns to scale of the urban sector allows it to pay higher relative wages, which in turn tend to attract workers towards these urban areas, depending on the costs of migration relative to the real income differential. Urban growth, thus, may occur depending on this real wages differential, which by its turn is determined by the size of the increasing returns and other parameters within the model. These parameters may vary from one city to another, which by its turn permits that several equilibriums may occur, each with different levels of population concentration between regions. Therefore, these models bring two very important contributions to understand urban growth: they include the needed rural sustainability aspect of urbanization, and also serve as a baseline to explain the co-existence of cities with different degrees of urbanization.

These several cities' sizes possibilities, more than simply corroborating one real aspect of urbanization throughout the world, also bring into perspective another aspect which needs to be taken in consideration: cities are usually a part of a urban system within one region or country, and differences in size, economic development, growth and other variables may be explained not only by internal determinants such as agglomerations and wage differentials, but also by the interconnectivity between these cities as parts of a system. In this sense, a third (and more recent) category of models regarding urbanization growth emerge: the models which study urban systems as a whole. We refer to this group of models as the "urban system" ones (see Duranton & Puga, 2001; Duranton, 2007; Findensen & Suedekum, 2008 and Igliori et al., 2012).

This group of models consider urban systems as dynamic entities, with each urban node working as a small open economy, trading firms and workers with other cities. In this sense, cities usually grow according to their capacity of providing technological innovations which attracts firms and/or workers. In this sense, more innovative cities tend to increase by capturing agents from less innovative municipalities. The logic behind this argument is that firms and industries are able to move quickly from one city to another. City growth, however, is slow, due to the fact that, in general, whereas one group of industries prosper, others decline at the same time. Finally, cities' size distribution tend to be is still, as most of the times populations simply move from one city to another, without interfering in the distribution of these.

As a final regard, we remind that this classification of urbanization models in three categories is made only in order to ease the comprehension of the main differences between the several argument presented by literature. However, it is important to highlight that, in fact, all these models are deeply interconnected, and most of the times rely on the same set of hypothesis and conclusions. As Henderson (2010) argue, a good example of this interconnectivity is the explicit link between the Urban models and the NEG Core-Periphery models. According to the author, both of these groups of models rely on knowledge spillovers as being the basic pillar among the urban characteristics, since they are responsible for the attraction of firms and individuals towards cities in Urban models, whereas in the NEG models, they are support the argument that the urban sector presents increasing returns to scale.

How urbanization affects economic growth and development?

Almost all studies about urbanization tend to agree that urbanization and economic growth and development are positively correlated. The very models which explain urban growth, reviewed in the last subsection, also argue that this positive correlation is usually a natural consequence of urban growth. In Urban Models, an increase in economic growth is considered the result of the agglomeration externalities nature, since these effects turn out to reduce costs of production, increase labor productivity and create innovations. For example: agglomeration externalities are composed by knowledge spillovers, which by their turn raise the creation of technological innovations, thus resulting in a following increase of regional economic growth. Thick input markets, as another example, tend to reduce costs of production, therefore also increasing the economy’s output level. Jacobs externalities, in turn, are naturally associated with creation of knowledge and increases in productivity. And so on.

New Economic Geography models, by their turn, also imply that urbanization comes along with economic development. Population movements towards urban nodes with higher real wages increase are considered to raise the economy’s overall productivity, as the core urban sector which absorbs the workforce presents increasing returns to scale, whereas the rural sector is considered to be less productive. Also, some NEG models, such as in Fujita & Thisse (2003), migrants are considered to be more skilled[4] workers than non-migrants. And as one of the main results of the NEG models concerns the fact that migrants usually move towards the core urban regions (which are also the more productive ones), such migrations flows tend to lead to an increase of labor productivity, which by its turn ends in elevating economic growth.

Furthermore, “Urban system” models such as Duranton (2007) and Duranton and Puga (2001) are based on the hypothesis that urbanization is led by industries’ migration flows, which by their turn follow the pattern of emergence of technological changes throughout space. Thus, urban growth itself is already considered to be a part of technological progress, and therefore, is naturally positively correlated with economic development. Furthermore, in these systems, whereas some industries grow driven by the innovations shocks, others simply grow due to the economic multipliers which are generated by the initial growth of the innovative industries. This, in turn, generates new positive demand shifts for all industries within the municipality, due to the resulting increase in income and employment driven by these initial growth and the economic multipliers generated through the process, in several sectors. All of this dynamic, in turn, results in economic growth and development of the city as a whole. Furthermore, these “Urban system” models still argue that such growth may not be restricted to the municipalities where innovations emerge: according to those models, the initial development of one urban node tend to sprawl throughout the whole urban system, due to the several economic linkages between municipalities, such as the trade of goods and services, tourism, neighbor spillovers, and so on.

[4] A result which also follow extensive literature on the subject. See, for example, the classic article from Mincer (1978) for more details on this specific issue.

Finally, other types of models (which do not fit exactly in any of the three categories considered here so far) also agree with the perspective that urbanization and growth are positively correlated (see, for example, Gleaser, 2008; Henderson & Thisse, 2004). Ioannides and Rossi-Hansberg (2005), for example, argue that many economic activities can only take place within urban environments: personal services, financial activities, real estate markets and many others activities depend on face-to-face interactions to exist, and are only feasible in urban conglomerates. These activities, by their turn, create economic multipliers which naturally drive economic growth and development.

Justifying structural forms: Empirical literature on Economic Growth as the dependent variable

As presented in the previous section, theoretical literature about urbanization, in general, seems to converge towards the existence of a positive correlation between urbanization and economic growth. More than that, it is possible to argue that these models imply an even stronger relationship between those two variables: a relationship of causality, in the sense that urbanization causes growth and development. Apparently, such causality seems to be straightforward in the models presented so far: positive agglomeration externalities, increasing returns to scale, knowledge spillovers are all characteristics attributed to urbanization in these models. And these characteristics, by their turn, are the ones which create economic multipliers, which increases real wages and productivity, which reduces costs of transportation and production, and which promote the creation of markets that need face-to-face interactions between agents to exist.

However, due to the complexity of urbanization processes, since they involve many interconnected variables, it is important that we ground the econometric structural form used in this study on empirical models which take such causality into consideration. Keeping it in mind, we review the empirical literature on the determinants of economic growth and development, focusing on how these models include growth and development as endogenous variables caused by exogenous explanatory variables which encompass urbanization representatives among them. This empirical literature will serve as basis to provide the structural forms which we implement on the regressions that we estimate in our econometric approach.

Specifically, in this literature, empirical estimations of regressions with economic growth being the dependent variable follow a strategy which relies on neoclassical macroeconomic growth models (see Fingleton 2003, Barro and Sala-i-Martin 1995, Armstrong 1995). The basic form of these models refer to Solow (1956) classic macroeconomic growth structure, in which a single economy exhibits a production function involving capital and labor, with constant returns to scale. Output in this economy is of a single sector which produces a homogenous good, which by its turn is either invested or consumed. Investment rate equals savings rate, capital depreciation is constant over time, and both population and workforce growth are determined exogenously at a constant rate. Given those conditions, economy converges to a stable point where investment-depreciation relationship is at balance: if output is above the steady state point, savings are also higher, and due to diminishing productivity of capital (whereas the depreciation rate is constant), such savings will have to decrease in time, for its returns are lower than the depreciation schedule. The opposite similar argument is true if output is below the steady state point. For such, investment-savings rate equals the depreciation rate when convergence is achieved. Also, at this steady state point, when there is no technological progress, consumption, output and capital stock each grow at the same rate of population, which means that per capita values are constant. In other words, in this set up, per capita growth in the long run is only possible if there is also technological progress.

Following Fingleton (2003), we can express this basic model's steady state growth rate of the GDP per capita by a Taylor linear approximation around the steady state point. This is expressed in equation 2.1.

$$p_t = -(1-\pi)(\theta+\phi+\delta)(\ln P_{t-1} - \ln P^*_{t-1}) + \phi \qquad (2.1)$$

In 2.1, p_t represents the per capita output growth at time t, P_{t-1} is the level of output per capita at t-1, P^*_{t-1} is the output per capita in steady state equilibrium, θ is the population (workforce) growth rate, ϕ is the rate of technical progress, π is the capital share parameter in a Cobb-Douglas production function, and δ is the depreciation rate. Integrating 1.1 leads to 1.2:

$$\ln(P_t / P_{t-T})/T = \alpha - (1-\exp(-\beta T))\ln(P_{t-T})/T + u_t \qquad (2.2)$$

Where $\beta = (1-\pi)(\theta+\phi+\delta)$ and represents the rate of convergence of this economy towards the steady state point, u_t is a disturbance term, and α is the constant term to which the economy converges (Fingleton, 2003). As π represents a share, it is lower than 1, and for such, one can easily notice that β is negative, representing that if the economy has an initial level of output per capita higher than the steady state level, then GDP per capita must decrease in order to reach the steady state savings-depreciation balance, and vice-versa.

This basic framework represented by equation 2.2 is the most basic convenient reduced form for empirical modeling, for it represents a simple equation with per capita GDP growth rate as the dependent variable regressed against an initial level of GDP. If we were to apply equation 2.2 as the structural form of our regressions, as the units of observation of our sample are the municipalities of Brazilian Amazon, we would be able to say that equation 2.2, as it stands, would be representing a basic neoclassical growth model with the assumption that α is the same for any city *i* in Brazilian Amazon. This, in turn, would mean that we would be assuming absolute convergence in the long run, because all cities' economies would be tending to the same steady state point. However, this is not a very realistic set up, since Amazon cities pursue different rates of investment, initial capital level, depreciation rates, workforce growth, and several other heterogeneities. In order to capture these heterogeneities of the municipalities in the Brazilian Amazon region, and also in order to avoid any kind of omitted variables bias in estimations, we keep on following literature (see Fingleton 2003, Barro and Sala-i-Martin 1995, Armstrong 1995), and assume that each city will be converging to its own steady state. This means assuming the so called relative convergence assumption in growth literature. In order to do so, this literature suggests to include, in equation 2.2, other explanatory variables which capture the specific characteristics of each city, such as economic indicators, geographical characteristics, social patterns, institutional differences, and so on, all referring to the initial period of the analysis. In our estimations, this means expanding equation 2.2 towards equation 2.3.

$$\ln(P_{t,i} / P_{t-T,i})/T = \alpha - (1-\exp(-\beta T))\ln(P_{t-T,i})/T + \gamma X_i + u_{t,i} \qquad (2.3)$$

where *i* refers to each city of Brazilian Amazon, X_i represent the matrix of exogenous characteristics of city *i*, with γ being the vector of partial correlation coefficients of each of these characteristics. Note that once we control for the vector of characteristics X_i, then α becomes the average steady state equilibrium point of growth which each city would achieve if their steady states were all equal. However, once specific characteristics are included, equation 2.3 will result in different equilibrium points for each city i, which means relative convergence of growth, as desired. Finally, by calling $b = (1-\exp(-\beta T))/T$, we attain:

$$\ln(P_{t,i} / P_{t-T,i}) / T = \alpha - b\ln(P_{t-T,i}) + \gamma X_i + u_{t,i} \qquad (2.4)$$

which gives us a linear structural form which clearly can be estimated through standard econometric regressions methods, such as Ordinary Least Squares (OLS).

Therefore, equation 2.4 is the basic structural form which we chose for our estimations. That being put, we now describe specifically the set of variables (dependents and explanatory) which we use in our estimations.

As dependent variables, we chose three different variables whose growth rates represent economic growth and development. The first one is given directly by the left-hand side of equation 2.4 itself, which is the per capita GDP growth rate, used in basic neoclassical growth models (such as the ones presented here). The other two are the employment growth rate, and the HDI-M (Human Development Index per Municipality[5]) growth rate. Employment growth was chosen because of two main reasons. First of all, employment is a variable which represents labor, which by its turn may be seen as one of the most important elements of growth and development: labor income may be considered the main driver of economic growth, once it is the main source of personal wealth for the majority of the population. Moreover, employment growth is associated with the creation of new jobs, which by its turn represent capital accumulation, according to neoclassic growth models. Also, positive shifts in employment are associated with forward and backward multiplier linkages in the economy, since creation of destruction of jobs openings may be the result of external positive or negative demand shocks, and job creation themselves increase the national income level. The second main reason for the use of employment growth as a dependent variable in our regressions is due to empirical literature (see, for example, Gleaser 1995 and 1991, and Igliori et al 2012, Fingleton, Moore and Igliori, 2003). For instance, equation 2.4 became famous in literature as the “Barro regression”, with growth being explained by several other initial conditions. And in empirical literature, this “growth” is often represented by employment instead of per capita GDP, not only due to the first reason gave in the previous paragraph, but also many times due to data scarcity of per capita GDP per municipality. In such literature, a structural form identical to 2.4 is used for regression estimations, but with employment in place of per capita GDP, which is also one of the possibilities adopted here.

[5] This Index is calculated per municipality of Brazil every ten years, and follow the same methodology of the HDI index calculated per country by the United Nations. The HDI-M Index itself, along with this specific methodology can be faound at PNUD (2013).

As for the choice to also estimate regressions with the HDI-M growth rate as the dependent variable, it was made in order to possibly capture aspects of economic development which might not be directly captured by GDP or employment. The reason for that is because the HDI-M index is composed not only by the population income (or employment) level, but also by its educational and health conditions levels. Therefore, in order to analyze if urbanization and other explanatory variables exert influence not only on growth, but also in other aspects of development, we have also estimated these HDI regressions in this study analysis. Furthermore, in order to be able to compare results from the HDI, employment and GDP regressions, we chose to implement the same structural form to HDI growth regressions as the ones used in employment growth and per capita GDP regressions, given by equation 2.4.

At this point, it is important to stress that we are aware that using an index (such as HDI) as an dependent variable may not be considered ideal by literature, mostly because indexes are by construction influenced by many variables included in the explanatory vector (which means that the correlations captured by estimations might be mathematical, and not statistical). However, we highlight that the great majority of the explanatory variables which we have used refer to the initial period of the explained growth rates. This, in turn, means that this kind of endogeneity through reverse causality is mostly being avoided, since it is difficult to argue that variables whose values were defined in previous periods may be explained by values of another variable defined in future periods. On the other hand, empirical literature which uses the HDI or other indexes growth rates as dependent variables is not abundant, which leads us to interpret these group of regression results mostly as complementary, with our main results referring to the per capita GDP and employment growth regressions, even though most of the results in our regressions proved to have similar interpretations for all dependent variables.

As for the explanatory variables, the choice of variables was made according to and the main goals of our analysis, and also based on literature on economic growth. In this sense, besides including the initial level of the dependent variable in the year 2000, we firstly have selected three variables to represent the urbanization rate of each municipality, in order to achieve the main objective of investigating if urbanization has been causing economic growth in Brazilian Amazon. The first one is the municipality’s share of urban labor force (or the share of urban EAP – Economic Active Population), which measures who many workers in each city live within the urban areas[6], among all workforce population. The higher this share, the greater the urbanization level. This first measure of urbanization captures the geographic aspect of it, as it is based on a geographical concept of urban areas. The second measure is the share of employment in the terciary sector within the city, and refers mostly to an economic perspective of urbanization, as the terciary sector basically consists of urban economic activities such as personal and public services, culture industry, etc. This aspect may be considered complementary to the first one, since it encompasses a different aspect of urbanization, and it does not rely on any mapping subjective concept. The third measure of urbanization included, by its turn, tries to capture the influence of the most densely and big urban complexes of Amazon, as it is a dummy variable representing whether or not a city belong to a metropolitan region. This effect is also complementary to the other two, since metropolitan regions not only are highly urbanized by definition, but they also exert influence on their neighbors' urbanization rate, since spatial spillovers from these areas tend to be elevated.

In order to be able to compare the relative importance of internal drivers and external drivers over local economic growth, we have also included explanatory variables which represent the external links of the Amazon municipalities' economies with the ones from the Rest of Brazil and the exterior. The main idea behind this inclusion is that, if these variables' coefficients prove to be significant, then economic linkages with external economies are to be considered relevant as drivers of growth and development. If coefficients from internal characteristics are also significant, then we may consider that both internal and external economic variables are relevant in terms of promoting growth and development within Brazilian Amazon. In this sense, the specific variables we have included are the amount of output exports from Amazon cities to other countries, and the average transportation costs from each city to São Paulo[7]. In sum, it is expected that the greater the economic dependency of Amazon on other countries and regions' economies, the bigger will be the coefficients of the exports to other countries, and the more negative will be the coefficient of transportation costs to São Paulo.

[6] According to the IBGE criteria of urban areas. This criteria is criticized by Veiga (2002), who argues that it may overestimate the size of brazilian cities urban areas. However, even if this critique is correct, as we use the same criterion for all municipalities, we are most likely committing the same average error to all cities, which means that the variance of this urban explanatory variable may not be altered relevantly by this error. And as the explanatory variables' variance is what matters for the econometric methods to be consistent, then we argue that this measurement error may not affect our results relevantly.

[7] We chose São Paulo as the representative destination for the transportation costs, because it is by far the biggest Brazilian city, thus it represents the largest market to which Amazon cities may sell goods.

Furthermore, we also have included other control explanatory variables in our regressions according to their availability, and also to empirical literature recommendations on the subject. In this sense, we have included a set of variables representing the educational level of the municipality, measures of income distribution, the deforestation level of the city and its neighbors, public social policies, labor market characteristics, sectorial GDP shares and spatial neighbor relationships. Education variables are justified because literature indicates that higher education levels may lead to more technical progress and higher productivity (see Duranton, 2007; Fingleton, 2003; Becker, 2013). The city's deforestation level, by its turn, is expected to partially explain growth because deforestation is naturally associated to land-use change and production, thus, cities with higher levels of deforestation are expected to present higher levels of economic development too.

The labor market main characteristic which we have included regards its level of informality, as it is an institutional aspect which measures the degree of maturity of one city's economy. Thus, it is expected that more informality in labor markets should lead to lower development. As for the public social policies, in fact we have only included one variable, which is the social assistance program "bolsa família", which is a social policy of income transfer from Brazilian Federal government to the poorest share of the population. The main goal of this program is to increase equality in Brazil as a whole, through the direct reduction of poverty. Therefore, it may also be directly connected to growth, as it tends to increase consumption from the poorest share of the population, even though it may also bring a downshift in labor supply, as in any social secure policy.

Even though growth and equality should not present any kind of expected correlation, from a neoclassic microeconomic perspective[8], we chose to also include some indicators of inequality within the municipalities as control variables, as they might be associated with other omitted variables which also affect growth. These variables are the Theil index of inequality and the share of the HDI index which refers to income distribution. Also, still following growth literature (see Fingleton, 2003), we have included the shares of each great sector in the GDP composition in the base year of 2000 as explanatory variables. The reason is: by including all three great sectors (agriculture, manufacturing and services) as explanatory variables of growth, we are possibly capturing tendencies of development related to at least one of these sectors. In this sense, for example, if a specific new technology in one of the

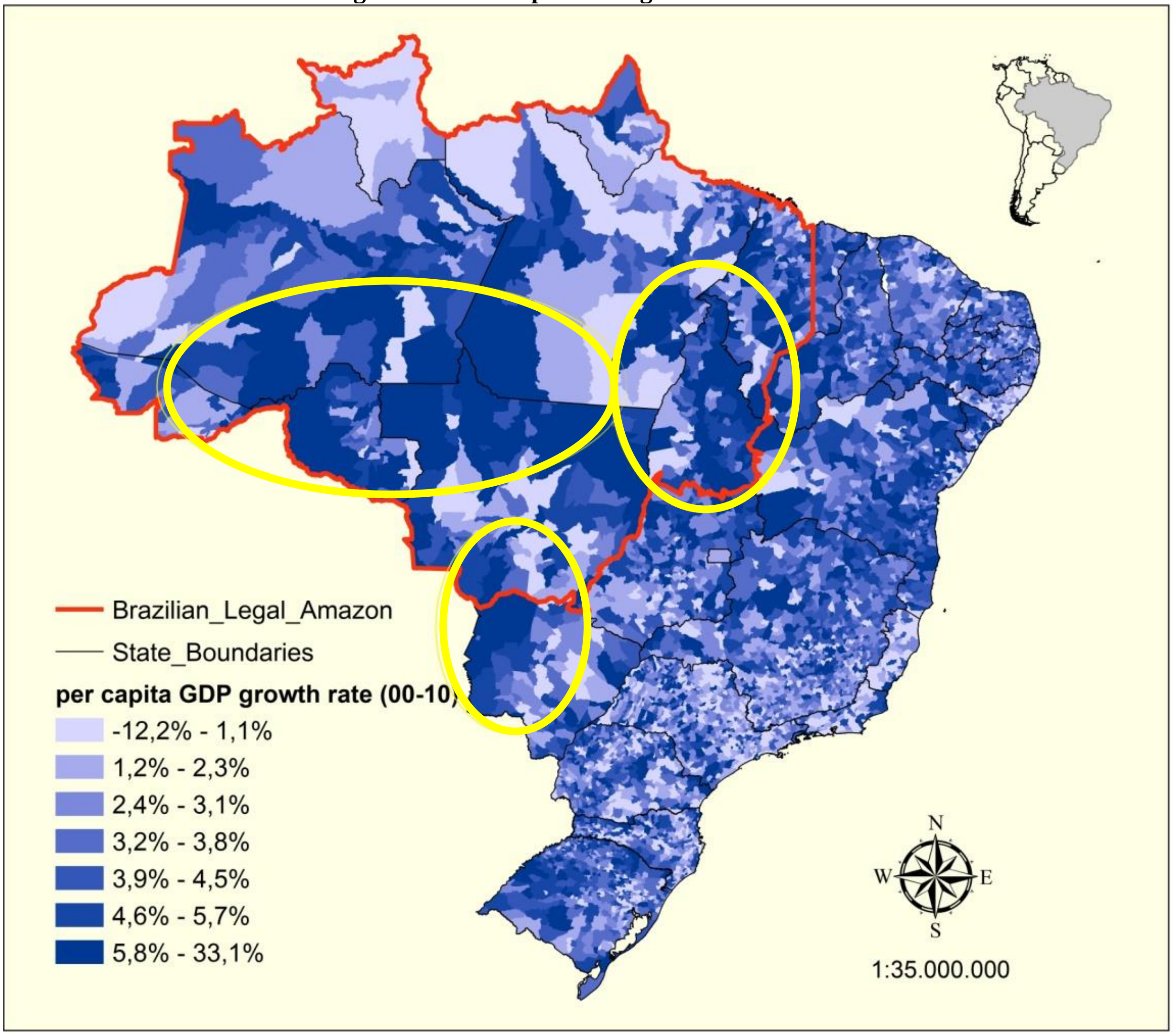
[8] The main example of this affirmation would be the Pareto (1964) efficiency concept, in which consumption or profit maximization may be achieved independently of the income or endowments distribution.

three sectors emerges, and this boosts growth in all cities in which this sector has the major share, then this variables will capture this specific effect. As we will see later on, this do not seem the case in our estimations.

Finally, one more time guided by literature, we chose to include spatial neighbor effects among the explanatory vector. The reason for that is because in almost all theoretical models which guide this study, growth and development are spatial phenomena. Empirical literature on the subject shows, most of the times, that growth occur in a clustering dynamics. Among several studies, Fingleton (2003) shows explicitly that the dynamic of per capita GDP growth in Europe, recently, follow a pattern in which one can easily identify spatial agglomerations of the economies with the highest growth rates. Fingleton, Igliori and Moore (2005) find a similar evidence for English cities. According to this point of view, it is impossible to isolate the economic growth of one region (city, country) from its neighbors. And these neighborhood effects are present for several reasons. Cities and countries are small open economies which trade goods and services among themselves, therefore, growth in one region tend to increase its importations of gods and services produced by its neighbors. Also, technological progress in one city creates multiplier effects which sprawls in a geographical scale, inciting migration and capital flows between these economies. In dense metropolitan areas, it is common that some peripheral cities serve as workers "dormitories", while central municipalities cluster the jobs of these individuals, which reflects the neighbors' influence on growth as a necessary condition in this case.

Empirically, in our case, this clustering pattern is not difficult to observe. Similarly to Fingleton (2003) and Fingleton, Igliori and Moore (2005), Figure 2.4 brings the per capita GDP growth rate (yearly) between 2000 and 2010, per municipality, in Brazil, focusing on Brazilian Amazon. It is easy to identify positive clustering patterns in the figure. Yellow ellipses indicate the most visible positive ones within Brazilian Amazon.

Figure 2. 4: Per capita GDP growth in Brazil.



SOURCE: Own Elaboration, IBGE Census data.

Therefore, econometrically it seems correct to include this spatial neighbor's pattern among the explanatory vector of variables, in order to avoid relevant aspects of growth which, if omitted, might cause bias in estimations. Thus, we chose to include spatial lags of both dependent variable and the error term, as in most of the traditional spatial econometric approaches (see Anselin, 1988; LeSage 2008). The specific estimators and econometric strategies are discussed in the next sections of this chapter.

The final consideration we make regarding the set of explanatory variables included in our models refer to one deficiency which we could not account for. As most of our explanatory variables refer to the base year of the growth rates studied (2000), it is possible that our model might not be accounting for a positive or a negative macroeconomic conjecture which Brazilian municipalities might have passed through in this year. For example, it is possible that the year 2000 was an exceptionally positive year in terms of the employment growth in

Brazil as a whole, which by its turn might have caused lower subsequent growth rates. Even though we cannot include these effects due to data restriction (since we work with Census data, which is decennial), we argue that this might not be considered a major issue in our estimations, as our main strategy relies on comparing between Amazonian municipalities. This means that, at least, macroeconomic shocks which affected similarly all Amazon municipalities might not cause bias in our estimations, given our goals and methods. Moreover, many of our explanatory variables included are not defined by short term conjectures, including the main interest urbanization ones, which represent part of a long run process. In any case, we are aware that our results may be subject to cyclical economic variations in a certain degree.

2.3. Historic Background, database and exploratory analysis of the Amazon urbanization process

2.3.1. Amazon Geography and Historic Background

The Brazilian Amazon region is immensely important in many ways. Geographically, it covers an area of approximately 61% of Brazil, with 5,217,423 km ² of extension, divided in three different kinds of vegetation: rainforest, savannah ("*cerrado*") and wetlands ("*pantanal*"). Economically, it is the region where the Brazilian agricultural frontier is located at, nowadays, also being where the Brazilian soybean belt is located at (especially at the Mato Grosso State). It also holds some of the largest mineral deposits in the world, particularly in the State of Pará. Environmentally, it holds the largest tropical rainforest in the world, whose burnings and deforestation are responsible for about 58% of emissions of greenhouse gases in Brazil, and about 2% of global emissions (Imori et al., 2011). This deforestation process, by its turn, is one of the main global environmental concerns in the world nowadays, not only due to such greenhouse gases emissions, but also because of the potential irreversible biodiversity loss associated with it.

This deforestation process, as explained in the previous section, is often attributed, by literature, to the supply-side of local economy, which points to soybean production and cattle ranching as being its' main causes. However, as already mentioned, literature seem taking a wrong direction by overlooking another historically recognized driver of environmental impacts and deforestation, which is the local urbanization process and population growth taking place in the region. As an example, it is a well-known fact that China is been going through a huge urbanization process within its borders, and such urbanization is often held responsible for enormous macroeconomic and environmental impacts all over the world, especially in terms of global emissions. That being said, it seems a little naïve, environmentally speaking, to ignore a local urbanization process taking place at the same region where the major lasting rainforest of the world remains.

That being put, it is easy to notice that the initial stages of such urbanization process are largely due to occupation policies of the Brazilian federal government during the decades of 1960, 1970 and 1980. During those years, many regional development programs were created aiming to integrate the Amazon region to the national economy, for this part of the country was considered to be very isolated and underdeveloped at that time. Since then, government invested heavily on roads and infra-structure, promoted new lines of credit to subsidize production, created the "Manaus Free Zone", an area of free commerce with tax exemptions, among other policies. After that period, during the decade of 1990, such occupation policies have lost strength due to the international economic crisis at the time, which affected heavily all Latin-American governments, and also due to new emerging eco-friendly visions spreading all over the globe, which defended that Amazon deforestation process was extremely correlated to those policies.

However, as the brief census data analysis from previous sections showed, urbanization and population growth within the region did not decreased after government incentives ceased: on the contrary, local population and urban centers continued to grow, sometimes at a higher pace than before. And the economic consequences of it are exactly what this thesis aims to investigate.

2.3.2. Database

The database used in our estimations was almost entirely obtained from the IBGE[9] Census and from IPEA[10], with only one exception: deforestation data was obtained directly from INPE[11]. Table 2.2 brings the basic description for all variables used in our regressions, as well as the units for all variables.

---

[9] "*Instituto Brasileiro de Geografia e Estatística*". Available at www.ibge.gov.br
[10] "*Institutos de Pesquisa Econômica Aplicada*" Available at www.ipeadata.gov.br
[11] "*Instituto Nacional de Pesquisas Espaciais".* Available at www.inpe.br

Table 2. 2: Description of the variables

	Brazilian Amazon	Mean	Std. Deviation	Min	Max
Employment growth (% yearly, from 2000-2010)	3.3%	2.7%	3.1%	-5.7%	16.8%
GDP per capita growth rate (yearly) (% yearly, from 2000-2007)	3.6%	4.8%	4.2%	-8.8%	32.0%
HDI-M Growth (% yearly, from 2000-2010)	2.5%	3.2%	1.0%	1.1%	7.7%
HDI-M in 2000 Index	0.51	0.45	0.08	0.22	0.69
HDI-M in 2010 Index	0.66	0.61	0.06	0.42	0.79
GDP per capita A thousand Constant 2000 R$ per inhabitant, in 2000	3.8	2.9	3.2	0.6	39.2
Share of Urban EAP (% of EAP, in 2000)	71.8%	54.1%	21.7%	1.5%	99.8%
Metropolitan Region dummy (1 if MR, 0 otherwise)	3.3%	NA	NA	NA	NA
Share of Employment in the Services Sector (%, in 2000)	71.6%	53.9%	18.1%	13.9%	99.1%
Theil Index (Index, in 2000)	0.62	0.57	0.11	0.29	1.13
Income HDI (Index, in 2000)	0.61	0.56	0.09	0.33	0.82
Longevity HDI (Index, in 2000)	0.69	0.66	0.07	0.49	0.82
"Bolsa Família" (04-07) per capita (Constant 2000 R$, from 2004-2007)	11.9	14.6	6.3	1.4	34.5
Exports per capita (Millions of FOB US$ (2003-2010))	5.74	4.08	24.10	0.00	322.21
Informality (% of Employment, in 2000)	76.1%	90.0%	14.5%	0.0%	100.0%
Transport Costs Index to São Paulo (Index, in 1995)	3234	3234	1419	1188	10512
Transport Costs Index to the nearest Capital (Index, in 1995)	941	941	801	0	5949
Deforested Area (% of Area, in 2001)	14.8%	43.8%	38.1%	0.0%	100.0%
Share of Agriculture GDP (% of GDP, in 2000)	31.5%	12.7%	15.8%	0.1%	77.0%
Share of Services GDP (% of GDP, in 2000)	55.9%	56.2%	13.7%	15.6%	84.8%
Share of Manufacture GDP (% of GDP, in 2000)	8.9%	20.9%	8.2%	1.2%	79.9%
Share of Superior education (8-12 years of study) (% of population, in 2000)	29.9%	16.4%	8.5%	2.6%	68.4%
Share of High Education (4-8 years of study) (% of population, in 2000)	26.5%	25.9%	7.0%	7.0%	42.9%
Share of illiterates (% of population, in 2000)	18.0%	25.3%	10.4%	5.0%	60.7%

SOURCE : IBGE, IPEA and INPE, own elaboration

As Table 2.2 points out, urbanization rates measured by all three variables at the year 2000 in Brazilian Amazon may be considered elevated, even though they are still lower than the rates for Brazil as a whole. Moreover, in terms of the local labor market, Amazon economy is still highly informal (76% of the local workers are informal), even though the variance of this indicator is also high in our sample. Moreover, inequality measures point to a structure of relatively high inequality in 2000, something that in fact, was a reality for the whole country at that period. Finally, the Brazilian Amazon educational level can be considered low in global terms, both according to the high shares of illiteracy and the low share of population with superior education. To what concerns our estimations, and also according to literature on the subject, this means that technical progress due to innovations should probably be discarded as the possible main drivers of growth within the region. However, as we will see in the next sections, in spite of such relatively low educational level within the region, education is still responsible for at least part of the development and growth which Brazilian Amazon has been going through in the last decades, corroborating the main theoretical models discussed previously.

A few important remarks about some variables must be made before we move for the empirical methodology and its results. Firstly, it is important to highlight that as the employment level and growth were computed using Census data, they take into account both formal and informal jobs, which is a very desirable characteristic, since the Amazon region exhibit a very high informality rate in the labor market. Secondly, exports data refer only to commerce with other countries, and not between municipalities within Brazil. Thirdly, our database contains data from 1995 to 2010. Between those years, 15 new municipalities[12] emerged in the region, originated from dismembered neighborhoods of previously larger cities. To deal with that without losing neither information nor observations, we treated these new municipalities as still being a part of the original cities from which they were dismembered of, aggregating and averaging (weighting by the population) data from the emerged municipalities to data from the original cities.

Finally, the Transport Costs Indexes used in the regressions were computed for the year of 1995 by De Castro (1999), with no available updates for the year 2000. Even though this earlier initial year of 1995 is not ideal, these indexes were still computed for a time period which occurred previously than the period of growth studied, and for such, still fulfills the theoretical framework hypothesis of using initial conditions of the exogenous explicative variables in our regressions. Of course, these transportation costs might have changed from 1995 to 2000, but we will assume that those changes were completely random among the municipalities, and for such, no explicative power could be gained from those five missed

[12] These municipalities are: *Vale de São Domingos; Serra Nova Dourada; Santo Antônio do Leste; Santa Rita do Trivelato; Santa Cruz do Xingu; Rondolândia; Novo Santo Antônio; Nova Santa Helena; Nova Nazaré; Ipiranga do Norte; Itanhangá; Curvelândia; Conquista D'Oeste; Colniza*

years of observations, even though we are aware that this may be considered a strong assumption. Still, it seems better to use this available transportation costs than to simply omit such an important explicative variable from estimations.

Moreover, we have also tried to replace these transportation costs to São Paulo and to the nearest capital, respectively, by the distance to São Paulo and the distance to the nearest capital. And by doing so, the results obtained were practically the same in terms of significance and signal of the exogenous variables' coefficients, as the ones that we will show in the following sections. Hereupon, we have opted to maintain such transportation costs indexes because of two main reasons: 1) Distance may be considered as only a proxy to the real transportation costs, as it is possible that infrastructure investments between two cities result in lower transportation costs, without changing the distance between them; (2) As we will discuss later on, we will make use of distance as instruments for Transportation Costs in two of our estimators, since such distances exhibit the desirable characteristics of a good instrument which attends to our purposes, as we will argue in the next subsections.

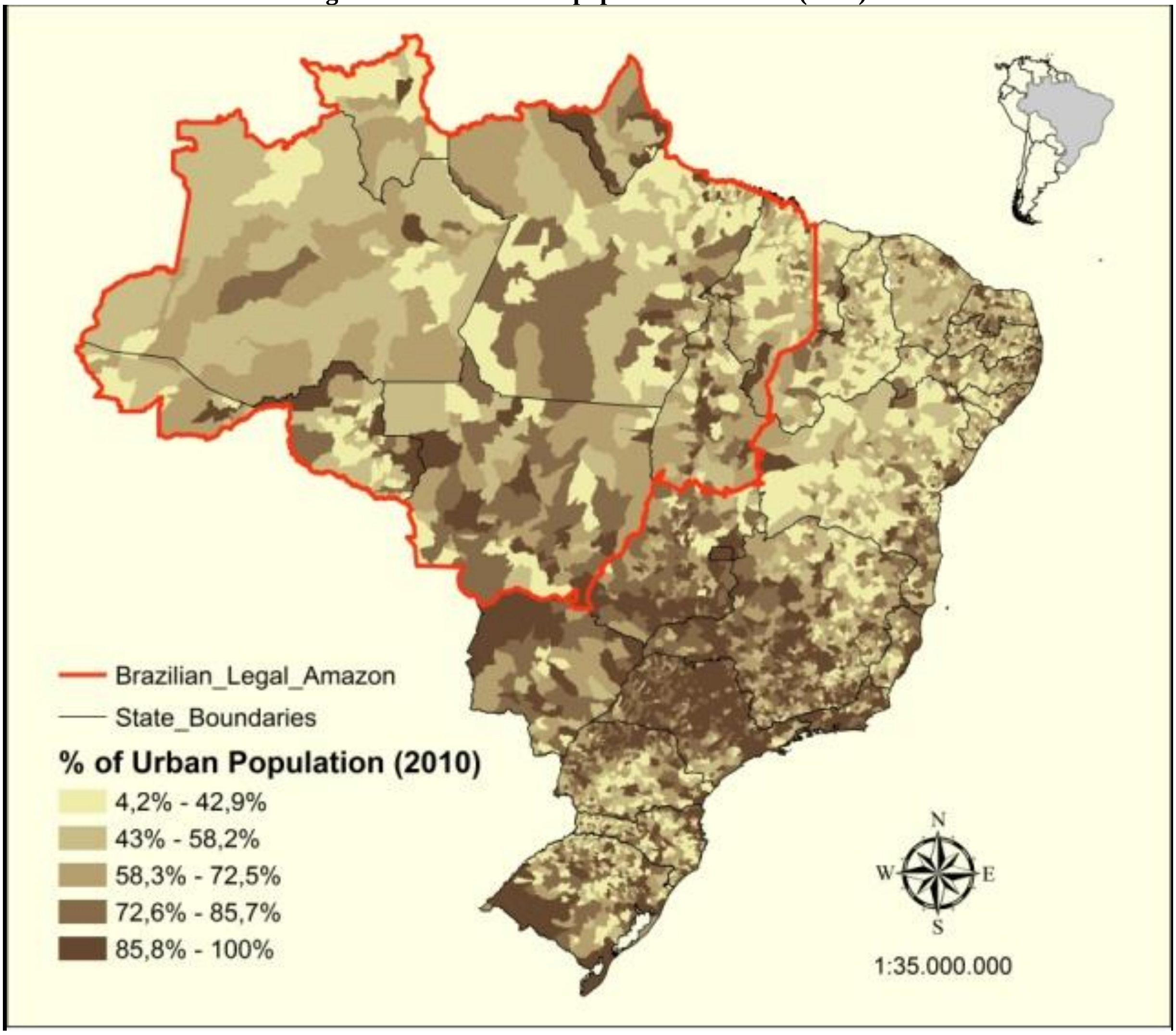
2.3.3. Exploratory Analysis

As already mentioned in the introduction of this paper, a first glimpse at the IBGE census database reveals that urbanization seem to be evident at the Brazilian Amazon: more than 70% of the population lived inside the cities' urban areas in 2010, with approximately 37% of this population living in dense metropolitan areas. Besides, urban population also have increased 30% in the last 20 years. That being put, we analyze a few descriptive statistics which may shed light on some interesting aspects of the relationship between development and urbanization.

In this sense, Figures 2.5 to 2.8, and also tables 2.3 to 2.5 bring some indicators of urbanization and economic development within Brazilian Amazon. More precisely, in Figures 2.5 and 2.6 we show maps of urbanization in Brazil measured by the share of urban population in each municipality, and the share of employment in the terciary sector, respectively. In figures 2.7 and 2.8, in turn, we show two variables representing economic development, at the municipality level: the HDI (Human Development Index) and the per capita GDP. All these variables are considered for the year of 2010, and data comes from IBGE's Census. By comparing these figures, we are able to get a good glimpse at the apparently positive relationship between urbanization and growth: it is easy to note that, at least visually, those cities with higher level of urbanization (measured by both variables in Figures 2.5 and 2.6) are also the ones that exhibit the greater HDI-M and per capita GDP.

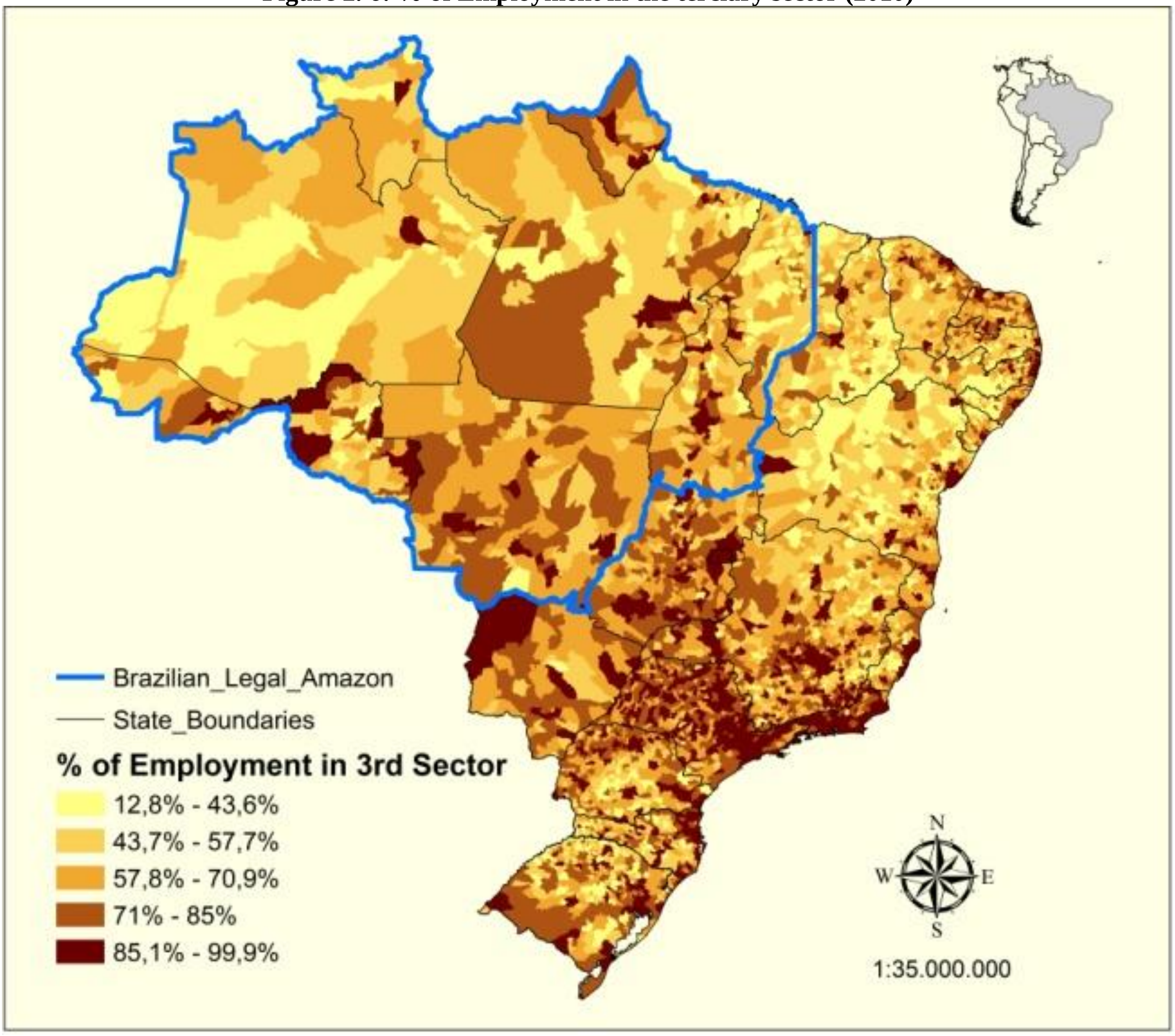
A closer look at descriptive analysis of these variables confirms this visual evidence. Table 2.3 ranks the top 20 most populated cities in Brazilian Amazon, as well as it brings information about urbanization and development rates for each of these municipalities. The first result that calls attention in this table is a very well-known fact for many, but a still overlooked aspect worldwide: three different cities of the region (Belém, São Luís and Manaus) exhibit a population size above 1 million inhabitants, with Manaus leading the rank with over 1.5 million people. By no coincidence, these three cities are their States' capitals, which evidences the relatively high importance of regional political centers. Moreover, this very first evidence, by itself, might be considered enough to raise a doubt on any supposed general beliefs that the Brazilian Amazon economy might be exclusively dependent on external demands from markets located in southern Brazil: it seems hard to trust that internal demand vectors from more than 1 million inhabitants in one city shall not create relevant multiplier effects an agglomeration externalities which could relatively endogenize part of the region's economic growth.

Figure 2. 5: % of Urban population in Brazil (2010)



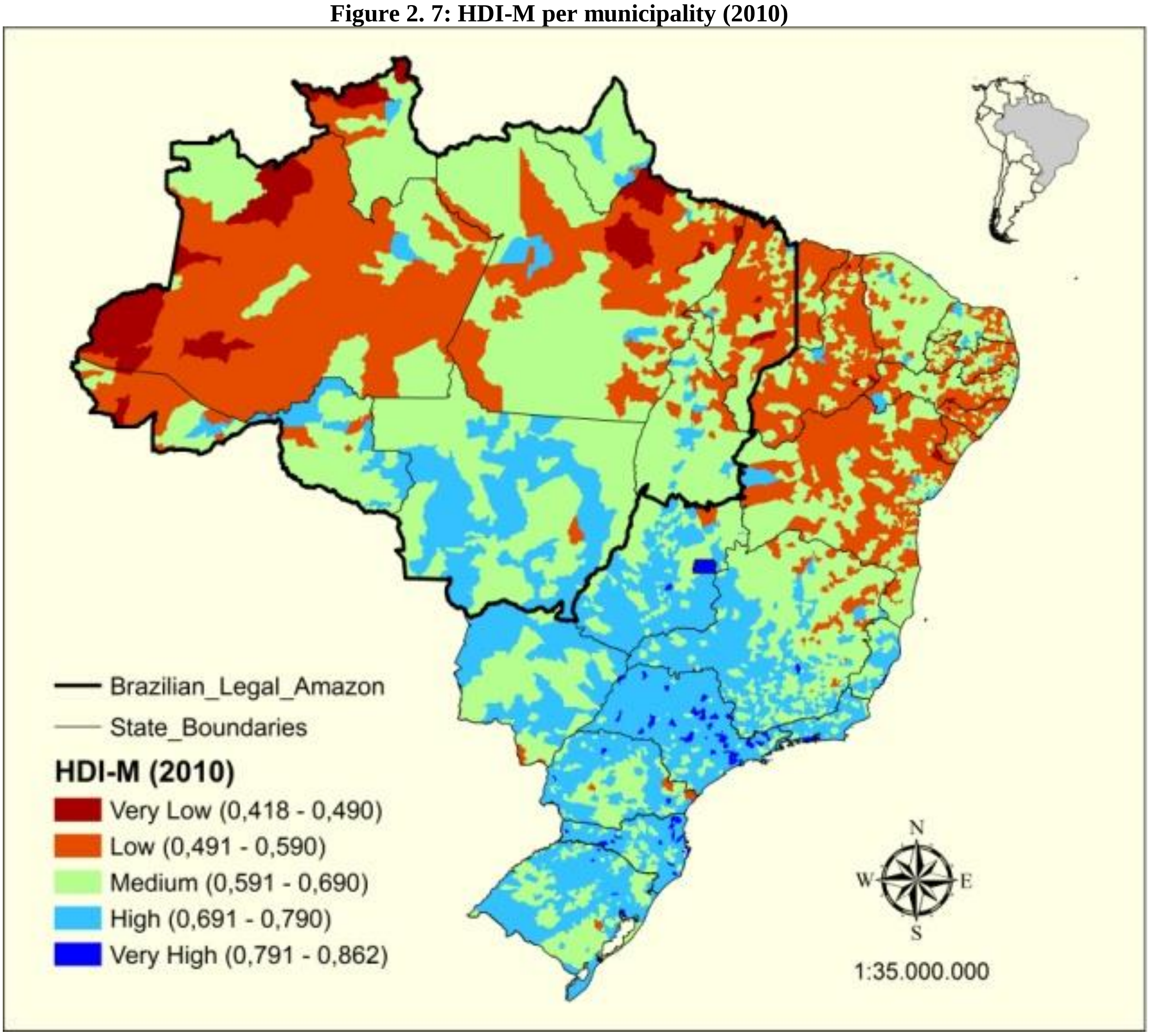
SOURCE: IBGE, own elaboration.

Figure 2. 6: % of Employment in the terciary sector (2010)



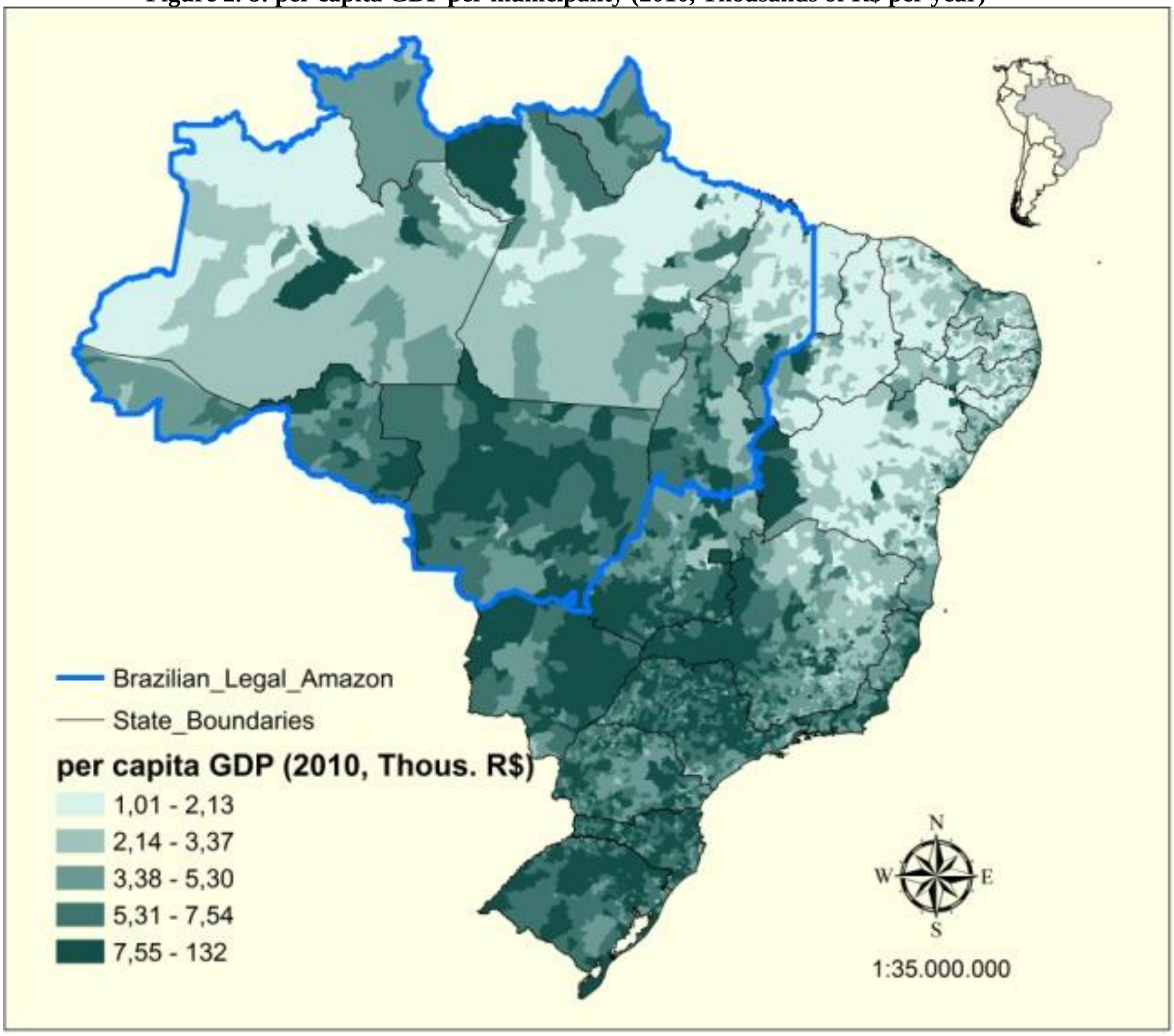
SOURCE: IBGE, own elaboration.

Figure 2. 7: HDI-M per municipality (2010)



SOURCE: IBGE, own elaboration.

Figure 2. 8: per capita GDP per municipality (2010, Thousands of R$ per year)



SOURCE: IBGE, own elaboration.

Table 2. 3: Top 20 Population Ranking in Brazilian Amaozn (2010)

State	City	Population (2010)	Metropolitan Region*	Employment Growth (2000-2010, yearly, %)	% of Urban Population (2010)	HDI-M (2010)
Brazilian Amazon		**24,380,146**	**-**	**3.3%**	**72.6%**	**0.660**
Amazonas	Manaus	1,802,014	1	4.7%	99.5%	0.737
Pará	Belém	1,393,399	1	2.6%	99.1%	0.746
Maranhão	São Luís	1,014,837	1	3.9%	94.5%	0.768
Mato Grosso	Cuiabá	551,098	1	3.5%	98.1%	0.785
Pará	Ananindeua	471,980	1	3.7%	99.8%	0.718
Rondônia	Porto Velho	428,527	0	4.8%	91.2%	0.736
Amapá	Macapá	398,204	1	5.7%	95.7%	0.733
Acre	Rio Branco	336,038	0	4.2%	91.8%	0.727
Pará	Santarém	294,580	0	2.6%	73.3%	0.691
Roraima	Boa Vista	284,313	0	4.5%	97.7%	0.752
Mato Grosso	Várzea Grande	252,596	1	3.6%	98.5%	0.734
Maranhão	Imperatriz	247,505	0	2.5%	94.8%	0.731
Pará	Marabá	233,669	0	4.7%	79.7%	0.668
Tocantins	Palmas	228,332	0	6.9%	97.1%	0.788
Mato Grosso	Rondonópolis	195,476	0	4.7%	96.2%	0.755
Pará	Castanhal	173,149	0	4.0%	88.6%	0.673
Maranhão	São José de Ribamar	163,045	1	7.1%	23.1%	0.708
Pará	Parauapebas	153,908	0	8.9%	90.1%	0.715
Tocantins	Araguaína	150,484	0	5.0%	95.0%	0.752
Pará	Abaetetuba	141,100	0	2.9%	58.8%	0.628

SOURCE: Own Elaboration, IBGE Census data;

*Note: *1 if the city belongs to a Metropolitan region; 0 otherwise. Legend:* Above Brazilian AMZ

Table 2. 4: Top 20 HDI Ranking in Brazilian Amazon (2010)

State	City	Population (2010)	Metropolitan Region*	Employment Growth (2000-2010, yearly, %)	% of Urban Population	HDI-M (2010)
Brazilian Amazon		**24,380,146**	**-**	**3.3%**	**72.6%**	**0.660**
Tocantins	Palmas	228,332	0	6.9%	97.1%	0.788
Mato Grosso	Cuiabá	551,098	1	3.5%	98.1%	0.785
Maranhão	São Luís	1,014,837	1	3.9%	94.5%	0.768
Mato Grosso	Lucas do Rio Verde	45,556	0	9.8%	93.2%	0.768
Tocantins	Paraíso do Tocantin	44,417	0	3.9%	95.6%	0.764
Tocantins	Gurupi	76,755	0	3.4%	97.7%	0.759
Mato Grosso	Nova Mutum	34,140	0	10.7%	79.8%	0.758
Mato Grosso	Rondonópolis	195,476	0	4.7%	96.2%	0.755
Mato Grosso	Sinop	113,099	0	5.9%	82.9%	0.754
Roraima	Boa Vista	284,313	0	4.5%	97.7%	0.752
Mato Grosso	Primavera do Leste	52,066	0	3.9%	94.6%	0.752
Tocantins	Araguaína	150,484	0	5.0%	95.0%	0.752
Mato Grosso	Campo Verde	31,589	0	8.1%	80.6%	0.750
Mato Grosso	Barra do Garças	56,560	0	1.3%	90.1%	0.748
Pará	Belém	1,393,399	1	2.6%	99.1%	0.746
Mato Grosso	Sorriso	66,521	0	7.5%	87.7%	0.744
Mato Grosso	Campos de Júlio	5,154	0	7.2%	77.9%	0.744
Tocantins	Guaraí	23,200	0	3.4%	91.1%	0.741
Tocantins	Porto Nacional	49,146	0	3.4%	86.3%	0.740
Amazonas	Manaus	1,802,014	1	4.7%	99.5%	0.737

SOURCE: Own Elaboration, IBGE Census data

*Note: *1 if the city belongs to a Metropolitan region; 0 otherwise. Legend:* Above Brazilian AMZ

Table 2. 5: Top 20 HDI Ranking (2010) for medium cities in Brazilian Amazon

State	City	Population (2010)	Metropolitan Region*	Employment Growth (2000-2010, yearly, %)	% of Urban Population	HDI-M (2010)
Brazilian Amazon		**24,380,146**	**-**	**3.3%**	**72.6%**	**0.660**
Tocantins	Palmas	228,332	0	6.9%	97.1%	0.788
Tocantins	Gurupi	76,755	0	3.4%	97.7%	0.759
Mato Grosso	Rondonópolis	195,476	0	4.7%	96.2%	0.755
Mato Grosso	Sinop	113,099	0	5.9%	82.9%	0.754
Roraima	Boa Vista	284,313	0	4.5%	97.7%	0.752
Tocantins	Araguaína	150,484	0	5.0%	95.0%	0.752
Mato Grosso	Barra do Garças	56,560	0	1.3%	90.1%	0.748
Mato Grosso	Várzea Grande	252,596	1	3.6%	98.5%	0.734
Amapá	Macapá	398,204	1	5.7%	95.7%	0.733
Rondônia	Vilhena	76,202	0	4.9%	94.8%	0.731
Maranhão	Imperatriz	247,505	0	2.5%	94.8%	0.731
Mato Grosso	Tangará da Serra	83,431	0	4.6%	91.0%	0.729
Acre	Rio Branco	336,038	0	4.2%	91.8%	0.727
Maranhão	Paço do Lumiar	105,121	1	4.9%	75.0%	0.724
Rondônia	Cacoal	78,574	0	1.9%	78.8%	0.718
Pará	Parauapebas	153,908	0	8.9%	90.1%	0.715
Rondônia	Ji-Paraná	116,610	0	2.5%	89.9%	0.714
Mato Grosso	Cáceres	123,538	0	1.9%	84.0%	0.708
Maranhão	São José de Ribamar	163,045	1	7.1%	23.1%	0.708
Rondônia	Ariquemes	90,353	0	3.5%	84.7%	0.702

SOURCE: Own Elaboration, IBGE Census data

*Note: *1 if the city belongs to a Metropolitan region; 0 otherwise. Legend:* ▢ Above Brazilian AMZ

Secondly, it is easy to notice in table 2.3 that the most populated cities within Brazilian Amazon are spread all over the region's different States, since the top 8 most populated municipalities belong to 8 different Federal Units. This shows that the region cannot be considered monocentric in any spatial analysis. To what concerns our goals, it seems reasonable to believe that commerce between those centers might also exert relevant impulses to local growth and development, which once again favors the argument that internal demand drivers and multipliers may be significantly relevant to boost local economic growth

Still regarding table 2.3, a closer look at the results evidences even more the urbanization process by which the region is passing through recently: eight among the 20 biggest Brazilian Amazon cities belong to one of the five official[13] Metropolitan regions in Amazon. Also, in these 20 municipalities, more than 90% of the population lived within urban areas in 2010, a percentage much higher than the average 73% of the whole region, and even higher than the average for Brazil.

Specifically in respect to the correlation between regional growth and urbanization, tables 2.3 to 2.5 evidences that Urban, NEG and "urban system" models reviewed in previous sections

[13] According to IBGE criteria.

seem to make perfect sense for the case of Brazilian Amazon. In Table 2.3, we can see that Human Development Indexes in 2010 for the top 20 major population cities are all above the regional average of 0.66, with the exception of the city placed in 20th in this rank. Also, employment growth in the last decade is below the region's average of 3.3% per year in only 4 out of the 20 most populated cities. By their turn, urbanization variables are also mostly above Brazilian Amazon average, with employment in the terciary sector being superior to the region's average in 19 out of 20 cases for the top 20 most populated cities of Brazilian Amazon. Thus, one might say that at least for the most populated cities in Amazon, this is another evidence that the theories which argue that higher urbanization levels are accompanied by higher growth and development might be correct in the case of Brazilian Amazon.

In order to visualize a different perspective of this result, table 2.4 ranks the top 20 most developed cities of Brazilian Amazon, according to the HDI (Human Development Index) for the year of 2010. And once again, a strong correlation between urbanization (measured by the share of employment in the terciary sector and the share of urban population) and economic growth and development is found: only one out of the 20 most developed cities of Brazilian Amazon showed to be less urbanized than the region's average, according to the share of urban population in these cities. According to the share of employment in the terciary sector, all these 20 cities are more urbanized than the region's average. Moreover, in both tables 2.3 and 2.4, the levels of employment growth in the last decade for the top 20 municipalities of each table proved to be, most of the times, above the region's average. This, once again, represents an evidence showing a positive correlation between urbanization and growth within the Brazilian Amazon. Furthermore, evidence from these tables might also be considered as long-run equilibriums, since the per capita GDP level of these cities are also all above the Brazilian Amazon's average.

In order to expand the results from these top 20 tables to numbers which consider the full set of Brazilian Amazon municipalities, we have directly calculated simple correlation coefficients between urbanization and growth in the Amazon region, measured by the same variables as in tables 2.3 and 2.4. These calculations resulted in a positive significant[14] correlation coefficient of 0.3 between the share of employment in the terciary sector and the

[14] All "significant" terms at this paragraph refer to an 1% level of significance of an unicaudal F test of the null hypothesis Ho: $corr(x_u, x_g) > 0$, with x_u representing the "urbanization variable, and x_g representing the economic growth (or development) variable.

HDI level in the year of 2010, a positive and significant correlation of 0.21 between HDI and the share of urban population in the year of 2010, a positive and significant correlation coefficient of 0.15 between per capita GDP growth (from 2000 to 2010) and the share of employment in the terciary sector in the year of 2010, a positive and significant correlation coefficient of 0.34 between employment growth (from 2000 to 2010) and the share of employment in the terciary sector in the year of 2010, and a positive and significant correlation coefficient of 0.21 between employment growth (from 2000 to 2010) and the share of urban population in the year of 2010. Thus, all results point to a simple positive relationship between urbanization, growth and development.

Finally, results from table 2.3 and 2.4 refer mainly to the most populated municipalities of Amazon. Thus, in order to test if these positive correlations found between urbanization and growth remain positive and significant also in the medium and small cities of Brazilian Amazon, we have divided our sample in three parts: one containing only cities with more than 300,000 inhabitants (large cities); other containing only municipalities whose population was between 50,000 and 300,000 inhabitants in 2010 (medium cities); and another containing only municipalities with less than 50,000 inhabitants (small cities). In table 2.5, we only exhibit the top 20 most developed cities among the medium-sized cities, but results found for small and large cities were very similar to the ones presented in this table, to which concerns the focus of our analysis. Specifically, all positive correlations between urbanization, development and growth remain very similar to the ones observed in the previous tables. Thus, predictions from theoretical models of urban economics, NEG and macroeconomic growth seem to apply for almost all Amazon cities, according to this preliminary approach. Still, we emphasize that this analysis is purely preliminary and descriptive, with no regards on causality (between urbanization and development), consistency, or even the influence of other variables over growth and development. These statistical improvements are made on the next section through the econometric approach implemented here. However, we are able to say that this very first descriptive exploration already shed light on some of our goals concerning the empirical relationship between urbanization and growth in Brazilian Amazon.

2.4. Econometric Analysis

2.4.1. Estimators

In order to implement our econometric models, we have selected 6 different kinds of estimators. The reasons for such selection is because each present different characteristics, advantages and disadvantages, and the comparison between all of them permits us to measure the robustness of our methods.

The first one is the traditional OLS, with the coefficients' standard deviations robust to heteroskedasticity (through the also traditional White method) of the error terms. We called this estimator "White robust OLS". This estimator was computed for it is the easier to interpret, as well as the most traditional estimator used in econometrics. Thus, it serves as a the basic starting point of the analysis. However, it does not take into consideration possible important neighborhood effects or spatial correlations which are probably present when the dependent variable concerns economic growth and development, as showed by Fingleton (2003) and Fingleton et al. (2005). This may cause coefficients to be inefficient, if the error terms of the regressions are spatially correlated, or even biased, if the spatial lag of the endogenous variable is an omitted significant variable. Even if the Spatial lag of the endogenous variable is included in regression estimated by OLS, the results will still be biased, as shown by Anselin (1988). This happens because when the lag of the dependent variable is included among the explanatory variables' vector, then shocks in any of the exogenous variables of the regression will simultaneously affect y_i and its neighbors, since y_i itself composes the spatial lag of some y_j of the system, with $i \neq j$. This means that a right-hand side variable (the spatial lag) is determined simultaneously to the dependent variable of the model. And such simultaneity must be taken into account by the estimator, otherwise its estimated coefficients will be biased, which is exactly the case of the OLS estimator. For a detailed discussion of this point, see Anselin (1988).

Therefore, in order to be able to include consistently and efficiently the neighbor effects in our regressions, we also have computed spatial econometric estimators which take these into account. More specifically, all of our estimators, other than the White-robust OLS, were based on the so called SARAR (Spatial Autoregressive Regression with Auto Regressive disturbances) specification. This specification follows this compact notation:

$$Y = \lambda WY + X\beta + u \tag{2.5}$$

$$u = \rho Mu + \varepsilon \tag{2.6}$$

where *Y* is a *nx1* vector of observations of the endogenous variable; *X* is a *nxk* matrix of the n observations of the *k* right-hand side exogenous variables of the model; *W* and *M* are spatial-weights matrices with zero diagonal values, with each of their elements representing the weight that each neighbor of municipality *i* exert over the endogenous variables of this city; ε is a *nx1* vector of disturbances (also called innovations); *WY* and *Wu* are *nx1* vectors referred by literature as spatial lags, with ρ and λ being scalars parameters which represent the spatial correlations of the error term and the endogenous variable lag, respectively.

In all of our estimations, we assume that *W* = *M*, and also that each of element of this spatial weights matrix is given by $w_{ij} = 1/d_{ij}$, where d_{ij} is the distance between cities *i* and *j*. The interpretation of these weights is as follows: every municipality of the sample is considered to be a neighbor of any city *i*, however, the closer one neighbor city is, the higher is the weight of importance attributed to this city as a neighbor. This proximity assumption is a common feature on empirical spatial econometric analysis, and is supported by many models in which transportation costs are directly proportional to distance.

Our first SARAR estimator, which we call "SARAR ML" is simply the SARAR model as specified above, estimated by the Maximum Likelihood method, as in Anselin (1988). This method assumes specifically that $\varepsilon \sim N(0, \sigma^2 I)$, which means that innovations are homoskedastic and normally distributed. Such Maximum Likelihood method follows the estimation procedure proposed by Anselin (1988), which corrects the inconsistency caused by the simultaneity bias of the right-hand side spatial lag variable, when this is included among the regressors. Due to that, literature implemented this estimator extensively over the last two decades, and in order to make our results comparable to such literature, we also computed it. However, this estimator holds on two strong assumptions which may represent statistical issues to our results. Firstly, the homoskedastic assumption may be too strong for our database. As Kelejian and Prucha (2010) argue, "*Spatial units are often heterogeneous in important characteristics, e.g., size, and hence the homoskedasticity assumption may not hold in many situations (conditionally and unconditionally)*". This seem to be the case for our sample, since municipalities in the Brazilian Amazon usually differ in size, economic activities, area, and even vegetation, mostly because of its very extensive overall area. According to Lee (2004), when innovations are heteroskedastic, ML estimators are inconsistent. Secondly, the normal distribution of the disturbances might also not be the case of our regressions. If this is so, ML estimators are also biased, and some problematic results concerning the spatial lags may be observed. Specifically, we may face explosive patterns of the spatial lags' coefficients ($\lambda \rightarrow ^{+}_{-} 1$ and/or $\rho \rightarrow ^{+}_{-} 1$) if such distributional hypothesis does not hold, which once again imply in inconsistent estimations.

Therefore, in order to avoid all those problems, we have also implemented an estimator that we have called "SARAR GMIV", which is the estimator developed by Kelejian and Prucha (2010). This estimator follows the same structural form of the SARAR model, but is estimated by an Generalized Spatial Two Stages Least Squares (GS2SLS) in which innovations ε are assumed to be IID (identically and independently distributed) and heteroskedastically distributed, with heteroskedasticity of any unknown form. Thus, it corrects the bias caused by possible heteroskedastic innovations. Moreover, the natural simultaneity bias in the OLS estimator caused by the autoregressive right-hand side variable is also avoided in the GS2SLS method, by the use of a two stages instrumental variable method in the estimation of the autoregressive (spatial lag) parameter. The set of instruments used in this method are linearly independent columns of the set $\{X, WX, ..., W^{g}X\}$. Thus, this estimator can be considered more robust and consistent than the OLS or the ML estimators, under any circumstances. For a more detailed discussion of this estimator, see Kelejian and Prucha (2010).

One practical problem of implementing SARAR models in our estimations concerns the interpretation of the exogenous variables coefficients, β. This is so because including the spatial lag of the dependent variable in the regression implies simultaneity of outcomes determination. More specifically, if one observation value of an exogenous variable increases, the total resulting change in the endogenous left-hand-side variable will depend on this exogenous variable coefficient, but also on the change of the autoregressive right-hand side variable, which also varies, responding to the original exogenous variable shock, due to its very definition of being a simultaneously determined autoregressive term. For a detailed discussion of this topic, see Anselin (2003); Kelejian and Prucha (2007); and LeSage and Pace (2009). To deal with this matter we have implemented two different strategies. Firstly, when necessary, we followed literature's advice (see Drukker & Prucha, 2011), and computed the marginal effects (from the SARAR regressions) of the interest explanatory variables of which we wished to interpret the coefficients. The methods and formulas by which we implement the calculation of these marginal effects are detailed in the following section of this chapter. Secondly, we have also computed an estimator which we called "SEM GMIV". This estimator follows exactly the same GS2SLS method described above for the SARAR method, but with a fundamental difference in the structural form of the model: we assume that the spatial lag in the right-hand side of regressions equals zero (i.e. $\lambda = 0$), leaving all possible spatial correlations to the spatial error term (ρ), which justifies the use of the terminology SEM (Spatial Error Model) in the name of the estimator. Of course, this may be considered as a more restrictive method, especially if the spatial lag is a significant omitted variable, in which case estimations might be biased. However, it still serves as an easier way to interpret coefficients, since we are able to compare the coefficients estimated by this "SEM GMIV" method to the ones estimated by the "SARAR GMIV" method. In this case, if no big differences are observed among the respective coefficients of each method, then results may be considered robust, in general. More importantly, results from this SEM estimators proved to be practically the same as the ones observed in the SARAR models in terms of the β coefficients estimated by each method, which means that the choice for one method or the other does not matter significantly, in terms of analyzing the effects of the exogenous urbanization variables (which are main interest explanatory ones in our analysis, as previously discussed) over development and growth.

Furthermore, two[15] more estimators were computed. These are the ones we have called "SARAR GMIViv" and "SARAR GMIViv >10k". The only main difference between the two of them is that the "SARAR GMIViv >10k" was computed for a reduced sample in which we included only municipalities whose total population in 2000 was above 10,000 inhabitants, while the "SARAR GMIViv" was computed using the full sample. The reason for this sample cut option (in the case of the "SARAR GMIViv >10k" estimator) was in order to exclude from the sample some possible outlier values of growth, caused by the fact that some overly small cities may present an initial level of the endogenous variable (employment level, for instance) which is also excessively small, which by its turn may cause the growth rate represented by such endogenous variable to be excessively big simply due to the excessively small basis, and not because of an actual expressive growth. These two final estimators are estimated using Druker et al. (2010) and Arraiz et al. (2010) methods, which are implemented through a framework very similar to the one used for the GS2SLS's Keilejian and Prucha

[15] In fact, many other estimations were made, in which we have changed structural forms, the independent variables set, tested for other estimators, and so on. But for practical reasons, we have selected the estimators that we considered to be the most important in terms of robustness, and also in terms of completeness and possibility of comparison with the literature. Moreover, the results were practically the same whatever the estimator or specification we used, and for such, we have an additional reason to trust in the robustness of our results. These other discarded estimations results will show on the Appendix of this article in the future.

(2010) methodology. The only difference between this latter and the two former is that in the ones we called “GMIViv”, the presence of endogenous right-hand side variables is allowed, and the endogeneity bias that these variables would cause is treated by an instrumental Variables (IV) estimation, which adds up to the IV estimation already used to deal with the spatial lag simultaneity issue.

The reason why we also have computed these estimators is because two of the most important variables in our analysis, which are the transportation costs to São Paulo and to the nearest capital, may possibly be endogenous due to omitted significant variable bias: it is possible that a city with lower transportation costs to specific markets tend to grow faster than others exactly due to these reduced transportation costs, in which treating transportation costs as exogenous would be a correct approach. However, it is also possible that there is an unobserved variable, such as investments in infra-structure in the previous period, that determines simultaneously both lower transportation costs and the following GDP per capita and/or employment growth. That being put, we have chosen to instrumentalize transportation costs by the following instrumental variables: distance to São Paulo as an instrument for Transport Costs to São Paulo; and distance to the nearest market as an instrument for Transport Costs to the nearest market. These “distance” instruments are common in literature, and are considered to be strong instruments, since they are clearly highly correlated with the endogenous transportation costs, whereas they are not correlated to the same unobserved (as in the example, “previous investments on infra-structure”) omitted variable, once these distances are simply determined by the cities’ geographical localization, which by their turn are determined historically several years before our sample could be gathered.

Another reason by which we have estimated regressions with the shortened sample with only municipalities whose population were above 10,000 inhabitants was to check if results were robust even for cities which already reached a certain level of growth in the past. Or put in other words, we have estimated these regressions in order to to check whether the results applied to both smaller and bigger towns. As will be clear in the following sections, results did not prove to be much different between all these estimators, which shows that growth process taking place in the Amazon region seem to be homogenous in terms of the size of the cities being considered in the sample.

2.4.2. Econometric Results and Discussion

Tables 2.6, 2.7 and 2.8 bring the results concerning which are the main determinants of employment growth, per capita GDP growth and HDI-M growth in Brazilian Amazon, specifically shedding light on how the urbanization process taking place within the region, along with other variables, interfere the economic growth and development of the region. In what follows, we first highlight a few important remarks about the estimation process, and then discuss the results from these tables.

As a first remark, it is important to notice that all of our 6 chosen estimators resulted in similar coefficients between regressions for each dependent variable, both in terms of significance and signal. This is a positive characteristic in terms of the robustness of the results, for it indicates that in spite of the different assumptions behind each estimator, general evidence which has been found seem to persist. Moreover, it allows us to focus our discussion in the results without the need of focusing on one singular estimator, except for the cases where one particular variable's coefficient from two or more methods are statistically different (in terms of sign or significance).

A second important remark is that the endogenous variables growth rates (employment, GDP per capita and HDI-M growth rates) were calculated following literature on the subject (see Fingleton 2003, Gleaser et al. 1995 and Barro & Sala-i-Martin 1992) using equation 2.7:

$$\Delta\%Y = [\ln(Y_T) - \ln(Y_t)]/(T-t) \tag{2.7}$$

where Y can be either GDP per capita; employment or the HDI-M Index, T is the most recent time period (2010) and t represents the year 2000.

Another important remark regards exogeneity of the independent variables in our models (vector X in equation 2.5). Again, following literature, our right-hand side variables refer to a year which happened previously than the period to which the endogenous growth rates refer. This strategy implies that bias caused by reverse causality is most likely avoided for all variables, since it seems hard to believe that whatever occurred in future time periods might explain values which refer to prior dates.

However, two important exceptions had to be made, in the sense that two variables from the right-hand side vector do not refer to periods prior to the ones encompassed by the endogenous variables' growth rates. These variables are: exports per capita and *bolsa família* per capita. The reason for that is because these were not available for the year 2000, as the *bolsa família* government program started at the year of 2004, and no data on exports is available per municipality before 2004. Nevertheless, we chose to include those variables in the regressions either way, in order to avoid the classical omitted variable bias (see Wooldridge, 2002), which proved to be a good choice given that the respective coefficients of these showed statistical significance in almost every regression, as we will discuss in deeper detail in the following paragraphs. Moreover, we used the aggregated the level of those variables from 2004 to 2007, in order to avoid possible seasonal or outlier effects. This would also be an ideal strategy for the other exogenous variables However, it was impossible to do so, since most of them refer to Census data, which is decennial. Also, these two variables were included in per capita terms in order to avoid simply capturing "size" effects (for example: bigger municipalities might exhibit a higher level of exportations simply because they produce more to both internal and external markets) instead of the desired marginal effects.

A fourth important remark regards the interpretation of the SARAR coefficients. As we have already discussed in the previous sections, the coefficient from these estimators may not be interpreted directly in terms of their magnitude when the Spatial lag is significant. In order to deal with it, we have implemented SEM and OLS estimators to be able to compare the values and significance of these with the ones estimated by the SARAR methods. Regarding this matter, we emphasize that, to what concerns the specific objectives of this study, it suffices to analyze only the statistical significance and the signs of the explanatory variables' coefficients (if they are positive or negative), without having to take into consideration the magnitudes of these. This is because our main purpose here is to simply investigate which are the variables that possibly drive economic growth and development in the Brazilian Amazon region, without trying to measure or compare each of these effects' size.

Nevertheless, in the GDP per capita and the HDI-M regressions, the spatial lag did not proved to be statistically different from zero. Thus, interpretation of the coefficients from Tables 2.7 and 2.8 is, for all 6 models, as if these were all OLS models (see LeSage, 2009). On the other hand, the spatial lag is significant in the employment growth SARAR models from table 2.6, with a positive sign. This positive sign means that even though interpretation of the coefficients should not be made directly, at least their sign and significance do not change. Further explaining this point, if an exogenous variable shows a positive (negative) and significant coefficient on our employment growth SARAR regressions, then the employment growth (decrease) observed in city *i* will also affect positively (negatively) the employment growth in this city's neighbors, which by its turn cause an even higher (lower) growth in city *i* itself, due to the positive sign of the spatial lag variable. Such accumulation of positive effects is called spatial multiplier effects. Therefore, taking into consideration the spatial multiplier effects of the positive and significant spatial lag in SARAR regressions only increase the direct effect of each exogenous variable respective coefficient. And as we are not particularly worried about the size of these coefficients, then we can interpret directly their coefficients' signs on the SARAR GMIV and SARAR GMIViv estimators shown in all tables 2.6, 2.7 and 2.8.

Nevertheless, in order to bring robustness to our results, we have also estimated the marginal effects for some of the right-hand side variables representing urbanization in our employment growth regressions estimated by the SARAR methods, following the recommendation of Drukker et al (2010 and 2011b). More specifically, we have calculated these marginal effects for two variables representing urbanization (as these are our main interest variables in the analysis): the share of employment in the terciary sector and the share of the municipality urban population. For each of these variables and their respective coefficients, we have calculated two kinds of Marginal effects: the ATI (Average Total Effect) and the ATDI (Average Total Direct Effect). These were defined in LeSage and Pace (2009) as the following: the ATDI represents the average change in the endogenous variable y caused by a sequential change of magnitude δ (also called the "shock", which in our case was an increase of 1% in the value of the chosen explanatory variable) in each of the i observations of one chosen explanatory variable. The ATI, by its turn, represents the average change in the endogenous variable y caused by a simultaneous change of magnitude δ in each of the i observations of one chosen explanatory variable. Computation of these effects was made following Drukker et al. (2010b). The authors argue that these two marginal effects measures tend to present similar values, and recommend calculation of both of them in order to obtain robust results. The specific formulas of calculation can be found at Drukker et al. (2010b).

Table 2. 6: Determinants of Employment Growth

					Transport Costs Instrumented by Distance	
Explanatory Variable	**White Robust OLS**	**SARAR ML**	**SEM GMIV**	**SARAR GMIV**	**SARAR GMIViv**	**SARAR GMIViv >10k**
Employment Level	-0.035	-0.039	-0.038	-0.038	**-0.049***	**-0.070*** **
(Millions of jobs, in 2000)	(-1.472)	(-0.943)	(-1.586)	(-1.520)	**(-1.881)**	**(-2.738)**
Share of Urban EAP	**0.028*** **	**0.027*** **	**0.027*** **	**0.027*** **	**0.031*** **	**0.024** **
(% of EAP, in 2000)	**(3.207)**	**(3.617)**	**(3.038)**	**(3.093)**	**(3.374)**	**(2.396)**
Share of Employment in the Services Sector	**0.058*** **	**0.052*** **	**0.056*** **	**0.052*** **	**0.048*** **	**0.026** **
(%, in 2000)	**(5.252)**	**(5.406)**	**(4.977)**	**(4.594)**	**(4.201)**	**(2.274)**
Metropolitan Region dummy	**0.010** **	0.009	**0.010** **	**0.008***	0.007	0.004
(1 if MR, 0 otherwise)	**(2.384)**	(1.370)	**(2.263)**	**(1.847)**	(1.497)	(0.975)
Theil Index	**0.043*** **	**0.041*** **	**0.042*** **	**0.040*** **	**0.039*** **	**0.040*** **
(Index, in 2000)	**(4.285)**	**(4.281)**	**(4.157)**	**(3.984)**	**(3.795)**	**(3.628)**
Income HDI	**-0.157*** **	**-0.152*** **	**-0.158*** **	**-0.144*** **	**-0.129*** **	**-0.108*** **
(Index, in 2000)	**(-5.093)**	**(-5.486)**	**(-4.910)**	**(-4.643)**	**(-3.943)**	**(-3.567)**
Longevity HDI	-0.010	-0.019	-0.013	-0.022	-0.025	0.004
(Index, in 2000)	(-0.452)	(-0.845)	(-0.560)	(-0.970)	(-1.149)	(0.163)
"Bolsa Família" (04-07) per capita	**-0.002*** **	**-0.002*** **	**-0.002*** **	**-0.002*** **	**-0.002*** **	**-0.002*** **
(Constant 2000 R$, from 2004-2007)	**(-8.728)**	**(-8.856)**	**(-8.592)**	**(-8.246)**	**(-7.997)**	**(-6.917)**
Exports per capita	**0.268*** **	**0.266*** **	**0.267*** **	**0.266*** **	**0.258*** **	**0.197*** **
(Thousands of FOB US$ (2003-2010))	**(5.730)**	**(6.416)**	**(5.752)**	**(5.852)**	**(5.941)**	**(5.503)**
Informality	**-0.019** **	**-0.019** **	**-0.019** **	**-0.019** **	**-0.020** **	**-0.008***
(% of Employment, in 2000)	**(-2.077)**	**(-2.331)**	**(-2.236)**	**(-2.274)**	**(-2.464)**	**(-1.804)**
Share of Superior education (8-12 years of study)	0.001	0.001	0.001	0.001	0.001	-0.000
(% of population, in 2000)	(0.521)	(0.856)	(0.734)	(0.999)	(1.273)	(-0.467)
Share of High Education (4-8 years of study)	**0.001** **	**0.001** **	**0.001** **	**0.001** **	**0.001** **	0.000
(% of population, in 2000)	**(2.096)**	**(2.526)**	**(2.109)**	**(2.427)**	**(2.565)**	(0.201)
Share of illiterates	0.000	0.000	0.000	0.000	0.000	0.000
(% of population, in 2000)	(0.002)	(0.550)	(0.123)	(0.675)	(0.635)	(0.819)
Share of Low Education (0-4 years of study)	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
(% of population, in 2000)	(-0.459)	(-0.951)	(-0.610)	(-0.981)	(-0.969)	(-0.682)
Transport Costs Index to São Paulo	**0.005*** **	**0.006*** **	**0.005*** **	**0.006*** **	**0.009*** **	**0.005*** **
(Index, in 1995)	**(3.915)**	**(4.630)**	**(3.307)**	**(4.396)**	**(5.449)**	**(2.582)**
Transport Costs Index to the nearest Capital	**-0.010*** **	**-0.011*** **	**-0.010*** **	**-0.011*** **	**-0.014*** **	**-0.007** **
(Index, in 1995)	**(-5.372)**	**(-5.993)**	**(-4.811)**	**(-5.965)**	**(-5.257)**	**(-2.545)**
Share of Agriculture GDP	0.074	0.064	0.070	0.063	0.050	0.015
(% of GDP, in 2000)	(1.856)	(1.381)	(1.742)	(1.596)	(1.227)	(0.279)
Share of Services GDP	0.049	0.038	0.044	0.038	0.018	-0.012
(% of GDP, in 2000)	(1.125)	(0.788)	(1.009)	(0.869)	(0.395)	(-0.222)
Share of Manufacture GDP	0.038	0.030	0.035	0.029	0.010	-0.023
(% of GDP, in 2000)	(0.848)	(0.578)	(0.784)	(0.671)	(0.225)	(-0.419)
Deforested Area	-0.006	-0.006	-0.007	-0.005	-0.006	-0.009
(% of Area, in 2001)	(-1.609)	(-1.481)	(-1.440)	(-1.320)	(-1.569)	(-2.152)
Neighbours Deforested Area	0.002	-0.009	0.001	-0.016	-0.013	-0.009
(% of Area, in 2001)	(0.227)	(-0.888)	(0.061)	(-1.512)	(-1.230)	(-1.021)
Constant	-0.033	-0.026	-0.023	-0.036	-0.043	0.063
	(-0.559)	(-0.428)	(-0.372)	(-0.603)	(-0.709)	(0.956)
Lambda (Spatial lag)		**0.439***		**0.758** **	**0.928*** **	**0.803*** **
		(1.861)		**(2.460)**	**(3.069)**	**(2.612)**
Rho (Error lag)		0.395	**0.920*** **	0.206	0.106	0.283
		(1.293)	**(3.722)**	(0.596)	(0.308)	(0.835)
OBS	751	751	751	751	751	450

Notes: *t* statistic in parenthesis (); ***significant at 1%; **significant at 5%; *significant at 10%

SOURCE : Own elaboration.

Table 2. 7: Determinants of GDP per capita growth

Explanatory Variable	**White Robust OLS**	**SARAR ML**	**SEM GMIV**	**SARAR GMIV**	Transport Costs Instrumented by Distance **SARAR GMIViv**	 **SARAR GMIViv >10k**
GDP per capita	**-0.004*** **	**-0.004*** **	**-0.004*** **	**-0.004*** **	**-0.004*** **	**-0.006*** **
A thousand Constant 2000 R$ per inhabitant, in 2000	**(-4.951)**	**(-8.031)**	**(-4.993)**	**(-4.941)**	**(-4.604)**	**(-6.186)**
Share of Urban EAP	0.002	-0.003	-0.000	-0.001	-0.005	0.005
(% of EAP, in 2000)	(0.249)	(-0.322)	(-0.060)	(-0.140)	(-0.591)	(0.405)
Share of Employment in the Services Sector	-0.012	-0.006	-0.009	-0.008	-0.006	-0.017
(%, in 2000)	(-1.046)	(-0.611)	(-0.778)	(-0.729)	(-0.575)	(-1.359)
Metropolitan Region dummy	**0.016** **	**0.016** **	**0.016** **	**0.016** **	**0.018** **	**0.018** **
(1 if MR, 0 otherwise)	**(2.143)**	**(2.431)**	**(2.129)**	**(2.131)**	**(2.386)**	**(2.186)**
Theil Index	**-0.026** **	**-0.025** **	**-0.025** **	**-0.026** **	**-0.026** **	**-0.029** **
(Index, in 2000)	**(-2.350)**	**(-2.467)**	**(-2.314)**	**(-2.404)**	**(-2.347)**	**(-2.407)**
Income HDI	**0.072** **	**0.069** **	**0.070** **	**0.075** **	**0.072** **	**0.060**
(Index, in 2000)	**(2.323)**	**(2.307)**	**(2.192)**	**(2.377)**	**(2.039)**	**(1.394)**
Longevity HDI	**-0.076*** **	**-0.063*** **	**-0.068*** **	**-0.066*** **	**-0.070*** **	**-0.098*** **
(Index, in 2000)	**(-3.340)**	**(-2.727)**	**(-2.928)**	**(-2.796)**	**(-2.863)**	**(-2.698)**
"Bolsa Família" (04-07) per capita	-0.069	-0.019	-0.032	-0.035	0.029	0.297
(A thousand Constant 2000 R$, from 2004-2007)	(-0.279)	(-0.075)	(-0.130)	(-0.147)	(0.116)	(0.894)
Exports per capita	**0.460*** **	**0.458*** **	**0.460*** **	**0.461*** **	**0.448*** **	**0.579*** **
(Thousands of FOB US$ (2003-2010))	**(4.135)**	**(8.318)**	**(4.127)**	**(4.116)**	**(3.892)**	**(7.928)**
Informality	0.011	0.010	0.011	0.011	0.011	0.016
(% of Employment, in 2000)	(0.805)	(1.171)	(0.808)	(0.831)	(0.823)	(1.438)
Share of Superior education (8-12 years of study)	**0.002***	**0.002***	**0.002***	**0.003** **	**0.002***	0.002
(% of population, in 2000)	**(1.870)**	**(1.904)**	**(1.871)**	**(2.085)**	**(1.754)**	(1.497)
Share of High Education (4-8 years of study)	**0.001*** **	**0.001*** **	**0.001*** **	**0.002*** **	**0.001*** **	0.001
(% of population, in 2000)	**(3.289)**	**(2.993)**	**(3.145)**	**(3.470)**	**(2.854)**	(1.494)
Share of illiterates	0.000	-0.000	0.000	0.000	0.000	-0.000
(% of population, in 2000)	(0.164)	(-0.039)	(0.058)	(0.342)	(0.253)	(-0.225)
Share of Low Education (0-4 years of study)	**-0.001** **	**-0.001***	**-0.001** **	**-0.001** **	**-0.001** **	-0.001
(% of population, in 2000)	**(-2.099)**	**(-1.776)**	**(-1.965)**	**(-2.215)**	**(-1.987)**	(-1.335)
Transport Costs Index to São Paulo	**-0.011*** **	**-0.011*** **	**-0.011*** **	**-0.010*** **	**-0.013*** **	**-0.011*** **
(Index, in 1995)	**(-7.232)**	**(-7.759)**	**(-6.314)**	**(-5.662)**	**(-6.334)**	**(-4.219)**
Transport Costs Index to the nearest Capital	**0.012*** **	**0.011*** **	**0.011*** **	**0.010*** **	**0.015*** **	**0.011*** **
(Index, in 1995)	**(6.471)**	**(5.434)**	**(5.552)**	**(4.969)**	**(4.679)**	**(2.648)**
Share of Agriculture GDP	**0.125*** **	**0.146*** **	**0.136*** **	**0.142*** **	**0.154*** **	**0.169** **
(% of GDP, in 2000)	**(2.592)**	**(3.018)**	**(2.813)**	**(2.895)**	**(3.068)**	**(2.553)**
Share of Services GDP	**0.197*** **	**0.218*** **	**0.208*** **	**0.214*** **	**0.233*** **	**0.232*** **
(% of GDP, in 2000)	**(3.713)**	**(4.266)**	**(3.915)**	**(3.979)**	**(4.221)**	**(3.390)**
Share of Manufacture GDP	**0.153** **	**0.178*** **	**0.166*** **	**0.172*** **	**0.191*** **	**0.193** **
(% of GDP, in 2000)	**(2.569)**	**(3.280)**	**(2.785)**	**(2.872)**	**(3.109)**	**(2.464)**
Deforested Area	-0.004	-0.003	-0.003	-0.001	-0.001	-0.002
(% of Area, in 2001)	(-0.954)	(-0.607)	(-0.644)	(-0.307)	(-0.291)	(-0.410)
Neighbours Deforested Area	**-0.020** **	-0.022	**-0.022***	**-0.039*** **	**-0.031***	-0.024
(% of Area, in 2001)	**(-2.207)**	(-1.544)	**(-1.655)**	**(-2.652)**	**(-1.860)**	(-0.986)
Constant	**-0.138** ** **(-2.021)**	**-0.159** ** **(-2.400)**	**-0.151** ** **(-2.204)**	**-0.176** ** **(-2.491)**	**-0.164** ** **(-2.274)**	**-0.090** **(-1.048)**
Lambda (Spatial lag)		-0.003 (-0.017)		0.338 (1.611)	0.122 (0.492)	-0.268 (-0.584)
Rho (Error lag)		0.766 .	**0.826*** ** **(2.582)**	0.481 (1.087)	**0.759*** **(1.720)**	**1.327** ** **(2.365)**
OBS	751	751	751	751	751	450

Notes: *t* statistic in parenthesis (); ***significant at 1%; **significant at 5%; *significant at 10%

SOURCE : Own elaboration.

Table 2. 8: Determinants of HDI-M growth

HDI-M growth (yearly, %)						
					Transport Costs Instrumented by Distance	
Explanatory Variable	**White Robust OLS**	**SARAR ML**	**SEM GMIV**	**SARAR GMIV**	**SARAR GMIViv**	**SARAR GMIViv >10k**
HDI-M in 2000	**-0.161*** **	**-0.172*** **	**-0.166*** **	**-0.166*** **	**-0.167*** **	**-0.162*** **
Index	**(-21.685)**	**(-31.101)**	**(-22.245)**	**(-22.559)**	**(-21.565)**	**(-19.630)**
Share of Urban EAP	**0.004*** **	**0.002***	**0.003*** **	**0.003*** **	**0.003** **	**0.004** **
(% of EAP, in 2000)	**(3.359)**	**(1.801)**	**(2.711)**	**(2.636)**	**(2.304)**	**(2.555)**
Share of Employment in the Services Sector	-0.003	0.001	-0.001	-0.000	-0.000	-0.000
(%, in 2000)	(-1.721)	(0.861)	(-0.465)	(-0.271)	(-0.127)	(-0.147)
Metropolitan Region dummy	0.001	0.001	0.001	0.001	**0.002***	**0.002*** **
(1 if MR, 0 otherwise)	(1.451)	(1.318)	(1.577)	(1.632)	**(1.890)**	**(2.589)**
Theil Index	-0.002	-0.001	-0.001	-0.001	-0.001	-0.001
(Index, in 2000)	(-1.344)	(-0.687)	(-1.159)	(-1.058)	(-1.018)	(-0.545)
"Bolsa Família" (04-07) per capita	**0.067***	0.033	**0.061***	0.051	**0.058***	**0.079***
(A thousand Constant 2000 R$, from 2004-2007)	**(1.892)**	(0.970)	**(1.841)**	(1.508)	**(1.722)**	**(1.955)**
Exports per capita	**0.015*** **	**0.017*** **	**0.017*** **	**0.017*** **	**0.017*** **	**0.020*** **
(Thousands of FOB US$ (2003-2010))	**(2.672)**	**(2.881)**	**(3.326)**	**(3.192)**	**(3.232)**	**(3.506)**
Informality	0.001	0.001	0.001	0.001	0.001	0.001
(% of Employment, in 2000)	(1.593)	(0.460)	(1.041)	(0.902)	(0.832)	(1.302)
Share of Superior education (8-12 years of study)	**0.001*** **	**0.001*** **	**0.001*** **	**0.001*** **	**0.001*** **	**0.001*** **
(% of population, in 2000)	**(4.598)**	**(4.259)**	**(4.653)**	**(4.840)**	**(4.510)**	**(3.373)**
Share of High Education (4-8 years of study)	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
(% of population, in 2000)	(-0.527)	(-1.138)	(-0.627)	(-0.413)	(-0.865)	(-0.616)
Share of illiterates	-0.000	**-0.000*** **	-0.000	-0.000	-0.000	**-0.000***
(% of population, in 2000)	(-0.959)	**(-2.879)**	(-1.431)	(-1.546)	(-1.538)	**(-1.770)**
Share of Low Education (0-4 years of study)	**-0.000***	**-0.000** **	**-0.000***	**-0.000***	**-0.000***	-0.000
(% of population, in 2000)	**(-1.768)**	**(-2.056)**	**(-1.888)**	**(-1.931)**	**(-1.835)**	(-0.921)
Transport Costs Index to São Paulo	**-0.002*** **	**-0.001*** **	**-0.001*** **	**-0.001*** **	**-0.002*** **	**-0.001*** **
(Index, in 1995)	**(-6.699)**	**(-6.124)**	**(-4.616)**	**(-5.119)**	**(-5.341)**	**(-3.354)**
Transport Costs Index to the nearest Capital	**0.001*** **	**0.001*** **	**0.001** **	**0.001** **	**0.002*** **	**0.001** **
(Index, in 1995)	**(3.484)**	**(3.126)**	**(2.297)**	**(2.525)**	**(3.402)**	**(1.970)**
Share of Agriculture GDP	**-0.026*** **	**-0.014** **	**-0.020*** **	**-0.018*** **	**-0.017*** **	**-0.035*** **
(% of GDP, in 2000)	**(-4.154)**	**(-2.060)**	**(-3.304)**	**(-3.084)**	**(-2.806)**	**(-4.338)**
Share of Services GDP	**-0.029*** **	**-0.015** **	**-0.022*** **	**-0.020*** **	**-0.018*** **	**-0.037*** **
(% of GDP, in 2000)	**(-4.221)**	**(-2.086)**	**(-3.284)**	**(-3.049)**	**(-2.688)**	**(-4.233)**
Share of Manufacture GDP	**-0.027*** **	**-0.012**	**-0.019*** **	**-0.017*** **	**-0.015** **	**-0.036*** **
(% of GDP, in 2000)	**(-3.794)**	**(-1.645)**	**(-2.890)**	**(-2.646)**	**(-2.303)**	**(-4.077)**
Deforested Area	0.001	**0.002*** **	**0.002*** **	**0.002*** **	**0.002*** **	**0.002*** **
(% of Area, in 2001)	(1.402)	**(2.647)**	**(2.765)**	**(3.011)**	**(3.016)**	**(3.054)**
Neighbours Deforested Area	-0.002	0.000	-0.002	-0.005	-0.004	0.002
(% of Area, in 2001)	(-1.622)	(0.022)	(-1.238)	(-1.427)	(-1.038)	(0.412)
Constant	**0.144*** **	**0.139*** **	**0.140*** **	**0.138*** **	**0.140*** **	**0.150*** **
	(14.166)	**(13.803)**	**(14.010)**	**(13.302)**	**(13.293)**	**(11.844)**
Lambda (Spatial lag)		-0.028		-0.004	-0.024	-0.124
		(-0.485)		(-0.047)	(-0.326)	(-0.919)
Rho (Error lag)		**1.991*** **	**1.637*** **	**1.066*** **	**1.131*** **	**1.551*** **
		(37.543)	**(5.025)**	**(5.982)**	**(6.183)**	**(5.830)**
OBS	751	751	751	751	751	450

Notes: *t* statistic in parenthesis (); ***significant at 1%; **significa

SOURCE : Own elaboration.

By observing the results in tables 2.6 to 2.8, the first evidence which calls attention concerns the relative convergence result observed for the three dependent variables. Specifically, as the theoretical macroeconomic growth models predict, coefficients of the per capita GDP and HDI-M in 2000 are statistically significant and negative in all models. For the employment growth regressions, the initial level of employment in 2000 is negative in all models too, but it is significant only in the SARAR GMIV models in which transportation costs are instrumentalized by distance. Thus, in general, this may be interpreted as an evidence that the neoclassical growth models (see Fingleton, 2003) prediction of relative convergence in growth and development among different cities of the same region is confirmed within Brazilian Amazon. In other words, this means that development measured by per capita GDP and HDI-M index have been relatively higher in cities which presented a lower level of development in the baseline year 2000, in the last decade. Thus, one might say that our results suggest that there is a catching up process occurring in Brazilian Amazon region. In turn, this can be interpreted as first evidence that economic growth within the region may be being determined increasingly endogenously within Brazilian Amazon as time passes, as this evidences that all local markets seem to be converging to a higher level of development, in the region as a whole. Or, put in other terms, as this result points that regional markets seem to be converging in terms of growth and development, this may be interpreted as a sign that the region's economy will probably become less dependent of external markets as being the main determinant of local growth and development.

Moving straightforward to the main interest variables in our analysis, our results show robust evidence that the urbanization process taking place within the Brazilian Amazon region seem to be causing a relevant positive influence on both regional development and growth. For all three groups of endogenous variables, we find evidence that a higher level of urbanization in the initial period resulted in higher growth and development in the following 2000-2010 decade.

Specifically, we observe that a higher share of urban EAP resulted robustly in higher subsequent HDI-M and employment growth, according to all estimators. Furthermore, results evidence that cities located within the Amazon metropolitan regions showed higher growth rates of all three endogenous variables, especially in terms of per capita GDP, to which all estimators showed positive and significant coefficients for the metropolitan region dummy variable. In terms of employment and HDI growth, this specific metropolitan region dummy result was less robust, as significance only appeared after we controlled for possible endogeneity of transportation costs for the latter (and more robustly after we exclude from the sample the small cities with less than 10.000 inhabitants), whereas for the former, coefficients were only significant to estimators which did not controlled for the this possible endogeneity. Thus, this result may be considered weak evidence in terms of HDI-M and employment growth, and strong evidence in terms of per capita GDP. As for the third representative of urbanization in our estimations, we can see that the share of employment in the terciary sector (which is typically concentrated in urban areas) is only significant for the employment growth regressions, with positive sign, which means that a greater share of employment in the terciary sector in the initial period causes a higher employment growth rate in the subsequent period, probably due to the multiplicative nature of this kind of employment, and to the agglomeration externalities associated with the urbanization process that this variable represents.

Thus, summing up the results of the explanatory variables representing urbanization over growth and development, we conclude that urbanization seem to be positively causing both growth and development within Brazilian Amazon, especially in terms of employment growth, which showed positive results for all three variables. As we have also found a positive and significant spatial lag in the employment growth regressions estimated by the SARAR methods, we have chosen to calculate the marginal effects of these variables in these regressions, in order to bring more robustness to our results, because a simple analysis of the direct urbanization coefficients may not be capturing the whole effect that these exert on employment growth, even though the spatial lag proved to be positive in these regressions, and as we have argued, this means that direct positive coefficients of urbanization variables will only be intensified towards the same direction by this spatial lag. Results are shown in Table 2.9, which bring the two measures of the Marginal effects, ATDI and ATI, as previously described, for the share of EAP and the share of employment in the terciary sector in the year 2000.

We can see that the marginal effects of the two urbanization representatives are all positive, as expected. Thus, this also corroborates that higher initial urbanization tended to cause higher economic growth between 2000 and 2010, as theoretical models from urban and spatial economics, as well as NEG models would predict.

Table 2. 9: Marginal effects of Initial urbanization level over Employment growth

	ATI Marginal Effect		**ATDI Marginal Effect**	
	Employment Growth (yearly, %)		**Employment Growth (yearly, %)**	
Shocks	1% Raise in the Share of Employment in the Third Sector	1% raise in the Share of Urban EAP	1% Raise in the Share of Employment in the Third Sector	1% raise in the Share of Urban EAP
Marginal Effect	1.82%	1.15%	1.87%	1.18%

SOURCE: Own Elaboration

Furthermore, these results of positive causality and correlation between urbanization and growth in Brazilian Amazon may also be interpreted as another evidence which suggests that Brazilian Amazon seems to be passing through a process of endogeneisation of its economic growth and development, in the sense that local economic growth is becoming more self-sustainable, and less economically dependent on external markets from southern Brazil and abroad. Such interpretation is due to arguments given by the theoretical models which support our analysis: New Economic Geography (see Krugman, 1990, 1991 and 1994) models argue that urbanization is closely linked to one region's internal growth through increasing returns of scale; urban economic models (see Igliori, 2009, and Fujita & Thisse, 2002) argue that urbanization comes accompanied with agglomeration externalities which increase the economic growth of one region internally; Jacobs (1964) argue that urbanization creates opportunities for the development of innovations and the creation of "new work" (workforce with higher productivity) within the region borders, therefore, also increasing economic growth and development "internally". And our results suggest that all, or at least part of these theories predictions seem to be occurring in Brazilian Amazon from 2000 to 2010, in terms of the economic impacts that the region's urbanization might be bringing.

To which regards the econometric results of the other explanatory variables of our estimations, besides the ones that represent urbanization, we have classified these results in two groups, according to the objectives proposed in this chapter. These groups are: 1) the group of variables which serve mostly as "control" variables; whose results are interesting and relevant, but do not integer the specific aims of our research, and therefore need further research; 2) the group of explanatory variables whose results complement the analysis regarding the urbanization process as a whole, as in the case of the exports per capita and the other spatial variables that compose the exogenous vector.

Among the results that need further specific research, the first that we highlight regards the coefficients of informality in the labor market. On the one hand, estimations indicate that the share of informality in an average Amazon municipality in the year 2000 does not affect per capita GDP or HDI-M growth in the following years, but on the other hand, municipalities with higher levels of formality are positively affected in terms of higher employment growth rates. This result can be interpreted as an indicative that in terms of productivity and development (GDP and HDI), it does not matter if the Brazilian Amazon labor market is either formal or informal[16], whereas in terms of employment multiplier effects, it is possible to argue that stronger institutions in the labor market may push forward the creation of new openings in the future. Why this may be happening only can be answered with further research on the theme.

Another result which deserves further analysis, and is quite is difficult to interpret regards the shares of Services GDP, the share of Manufacturing GDP and the share of Agriculture GDP in 2000, for all three endogenous variables' regressions. More specifically, in tables 2.6 to 2.8 we can see that all these shares exert no signifficant influence on employment growth, but at the same time they all exert a positive influence on per capita GDP growth, as well as all of them influence negatively the HDI-M growth from 2000 to 2010. Therefore, it is not clear how the average GDP composition interferes on economic growth and development according to our results, since all of these shares coefficients in the each regression group (for each endogenous variable) exhibit the same signal. Or, in other words, as each initial GDP share exhibit the same coefficients' signs regarding the same endogenous variable, for each of the three endogenous variables considered in our analysis, it becomes difficult to compare how each GDP share, individually, affects growth or development, as they all affect the endogenous variable in the same direction. Still, in order to try to differentiate these individual effects, one could possibility compare between the size of each GDP share coefficient within the per capita GDP regressions, as well as within the HDI-M regressions (since for employment growth regressions, coefficients are not statistically significant). However, this strategy also leads to inconclusive and contradictory evidence: on the one hand, the services sector seem to be the one which increases per capita GDP growth (as its coefficient is the biggest among the three shares in all per capita GDP regressions), but on the other hand, this same services sector GDP share is the one that mostly causes HDI-M to decrease, as its coefficient is the lowest among the three sectors in HDI-M regressions.

[16] This can be due to an compensation effect, in which lower admission (demission) costs for hiring (firing) a new "informal" employee (instead of a "formal" one) are being compensated by a possible higher level of productivity of "formal" workers. But again, these are assumptions that deserve further research to be confirmed.

This contradictory result could be interpreted as if these shares, in fact, did not exert real influence on growth and development, probably due to the extraordinary diversity of successful (and non-successful) economic activities within the legal Amazon area, which is an enormous region with several kinds of population clusters, different types of soils, economic structures, and so on. For instance, within the Amazon region we can find cities whose economy is mostly based on agricultural activities, especially in the soybean belt region (Mato Grosso). On the other hand, the largest city of Manaus holds the biggest manufacturing pole of the region (*Zona Franca de Manaus*). Other cities, such as Santarém, have an economy which depends heavily on its harbor related services. These examples, together with our results regarding the influence of the initial GDP shares of each sector over local growth and development, leads us to believe that growth within the Brazilian Amazon may have flourished from industries of all kinds in the period of analysis. However, as these are not our main interest variables, and serve mostly as regression controls, we choose not to defend any conclusions in this sense. Still, we kept these variables among the explanatory vector in order to avoid the omission of relevant variables' bias.

Another interesting result regards the explanatory variable "*Bolsa Família*". As already mentioned, the "*Bolsa Família*" consists of a Brazilian Federal government social assistance program designed to reduce poverty and inequality throughout the country. However, evidence from our estimations indicate that higher levels of the "*Bolsa Família*" assistance in an average Brazilian Amazon municipality affects negatively employment growth in this city, does not influence its per capita GDP growth, and at the same time increases the HDI-M Growth. As this program is a subject covered by an extensive literature in Brazil, whose focus is exactly to measure the economic benefits and costs associated with "*bolsa família*" in terms of growth and equality, and this is not our primary goal here, once again we choose not to defend any conclusions regarding these coefficients, and simply treat them as controls for our main interest variables. However, one possible interpretation which we could attribute to our results is as the following: the primary goal of "*bolsa família*", which is to reduce economic inequality in Brazil, seem to be being achieved properly in terms of increasing the HDI in the poorest regions, as it affects positively HDI growth. However, as neoclassical models on labor supply would predict, one possible counter effect is that employment growth in cities which receive more "*bolsa família*" might face a downshift on their labor supply curve, as the "outside option" of laborers increases once the program is established, or its transfer levels increase within one city.

Furthermore, we once again call attention to the fact that our evidence regarding this variable might be considered weak, not only because our model do not exactly focus on it (i.e., might not include all relevant variables to reach conclusions about the efficiency of this social program itself), but also due to the fact that our variable representing "*bolsa família*", differently from the other explanatory variables, is built by the sum of the amount of this assistance per municipality between 2004 and 2007, whereas our endogenous growth rates variables regard the period 2000-2010. Thus, it may be the case that municipalities which received higher amounts of the "*bolsa família*" social assistance were exactly the ones with lower rates of employment or GDP growth recently, not only because the program caused it, but because "*bolsa família"* is exactly designed to be implemented in areas of higher unemployment and lower economic growth. Thus, the results concerning this social assistance program in employment growth regression might be biased due to reverse causality, in this sense. Bearing that in mind, in order to test if this possible reverse causality endogeneity could be affecting the other explanatory variables' coefficients, we have implemented all regressions from tables 4.2-4.4 without including "bolsa família" as a regressor. Fortunately, the other coefficients proved to be robustly estimated in this sense, since they did not change in terms of significance and sign in all regressions. Therefore, we still chose to keep the variable "bolsa família" in the explanatory variables' vector, in order to avoid omitting a possibly relevant variable, and also to shed light on the fact that in terms of increasing development measured by HDI-M, the social assistance program seem to be somewhat effective.

Another result that deserves further detailed analysis concerns the income distribution variables included in the model, more specifically, the Theil Index and the share of the HDI index regarding income distribution. First of all, we note that we have not included the income HDI variable in the HDI-M growth regressions, once we have already included the full HDI in 2000 as an explanatory variable in order to capture the convergence effects predicted by literature, and the income part of this index mathematically accounts for 1/3 of the HDI-M full Index. Either way, the remaining Theil index did not proved to be significant in the HDI growth regressions, probably because income distribution within each city is already being accounted by the HDI index in 2000. Thus, no significant result may be attributed to income in which concerns development growth measured by the HDI-M growth rate regressions.

As for the employment and per capita GDP growth, on the other hand, results concerning these two inequality explanatory variables proved to be significant, but each in an opposite direction: results from employment growth regressions showed robustly (i.e. for all estimators) that a higher initial inequality in 2000, represented by both income HDI and Theil indexes, tended to cause a higher employment growth rate between 2000 and 2010 in Brazilian Amazon, on average. As for the per capita GDP regressions, on the other hand, the evidence from these two same explanatory variables was that the higher the initial levels of inequality in one city, the lower the per capita GDP growth in the following years in that municipality. Thus, by one side, employment creation seems to be benefited by higher initial levels of inequality, while per capita GDP tend to grow more if inequality is lower within one city. This may indicate that growth within Brazilian Amazon might be unbalanced, with few individuals mostly benefitting from urbanization and multiplier effects. However, due to the controversy of this result, and to the complexity which involves inequality and growth, we choose not to draw any conclusions with respect to these variables' results, once again. As well as in the case of the "*bolsa família*" variable, in order to investigate more precisely this relationship, it would be necessary to focus the analysis on income distribution, considering the inclusion of other control variables, and focusing on testing theoretical models behind estimations more consistently, which is not the main goal of this paper.

By their turn, the evidence we have found concerning the educational level of the municipalities as explanatory variables of growth and development are much more expected: our estimations show, robustly (i.e. in all estimators for each dependent variable), that on average, higher shares of Illiterates and Low Education in 2000 caused lower HDI-M growth from 2000 to 2010, higher shares of Superior and High Education in 2000 caused higher per capita GDP growth from 2000 to 2010, and a higher share of High Education in 2000 also caused a higher employment growth in the last decade, in Brazilian Amazon. These are expected results according to literature on the subject (see Fingleton 2003, Romer (1989, 1990 and 1994), Krueger and Lindahl, 2001, Acemoglu and Angrist, 1999), since improvements in the population's educational level is expected to be accompanied by increases in labor productivity, knowledge spillovers and positive externalities. In fact, these results are quite similar to the ones found by Romer (1990) in his empirical analysis of returns to education. Furthermore, to what concerns our analysis, the specific evidence of a positive relationship between Superior education and per capita GDP growth in Brazilian Amazon may be considered associated with the urbanization process taking place within the region, as both

Universities and jobs openings which require College degree are typically located within urban areas.

Complementarily, these results regarding education might be also considered as evidence that economic development within Brazilian Amazon seems to be growingly endogenous and self-sustainable. Put in other terms, these results might be considered as evidence that the economic development of the Brazilian Amazon is increasingly becoming less dependent on external economies. The reason for that is because this evidence agrees with the endogenous growth perspective from Macroeconomic models such as in Romer (1990) and in Lucas (1988), which point out to education as the main determinant of the human capital level of one region, which by its turn increases its endogenous growth capacity. Jacobs (1970) also points to education as the main driver of human capital accumulation, which by its turn determines the capacity of one region to create innovative activities and promote self-sustained growth. Microeconomic models such as Mincer (1973) and Spence (1974) point to educational level as the main determinant of wages through signaling in job market, which by its turn lead to higher economic multipliers and externalities within one region. Thus, the positive correlation between education and growth found for Amazon might also be considered an evidence of the internal endogeneisation of the region's development towards self-sustainable growth.

Such self-sustainable endogeneisation dynamics is also corroborated by the spatial lag variables' coefficients in our estimations. This is so because positive and significant neighbor effects were found for all three dependent variables in our regressions, specifically for the spatial lag variable in the employment growth regressions, and for the spatial error lag in the HDI-M and per capita GDP growth regressions. In general, this means that growth and development within municipalities of Amazon tend to follow a spatial dynamic process of clustering, in which growth in one average municipality creates spillovers effects that intensifies the whole Amazon economy to grow. Such dynamic is typical of an economy which encompasses enough self-sufficiency in terms of demand vectors and local markets, so that the multiplier effects produced by one particular city may spread among its neighbor municipalities.

On the other hand, estimations show that although our results from urbanization, education and spatial neighbors exogenous variables point to an increasing pattern of endogeneisation of the local economic growth towards a self-sustainable path, the variables which represent the region's dependency on external markets still proved to be relevant when it comes to explaining such growth and development. Therefore, it is important to stress that such endogeneisation process should not be considered complete yet. More specifically, we find that all three dependent variables (HDI-M, per capita GDP and Employment growth) are positively (and robustly, i.e. in all estimators) correlated[17] with the level of exports per capita between municipalities of Amazon and countries from abroad. This indicates that commerce with external countries still play an important role in determining the growth of Brazilian Amazon's economy.

Moreover, the analysis of the other spatial variables results included in the regressions bring an additional aspect of complexity to this growing relative self-sustainable endogeneity aspect of Brazilian Amazon's growth and development: as we can see in table 2.6 (employment growth regressions), transportation costs to São Paulo's coefficient are positive and significant, whereas the coefficient of transportation costs to the nearest capital (within Amazon) are significantly negative. The exact opposite result is observed for the HDI and per capita GDP regressions. This may be interpreted as an evidence that the economic development measured by employment growth tend to be more self-sufficiently endogenous (or less dependent on external markets' economic linkages) than when it is measured by per capita GDP or HDI growth, as employment growth depends positively on local lower costs of transportation (nearest capital), and negatively on costs of selling (and buying) goods to markets outside the Amazon region (São Paulo), whereas local per capita GDP and HDI-M growth tend to depend positively on the external market of São Paulo (or negatively on transportation costs to this destination), and negatively on the main local markets (or positively on transportation costs to these destinations). This result is somewhat expected due to the conception of these three dependent variables: in general employment tend to be more spatially restricted, as it involves the displacement of families (i.e. all migration costs and other expenses concerning individuals' mobility), whereas per capita GDP and HDI are

[17] Note that for this variable, we do not use the term "causality", because as well as in the argument presented for the "*bolsa família*" variable, the exports per capita variable refers to the years of 2004 to 2007, which means that endogeneity through reverse causality might be an issue concerning this specific variable. Put in other terms, it is possible that some of the Amazon municipalities simply exported more goods between 2004 and 2007 due to an upward shift in their supply curve of tradable goods caused by a higher GDP or employment growth in the beginning of the 2000 decade. This, in turn, would means that growth would be the cause of higher exportations, and not the expected opposite, which by its turn may cause bias in estimations. As in the case of "*bolsa família*", in order to test for this possible endogeneity bias, we have estimated the regressions excluding exports per capita from the regressors' vector. Once again, fortunately, coefficients of the other explanatory variables did not change in terms of significance and sign, which brings robustness to our results, in spite of this possible econometric issue.

usually less spatially restricted, in terms of mobility, as they are composed of economic activities which includes industries with no transportation costs at all, such as online services, for example. Thus, it seems natural that employment growth would be encompassed firstly by a process of endogeneisation of the local economy towards a lower dependency on external economies, whereas per capita GDP and HDI-M will probably be more resilient in this sense, only adhering to this process in latter periods.

Finally, our last econometric results concern the accumulated deforestation level of each municipality and its' neighbors in the year 2000, as explanatory variables. In this regard, our results bring an ecologically pessimistic message: although these variables do not exhibit significant coefficients in the employment growth regressions, results from the other endogenous variables' regressions show that, on average, higher levels of deforestation in 2000 are associated with greater HDI-M growth rates within the same municipality, as well as cities whose neighbors showed a higher level of deforestation in 2000 tended to exhibit lower per capita GDP growth rates between 2000 and 2010. Once again, this result may be considered expected in theory, if we consider that land productivity tend to be higher in places where land clearing have occurred in earlier periods. As for the result concerning neighbors' deforestation, a similar theoretical argument may be presented: a city whose neighbors' deforestation level is higher than the level of deforestation of the municipality itself is expected to grow less than its neighbors' average, since this city does not pursue the same land-use competitive advantage (discussed above) as do its neighbors.

2.5. Conclusions and Final Remarks

In this chapter we have tried to investigate how the increasing urbanization process which has been occurring in the Brazilian Amazon recently has been influencing the region's economic growth and development. Specifically, based on economic theory from Urban Economics models, New Economic Geography and Urban System models, which defend that urbanization is deeply and positively connected with development and growth, we have purposed an exploratory and econometric approach by which we have tried to find causality links between urbanization and growth in Brazilian Amazon, throughout the last decade. Furthermore, given this scenario, we have also tried to investigate how much of the region's recent economic development may be attributed to internal drivers, such as education, urbanization and internal markets linkages, and how much still may be attributed to external dependency from Brazilian southern markets (such as São Paulo) and commerce with other countries.

More specifically, we have proposed highlighted a few indicators that show how fast this process is growing in Amazon, related this growth process with urban economics theory, provided preliminary statistical results, and investigate econometrically how close employment and GDP per capita growth are correlated to the growth of urban centers within the region.

In this sense, we have firstly described and evidenced the increasing force of the urbanization process happening in Brazilian Amazon recently: a simple descriptive analysis of the Brazilian Census data showed that urban population grew from 42% to 71% in the last four decades; Private services sector, typically related to urban centers, grew its participation on the region's GDP composition from 30% to 35% in the last 20 years; Urban population grew 29% in this same period; People living within metropolitan areas with millions of inhabitants has already reached about 30% of total Amazon population; some of its medium-sized municipalities, such as São Félix do Xingú, have experimented population booms whose growth rate surpassed 200% in 10 years.

Given that scenario, our exploratory analysis' results suggest that, as the economic theoretical literature revised here suggest, such urbanization process indeed seem to be positively correlated economic growth and development in the Brazilian Amazon: the most populated urban agglomerations present HDI-M levels far above the region's average; the top developed cities within the region tend to encompass more than 90% of their population living inside urban areas; all the three urbanization measures used in this study (the share of urban population, the share of employment in the terciary sector and the metropolitan regions' dummies) exhibit a significant and positive correlation coefficient with employment, per capita GDP and HDI growth from 2000 to 2010.

Furthermore, our spatial econometric analysis strongly suggests that recent economic growth and development in Brazilian Amazon is being caused, among other factors, by such urbanization process. Specifically, the evidence found in our regressions point out that all the three endogenous variables which represent economic growth and development from 2000 to 2010 (employment growth, per capita GDP growth and HDI-M growth) are positively (and significantly) explained by the initial levels urbanization of the Brazilian Amazon municipalities in 2000. Or, put in other words, our results show significantly that the higher the initial urbanization level of Brazilian Amazon municipalities in 2000, the greater tend to be the economic growth and development in these cities. Moreover, these same econometric results also show that even though the region's educational level is still incipient when compared to other areas from Brazil, higher levels of education are also correlated with higher growth and development within the region's municipalities. In addition, internal neighbor effects also contribute positively to local economic growth, according to our findings.

Finally, these positive connections between urbanization, education and internal neighbor linkages with economic growth and development found in our analysis suggest that the Brazilian Amazon region seems to be moving forward towards a path of a higher endogenous relative self-sufficiency, in which local economies seem to be reducing their dependency on external markets to grow and develop. However, our econometric results also show that such dependency still exists and plays a significant role in Brazilian Amazon's regional growth pattern: transportation costs to the State's to São Paulo, alongside with per capita exports to other countries, are still important explanatory factors of local growth and development, according to our results. Therefore, we may conclude that the Brazilian Amazon's economy seem to be increasingly self-sustainable on the one hand, but such internal endogeneisation process of local growth and development still may be considered at early or intermediate stages, on the other hand.

Nevertheless, as the urbanization rates in Brazil as a whole are still higher than the average Brazilian Amazon municipalities' rates, it seems likely that such urbanization process will keep up its pace for a few more decades within the region. Therefore, as this process seem to be accompanied by economic growth and development (according to our results), we conclude that the long term perspective concerning the region's economic development is positive, even though, as discussed in the previous chapters of this thesis, it will probably be accompanied by a rise in local deforestation, exactly because of such local urban sprawl.

3. LOCAL AND GLOBAL DRIVERS OF DEFORESTATION IN BRAZILIAN AMAZON

3.1. Introduction

The preservation of the Brazilian Amazon rainforest has certainly been one of the most important topics discussed at the global environmental agenda throughout the last decade. Several aspects have led to this given importance. First of all, recent literature found evidence supporting that the Amazonian deforestation process is highly correlated to global warming and climate change. Imori et al (2011), for example, use an Input-Output model, based on national data split into regional scale, to evidence that deforestation in Brazilian Amazon is responsible for about 58% of total greenhouse gases emissions in Brazil, and about 2% of total global emissions, in the year of 2004. Secondly, as the Amazon rainforest is the largest remaining tropical forest in the globe, it holds an immeasurable biodiversity whose conservation is undoubtedly fundamental to global ecological equilibrium.

Economically, the region hosts the current agricultural and cattle raising frontiers in Brazil, which is based on strong capital-intensive cultivation of soybeans, and whose exports have been boosting surpluses in the Brazilian trade balance recently (Morton et al. 2006; Vera-Diaz et al. 2009). Politically, it is considered an area of extreme strategic importance, due to its high availability of exploitable natural resources, especially in terms of the profitability potential for its extractive and mining industries. An example of such political importance is the recent plebiscite which occurred in the state of Pará, concerning a proposal of subdivision of this state into two parts, which would result in the creation of a new state called "Carajás". This has turned into a big debate in Brazil, as it was seen by many as a political maneuver coming from large mining companies installed on site, seeking higher tax liens and other political advantages.

Given such economical and ecological complexity, recent literature has been sharing a great deal of effort to evidence which are the main drivers of deforestation of the Amazon rainforest, and how to avoid it without compromising the region's economic development. In this regard, the majority of these studies point out to two main drivers of Amazonian deforestation, recently: the expansion of the grains agriculture, as well as the increase in land use by pasture for livestock (see, e.g., Morton et al, 2006; Vera-Diaz et al, 2009). Project Catalyst (2008) points out that the main drivers of deforestation in South America are pasture expansion for cattle raising (65%), subsistence agriculture (31%), forestry (3%) and intensive agriculture (1%). Chomitz & Thomas (2003), in turn, use a different approach, and find evidence that natural weather conditions also contribute to deforestation. Specifically, they find that rainfall regimes tend to determine land use in the Amazon, concluding that drier areas tend to be more rapidly deforested, due to the ease of fire use in order to clear the land.

Most of the times, studies such as the ones cited above share a common feature: in economic terms, they mainly focus on the "supply side" drivers of deforestation. To put it differently, these approaches systematically tend to base their conclusions on analyzing variables such as the direct land use transformation from forest regions into agricultural cultures or pastures; which techniques are employed in forest management practices; how local producers react to environmental policies and tax incentives for land occupation; or how deforestation responds to changes on local infrastructure (see, e.g., Walker et al. 2000; Binswanger 1991, Igliori et al. 2009b).

Despite bringing very important contributions, these studies alone should be considered incomplete by economic literature, since they fail to consider the effects from the "demand side" of economy. More specifically, literature still needs to take into consideration the impacts of local development, regional demand growth, and other demand driven aspects of both Brazilian and Amazonian recent contexts, in order to properly deal with the relationship between economic development and deforestation within the region. For example, one question still poorly addressed by literature is: how much each regional demand vector from different parts of Brazil can be held responsible for the expansion of production among industries within Brazilian Amazon which results in deforestation practices, such as cattle raising? From a very simplistic point of view, it seems reasonable that acknowledging where the output produced in previously forest covered areas is being consumed at, and by whom, may be as important as getting to know which are the industries responsible for this output production and the consequent deforestation.

These "demand-side" drivers become even more important if we take into account the increasing urbanization process and population growth that the region has been currently experiencing, as described in the previous chapter. From a theoretical perspective, many economic theories highlight the importance of the role played by these "demand side" effects in terms of land use practice and decisions, i.e., over one region's deforestation. Models from Urban Economics, New Economic Geography and Spatial Economics (see (Anselin (1988);

Von Thünen (1826); Hotelling (1931); Gleaser (2008); Krugman (1991); Fujita & Thisse (2002); Igliori, (2009)) point out clearly that when it comes to the matter of analyzing land use dynamics, location decisions always take into account two fundamental determinants: the size of relevant markets, in terms of number of consumers; and the distance between those markets and the productive units that supply them. Walker & Homma (1996), for example, when analyzing the contribution of these models in order to explain the dynamics of land cover by different uses, clearly conclude that this dynamics in one region is closely linked to transportation costs for disposing output and buying inputs, as well as to the development of local markets vis-à-vis exportations towards more distant ones.

Environmentally, these models may be interpreted as suggesting that larger markets should impose greater impacts over deforestation on a given region by the replacement of forest coverage by land use for productive activities designed to attend such demand. Moreover, this very models also recognize that markets located closer to the forest, even if smaller, should also impose a significant deforestation impacts, due to the lower transportation costs of selling these goods at local level instead of exporting it to farther areas (Igliori 2009). In a simple perspective, this implies that it might be possible that smaller but closer markets exert a deforestation pressure as big as, or even higher than larger but farther markets.

Urban Economics models also argue that not only distance and transportation costs play an important role on determining these demand-side effects over land use, but urbanization rates taking place at the different markets matter as well. Their main argument is that an urbanization process brings with it an increase in the population's consumption levels, not only because of the greater proximity to markets and reduced transportation costs for inputs and outputs, but mainly due to positive shifts on society's income level associated with higher productivity caused by increasing returns to scale and economies of agglomeration of various types (see, e.g., Fujita et al. 1999; Fujita & Thisse 2002; Gleaser 2008). Furthermore, urbanization by itself already stands for a major expansion of the construction sector, which increases consumption of steel and other materials, further elevating natural resources consumption to serve as inputs for this industry.

Literature on the recent concept of Ecological Footprint agrees with these messages brought by Spatial Economics analysis. Rees & Wackernagel (1996) use this specific methodology to conclude that more urbanized areas of the globe contribute significantly more to exhaust global natural resources than rural areas, due to the more increased consumption patterns that these urbanized regions present when compared to rural undeveloped regions. A good example of these demand-side impacts over land use associated with urbanization processes regards the recent urbanization boom in China. The effects of the substantial urbanization process caused by the expansion of Chinese cities is held responsible, by economists, for a great share of the production boost of several industries around the world, and consequently for the generation of enormous environmental impacts associated with such expansion, especially in terms of greenhouse gases emissions and consumption of natural resources as inputs for these industries.

With regard to mass media, although such effects of urbanization are also still mostly being overlooked, a recent “isolated” article in the New York Times (December 2012) presented one rare exception to this rule. This article, called "Swallowing Rain Forest, Amazon Cities Surge In" points out and briefly describes the process of urbanization and population growth that Amazon has been facing in recent decades, trying to comprehend the possible consequences in terms of local deforestation. Although this is not an academic study, the report is clearly concerned about drawing their conclusions based on opinions of the scientific community, illustrated by a statement by Phillip M. Fearnside, a renowned researcher of the Amazon region, in which he declares that "More population leads to more deforestation".

On the other hand, empirical academic literature about Brazilian Amazon deforestation is still very incipient when it comes to this matter of considering the “demand-side” impacts of markets from different areas of the country over forest conservation. Still, there seem to be a general belief, both in literature and common knowledge, that Amazonian local markets are practically unimportant as drivers of deforestation and land use change within the region, mainly because they represent only a very small share of the Brazilian population when compared to the southern regions of the country. In this sense, livestock and agricultural output produced in the Amazon region are assumed to be sold mainly to the larger markets in southern Brazil, or even abroad, even though the accurate calculations were never made in this respect. Therefore, this is a “general belief” which lacks empirical support.

There are, however, some important exceptions to this rule, which at least represent starting points in this sense. For instance, Faminow (1997) points out that deforestation in Brazilian Amazon, at that time, seemed to be closely related to the large inner regional demographic changes which we described so far. More specifically, his article firstly recognizes that in recent decades, urban population has been growing dramatically in the Brazilian Amazon (as we have seen, twice as fast as the rest of the country), which in turn results in an expressive expansion of the “local" demand vector for agricultural and cattle beef outputs coming from local producers. In this sense, the author shows evidence that this increase in local demand for beef has been accompanied by the expansion of pastures for cattle raising within the region, or, put in other words, local demand for cattle is increasingly being supplied by local producers. This local effect, argues the author, is the result of an economic structure based on high transportation costs of beef imports produced in the rest of the country, which turned cattle raising inside the Brazilian Amazon into a relatively profitable activity. Finally, as cattle raising is considered one of the main drivers of deforestation in terms of land use, the author concludes that this local demand effect causes a severe impact in terms of deforestation, but surprisingly, for some reason, it is still being neglected by literature.

Whereas for the Brazilian Amazon region the relationship between urbanization, population growth and deforestation remains mostly overlooked by literature, studies regarding such relationship are not as scarce in a global scale. In this sense, Geist and Lambin (2002), reviewed extensively literature on this matter, and found that among 153 studies about deforestation drivers throughout the world, 93 pointed out local demographic changes as one of the most important underlying driving forces of deforestation, being 58 related to in-migration towards the forest region, and 38 related to growth in population density (which may be considered as a measure of urbanization). Moreover, the author shows that another 58 studies pointed to urban growth as one of the main underlying drivers of deforestation as well. In a more recent study, using an econometric approach with data from 43 countries spread throughout the world, DeFries et al (2010) find robust evidence on the role of urban population growth as one of the main drivers of deforestation either.

A few other studies on this matter also shed light over this matter, sometimes even studying the Amazon case, but often treating it as a secondary result among many others from their main analysis structure. Andersen et al. (2002) is one of these exceptions. Using econometric methods and Census data, the authors conclude that the local urban population and GDP growth in the Amazon region are increasingly important when explaining recent patterns of land use and deforestation. Using similar methodology and database, Igliori (2009) finds evidence that higher levels of local agglomeration (population and economic activity) contribute positively to increase deforestation and economic growth.

Even though these referred studies bring important contributions and are pioneers in the sense of taking these local demand and urbanization effects into consideration, literature lacks a more detailed and complete approach of the subject. Faminow (1997), for example, studies only the direct impacts of a single industry (cattle raising) over local economic variables, without making any direct links to deforestation or any other environmental measures. Also, the author has not considered any multiplier effects with respect to other regions or other industries which this positive shift in local demand definitely may bring. Moreover, no studies try to compare the impacts of deforestation caused by local markets vis a vis the ones caused by foreign markets, in order to provide feedback to future policies of land occupation and urbanization of the region. And given these evidence on local demographic changes in within the region in the last couple of decades, it seems appropriate that literature stops overlooking these “demand side” effects on deforestation as soon as possible.

Given that picture of the problem, this chapter attempts to make a contribution to fill this gap in literature of Brazilian Amazon deforestation studies. Therefore, the main goal here is to try to measure the impacts, in terms of deforestation (but also concerning economic aspects), of the increasing local demand drivers associated to this urbanization process and the recent local population growth going through the region recently. More specifically, we use an Input-Output inter-regional model for Brazil, along with land use transition data, to measure the impacts driven by the demand vector of the 5 big Metropolitan Regions within the Brazilian Amazon, comparing these results to those driven by the demand vector from the Rest of Brazilian Amazon, which hosts a less urbanized share of the region, and also to those driven by the demand vectors from the rest of Brazil and the rest of the world (through exportations).

Apart from this introduction, this chapter is structured as follows: The next section provides details of the methodology adopted to achieve our goals, justifying this choice. Section 3.2 provides and a description of the database used, and discusses the main strategy adopted empirically. Section 3.4 supplies socioeconomic information and historical background of the Brazilian Amazon. Section 3.5 presents and discusses the results, and section 3.6 concludes.

3.2. Methodology and Theoretical Background

Prior to explaining the basic methodology of the Input-Output models used in this article, we justify the choice of such methodology. The main goal of this article is to measure the size of the impacts exerted by the local vectors of demand (associated to the urbanization process) over deforestation of the Amazon rainforest, comparing those to the impacts driven by the external demand vectors from the rest of Brazil and exports. In order to do so, firstly, we need a methodology that allows us to isolate the demand vectors from within the Amazon region, also isolating the vectors from its most densely urbanized areas from the ones of the rest of Brazil and other countries. Second, in order to accurately measure such impacts, we need to take into account not only the "direct" deforestation caused by production of the output designated to attend the final demand vectors, but also the "indirect" deforestation driven by the production of inputs used in the whole production process of such outputs, in each sector and in each region of Brazil.

The Inter-regional Input-Output Model encompasses both of these features: Firstly, it allows us to implement such regional division, isolating Amazon from the rest of Brazil and also isolating the Amazon Metropolitan regions (which, as will be justified ahead, will be our representative of the urbanization process) from the rest of the Brazilian Amazon. Secondly, it also allows us to calculate and isolate direct, indirect and induced impacts[18] on each sector of the economy caused by each demand vector from the chosen different regions.

In order to make this point clearer, we provide an example of result which may be achieved by such methodology: the Inter-regional Input-Output Model implemented here is able to measure the amount of deforestation in Amazon caused by the consumption of clothing accessories by households living within the Manaus Metropolitan Area, even if these goods are produced partially in São Paulo, and partially in Manaus itself. This is so because the inter-regional Input-Output matrices account for, among other things, the livestock production required in the Amazon region in order to provide inputs to be consumed by the footwear industry in São Paulo and Manaus. In this sense, the model approach is very similar to footprint calculation models (see Rees & Walbirgne, 1996), but it is also more complete and precise economically.

3.2.1. The Single-Region Basic Leontief Model

The theoretical basis adopted in this work is, as already mentioned, the Input-Output model. More specifically, we implement an Inter-Regional version of such model. However, for a matter of introducing the basic concepts of this methodology, we firstly introduce as a simple version of the model with a single region, and therein expand it to a multiple region approach.

[18] We define these 3 kind of impacts in the 2.2.1 subsection of this chapter.

The Leontief Basic Model with a single region is based on a system of multiple equations that represent the flows of goods and services between sectors and agents in the economy, with such flows explained by technological and economic factors (Miller & Blair, 2009). This system can be represented, in matrix notation, by equation 3.1:

$$X = AX + Y \tag{3.1}$$

where X is a (*nx1*) vector representing the total output produced by each of the n sectors of the economy, Y is a (*nx1*) vector representing the final demands of each *n* sector from families, government and exports, and A is a (*nxn*) matrix that contains the technical coefficients of production. These coefficients represent, for each sector *j*, $\forall j \in [1, n]$, the proportion of inputs that *j* must buy from each sector *i*, with $\forall i \in [1, n]$ with $i \neq j$, in order to produce one additional unit of its output *j*. At this point, the first important assumption of the Input-Output methodology is made: these coefficients are assumed to be fixed, independently of the amount of output which needs to be produced, and also regardless of to which branch of the supply chain the output is being produced for. Or in other words, it is being assumed that every sector of the economy exhibit constant returns to scale. In the Results Section we discuss what kind of implications this first assumption brings to our specific results.

Manipulating (1), we get equation 3.2:

$$(I - A)X = Y \tag{3.2}$$

or

$$X = (I - A)^{-1}Y = BY \tag{3.3}$$

where $(I - A)^{-1} = B$, and *B* is called the Leontief inverse matrix, or the matrix of direct and indirect coefficients, in which each element b_{ij} represents the total production which is needed to be produced in sector *i* in order to meet an additional unit of final demand of sector j, considering the total output needed to attend the final demand vector and also the inputs that needed to be produced to meet the intermediate consumption along the production chain.

From the Leontief inverse matrix *B*, the multipliers of type I (see Miller & Blair, 2009) for each sector can be calculated, and these are given by equation 3.4:

$$P_j = \sum_{i=1}^{n} b_{ij} \tag{3.4}$$

where P_j represents the total output generated in all sectors, as a result of an additional unit of sector *j*´s final demand.

Also from matrix *B*, we can calculate the employment, the value added and the deforestation generators for each sector i of the economy: calling V_i the total amount of deforestation produced by sector i in a given year, in order to calculate the deforestation generator for sector j, we firstly need to calculate the deforestation coefficient of each sector i, which is given by equation 3.5:

$$v_i = \frac{V_i}{X_i} \tag{3.5}$$

v_i tells us how much deforestation results from the direct production of one monetary unit of output i produced. Given such coefficient, the Leontief Inverse Matrix is used to calculate how much deforestation is generated by both direct and indirect output generated in all sectors in order to attend and additional unit of final demand of sector j, that is, how much deforestation is caused by the direct production of the output j which attends the final demand, and also by the production of the inputs necessary to build this output j. This is exactly the definition of the deforestation generator, and its calculation is given by equation:

$$GV_j = \sum_{i=1}^{n} b_{ij} v_i \tag{3.6}$$

where GV_j is the deforestation generator of sector j.

One extension that can be applied in the calculation of the multipliers and generators, and which we also use here, is to incorporate the income that families receive during the production process of the output intended to attend the final demands into the Input-Output system. Put in another way, it is possible to incorporate what literature refers to the induced multipliers effects which are generated in the economic system due to the payments that firms make to families in exchange of their labor in the production chain in order to produce the output to attend final demands. In order to incorporate these effects, we apply the methodology described in Miller & Blair (2009), known as endogeneisation of families consumption and income. In the following sections, we will show both kinds of results for the multiplier and generators calculated in this work: the ones which encompass, and also the ones which do not take into account these induced effects, along with the traditional direct and indirect effects.

3.2.2. The Inter-regional Input-Output Model

Let us now expand the single region model, and consider an Inter-regional Input and Output model, which is the one used in this chapter. In order to simplify, we here describe a model with only two regions, L and M, but we emphasize that the extension for a model with *m* regions, with $m \geq 3$, follows the same methodology (see Miller & Blair, 2009).

Therefore, consider the matrices and vectors from the basic model of Leontief as in the previous subsection, but now partitioned to represent our two region model. As Guilhoto (2009) shows, we can write these matrices as in equations 3.7, 3.8 and 3.9:

$$A = \begin{bmatrix} A^{LL} & \vdots & A^{LM} \\ \cdots & \cdots & \cdots \\ A^{ML} & \vdots & A^{MM} \end{bmatrix} \tag{3.7}$$

$$Y = \begin{bmatrix} Y^{L} \\ \cdots \\ Y^{M} \end{bmatrix} \tag{3.8}$$

$$X = \begin{bmatrix} X^{L} \\ \cdots \\ X^{M} \end{bmatrix} \tag{3.9}$$

where A^{ML} represents an *nxn* matrix, with *n* being the number of sectors of the economy, and with each of its elements a_{ij}^{ML} representing the technical coefficient which measures how much industry *j* from region *L* buys from sector *i* in region *M*, $\forall i, j \in [1, n]$. A similar interpretation is given to A^{LL}, A^{MM}, A^{LM} and their respective a_{ij}^{LL}, a_{ij}^{MM}, a_{ij}^{LM} elements.

Similarly, each x_j^L represents the elements of X^L, measuring the total output produced by sector *j* in region *L*, while y_j^L represents the elements of Y^L, and measure the final demand vector of sector *j* in region *L*.

Thus, following the same reasoning of the basic Leontief model with a single region, Guilhoto (2009) also shows that we can write the system previously given by equation 3.2 as in equation 3.10:

$$\left\{\begin{bmatrix} I & \vdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \vdots & I \end{bmatrix} - \begin{bmatrix} A^{LL} & \vdots & A^{LM} \\ \cdots & \cdots & \cdots \\ A^{ML} & \vdots & A^{MM} \end{bmatrix}\right\} \begin{bmatrix} X^L \\ \cdots \\ X^M \end{bmatrix} = \begin{bmatrix} Y^L \\ \cdots \\ Y^M \end{bmatrix} \tag{3.10}$$

Separating the partitioned parts of 3.10, we obtain:

$$(I - A^{LL})X^L - A^{LM}X^M = Y^L \tag{3.11}$$

and

$$-A^{ML}X^L - (I - A^{MM})X^M = Y^M \tag{3.12}$$

Equations 3.11 and 3.12 describe the Inter-regional Input Output system of equations to be estimated. In matrix notation, Miller & Blair (2009) show that this system can be written in the form $X = BY$, in which *X* and *Y* may be rewritten as in 3.8 and 3.9, respectively, and *B* follows equation 3.13:

$$B = \begin{bmatrix} B^{LL} & B^{LM} \\ B^{ML} & B^{MM} \end{bmatrix} \tag{3.13}$$

where *B* is the Inverse Leontief Matrix from the Inter-regional Input Output system. Again, each B^{ML}, B^{LM}, B^{MM}, B^{LL} is a *nxn* matrix, with *n* being the number of sectors of the economy, and each element b_{ij}^{ML} of the matrix B^{ML} represents the amount of output needed to be produced in sector *i* in region *M* to attend one unit of final demand of sector *j* in region *L*, $\forall i, j \in [1, n]$. A similar interpretation is given to the elements of B^{LM}, B^{MM} and B^{LL}.

Given the Inverse Leontief Matrix of the Inter-Regional system, Miller & Blair (2009) and Guilhoto (2009) show that it is possible to calculate the multipliers and generators of employment, value added and deforestation generators similarly to the ones from the basic single region Leontief model already presented, which is exactly what we implement here.

3.3. Empirical Strategy and database

3.3.1. Regional division of the Inter-regional Input-Output system

In order to measure and analyze the deforestation (and some other economic) impacts which urbanization and local demand growth in Brazilian Amazon may be causing recently, we followed the empirical strategy of building an inter-regional Input-Output model in which we divided Brazil into 3 regions, as follows:

• Region 1: comprises the 5 Metropolitan Regions[19] of the Brazilian Amazon[20], which are the Cuiabá-Várzea Grande Urban Conglomerate, the Metropolitan Region of Manaus, Macapá Metropolitan Region, the Grande São Luís Metropolitan Region, and the Metropolitan Region of Belém. Often during this work, we refer to this region as Amazon RMAM.

• Region 2: The remaining Brazilian Amazon. Often during this work, we refer to this region as REAM.

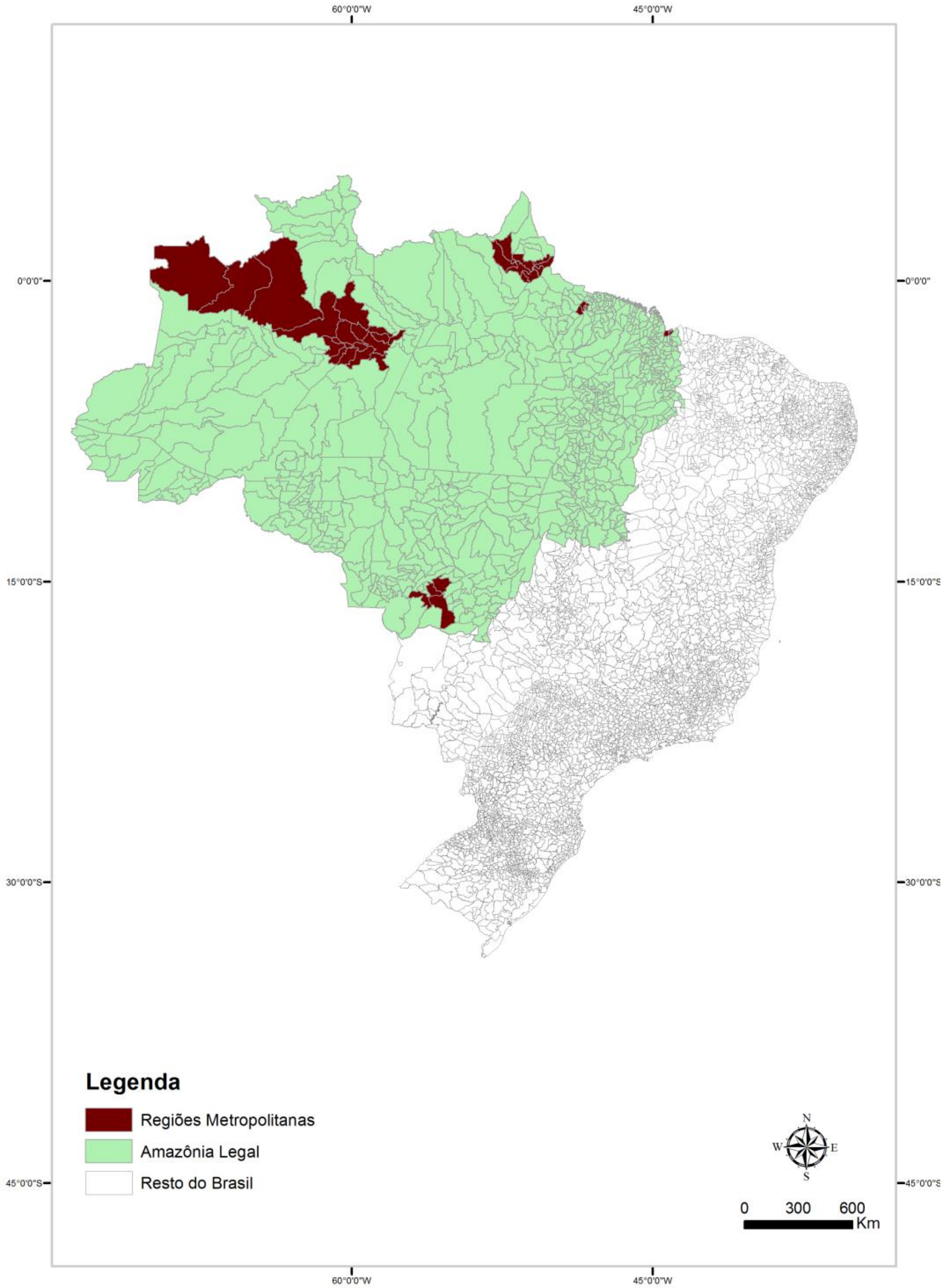
• Region 3: The Rest of Brazil. Often during this work, we refer to this region as RBR.

Figure 3.1 brings a map of these 3 regions. Note that they are not continuous geographically terms, since the Amazon metropolitan regions do share border between themselves, due to the region enormous extension.

[19] Due to data issues, we were not able to build Region 1 exactly with only the municipalities within each of the 5 Metropolitan Regions considered. A detailed discussion of this matter, along with the possible consequences to our results, can be found at Appendix A.1 of this thesis.

[20] Many researchers argue that these 5 Metropolitan regions may not be the only real Metropolitan Regions within Brazilian Amazon, as representatives of urban conglomerates. Due to that, in Appendix A.1 we briefly discuss the reasons and possible consequences of our choice of using use only these 5 regions.

Figure 3. 1: Metropolitan Regions of Brazilian Amazon, the remaining Brazilian Amazon, rest of Brazil



SOURCE: IBGE. Own Elaboration

The reason for adopting this criterion of division of the Brazilian territory between these three chosen regions is because we hope to measure two different types of demand impacts over local deforestation: a) the impacts that local markets exert, i.e., the impacts from the demand vectors from local consumers within the Brazilian Amazon borders; b) the impacts of the urbanization process occurring within the Brazilian Amazon region. In order to achieve that, we initially split the Brazilian Amazon from the rest of Brazil (Region 3), so we could isolate the vectors of demand for each of these two regions (Brazilian Amazon and Rest of Brazil). Then, in order to try isolate a representative of urbanization within the Brazilian Amazon region, we split it into two regions: Region 1, which encompasses the Amazonian Metropolitan Regions, and Region 2, which encompasses the rest of the Brazilian Amazon.

Separating Regions 1 and 2 in order to isolate the effects of urbanization effects was motivated by the fact that metropolitan areas represent the most urbanized areas of Brazilian Amazon, in addition to being the largest urban conglomerations within the area. Thus, is somewhat subjectively defined variable, these might be considered good representatives of the urbanization process. This statement can be easily checked at IBGE 2010 Census, by which we calculated that 90.4% of the population of Region 1 lived within the urban areas of their respective municipalities in 2010, while in the Rest of the Amazon (Region 2), this percentage was only around 60%. In fact, in 2010, more than 6 million out of the total of 24.4 million inhabitants of the Brazilian Amazon lived within its metropolitan areas (which represents about 25% of the region’s total population). Therefore, Region 1 seems to represent well the most densely urbanized centers of Brazilian Amazon, which means that isolating it from the rest of the Amazon can be considered a good strategy to try to capture the effects of urbanization in the Amazon region separately from the effects of local demand vectors over deforestation. Contributes positively to this choice the fact that in the rest of Brazil, where urbanization process is already more consolidated, history showed that the typical dynamic of the urbanization processes which other regions have passed through tend to present a tendency of reaching its to peak with the emergence and consolidation of giant and highly dense urban conglomerates and metropolitan regions, whose growth tend to walk side by side with the urbanization process in general.

Ideally, the urbanization effects would certainly be more precisely captured if we could split the local demand vector from the Amazon region as a whole into two: the demand vector of the Brazilian Amazon urban population, and the demand vector of the Brazilian Amazon rural population. Unfortunately, however, data disaggregation level does not allow us to implement such division.

Given the strategic division which we have chosen, we can rewrite equation 3.13 to represent the Leontief inverse Matrix of the Inter-Regional Input-Output model we use in this article, as equation 3.14:

$$B_3 = \begin{bmatrix} B^{RMAMxRMAM} & B^{RMAMxREAM} & B^{RMAMxRBR} \\ B^{REAMxRMAM} & B^{REAMxREAM} & B^{REAMxRBR} \\ B^{RBRxRMAM} & \cdot B^{RBRxREAM} & \cdot B^{RBRxRBR} \end{bmatrix} \quad (3.14)$$

where B_3 is the Inverse Leontief Matrix from the Inter-regional Input Output system with the 3 regions described above, and each element b_{ij}^{MxL} from B^{MxL}, represents the amount of output needed to be produced in sector *i* of region M to attend one additional unit of final demand of sector *j* in region *L*, $\forall i, j \in [1, n]$, with being *n* the number of sectors of the brazilian economy, and with being M and L each representing one of the three regions described in this section (RMAM; REAM; RBR).

3.3.2. Sectoral Division and Aggregation

Database sources used to construct the Input-Output Inter-regional system were obtained using data from the IBGE's Regional and National Accounts, for the year of 2004. Specifically, using this data, we have applied the methodology developed by Guilhoto and Sesso Filho (2005) in order to construct the Brazilian Input-Output Matrices at national level, at basic prices and with 56 industries and 110 commodities. Then, we have applied the methodology described in Guilhoto et al. (2010) in order to build an inter-regional Input-Output system for Brazil, with 56 industries and 558 (micro)regions of Brazil, always for the year of 2004. These (micro)regions were then aggregated into the 3 Regions described above, in compliance with all considerations we made previously. By their turn, the 56 industries were aggregated into 32 sectors.

The criterion adopted for such sectoral aggregation was the attempt to isolate sectors which present stronger relationships with deforestation, whereas grouping industries which produce similar goods and services, and are less directly correlated with deforestation. More specifically, firstly we isolated the two sectors that cause direct deforestation through land use competition, which are the Agriculture and Livestock industries. Secondly, we tried to isolate the sectors which are the main consumers of inputs (according to our initial 56 sectors Input-Output system) produced by these two, which means that we have tried to isolate the industries which tend to cause the greatest "indirect" deforestation. Finally, we group together into a smaller number of industries those sectors whose output is similar in terms of the characteristics of the good and services produced, and also in terms of their intermediate consumption of inputs produced by Agriculture and Livestock. As examples, we kept Agriculture, Forestry, and Livestock isolated as in the original 56 sectors system, whereas the various industries regarding specific personal services have been aggregated into one greater industry which we called the "Services" sector. Appendix A.2 brings a table in which we present the complete map of sectoral aggregation implemented here. Furthermore, in order to accomplish such aggregations correctly, we have followed the methodology described in Miller & Blair (2009).

The reason for incurring such aggregation is to ease the visualization of our results. This is because on the one hand, Inter-regional Input-Output models have the very desirable feature of capturing all multiplier effects between sectors and regions generated in all sectors along the production chain, but on the other hand, this same completeness of sectors and regions end in many results, making them difficult to be read and interpreted.

Yet, we observe that the aggregation criteria adopted do not bring large changes in our results, because in terms of deforestation, when we aggregate sectors different from Agriculture or Livestock, we are simply aggregating the intermediate consumption that those sectors demand from the first two, thus, in general, we do not underestimate the multiplier effects across sectors directly related to deforestation itself. This argument becomes clearer in the next sections, where we detail a description of the methodology used to measure the impact of each region's demand vector, in terms of deforestation.

3.3.3. Deforestation data

In order to calculate the deforestation impacts of each demand vector through an Input-Output system, we had to measure the amount of previously forest covered areas which became replaced by land use for Livestock and Agriculture production in the year of 2004 (which is the year to which our Input-Output data refers to). This was necessary because the Input-Output modeling allow us to measure how much all sectors and final demand vectors from each region in the economy consumed from Livestock and Agriculture in Brazilian Amazon.

Therefore, by measuring how much forest covered area was turned into pastures or agricultural land, we can also estimate how much of this land use change (deforestation) was driven by the consumption from families of each region, due to their direct consumption of Amazonian agricultural and livestock output, and also due to their consumption of every goods or service produced in the economy, which by their turn needed inputs from agriculture and livestock in the Brazilian Amazon.

Thus, we needed data on land conversion from forest covered areas which have turned into pastures or agricultural land within the Amazon region, on average, for the year as close as possible to 2004. Such data was obtained on the Second Brazilian Inventory of Emissions and Anthropogenic Removals of Greenhouse Gases, from where we have extracted Table 3.1, which sums up data on land transition from 1994 to 2002 found in Brazilian Amazon.

Table 3. 1: Deforestation and Land Transition in Brazilian Amazon (1994-2002)

		Land Use in 2002 (ha)					Total in 1994
		Forest Area	Reforestation	Pasture Area (for Livestock)	Agriculture and Forestry	Other Uses	
Land use in 1994 (ha)	Forest Area	345,400,858	27,264	15,294,488	2,275,242	45,847	**363,043,701**
	Reforestation	56	295,252	187	7,184	1	**302,680**
	Pasture Area (for Livestock)	772,591	12,296	25,791,281	987,198	67,368	**27,630,735**
	Agriculture and Forestry	73,057	753	1,332,935	3,083,190	5,626	**4,495,560**
	Other Uses	318	9,165	1,138,408	68,270	23,047,234	**24,263,398**
Total in 2002		**346,246,879**	**344,731**	**43,557,300**	**6,421,083**	**23,166,079**	**419,736,073**

SOURCE: Brazilian Ministry of Science and Technology. Own Elaboration.

Table 3.1 shows that between 1994 and 2002, 15,294,488 hectares of forest area had turned into pasture for livestock production in Brazilian Amazon, while 772,591 of pasture area for livestock had recovered back into forest land. Based on that, we considered that livestock production can be held responsible for 14,521,897 hectares of deforestation from 1994 to 2002 (which is the difference between 15,294,488 and 772,591) within the region. Making similar calculations, we estimate that in this very period, Agriculture and Forestry sector was responsible for 1,970,281 hectares of deforestation within the Brazilian Amazon region. Such calculations lead us to estimate and average of 1,613,544 and 218,920 hectares of deforestation per year, caused respectively by the Livestock, and the Agriculture and Forestry sectors.

As this is one of the only sources of land use transition data for Brazil, such specific data for the year of 2004 was not available, then in our calculations we have assumed that deforestation caused by these two sectors in 2004 had exactly matched these annual averages for each of the two sectors. We are aware that this may not be fully accurate, however, this was the only available method, at the moment of our calculations, in order to estimate our deforestation results.

3.4. Population, History, Socioeconomic and Environmental background.

Analyzing the impact of the local demand of Brazilian Amazon is a task that deserves some historical and geographic reflection beforehand. Most of the historical occupation of the Amazonian territory has occurred belatedly when compared to the rest of Brazil. As an example of it, Becker (2013) points out that in 1777 only three hundred people lived in Manaus, whereas in Rio de Janeiro population had already surpassed 35 thousand inhabitants. For centuries, regional occupation was based on transitory migration flows towards the region, almost all related to cyclical extractive activities, such as the Rubber Cycle[21], which poorly resulted in local population establishment, as these migration flows were significantly volatile, and tended to cease in periods of international market crisis of the product being extracted.

The real positive shift in Amazon occupation began around 1960, resulting from direct government intervention, when Brazilian military government created specific policies intended to promote the occupation and economic development of the region. These policies encouraged migrants to move towards Brazilian Amazon, through land concessions and federal infra-structure investments, in order to promote the settlement of the immigration flows. These land concession policies were based on the famous mote "land without men for men without land", which encouraged the departure of big miners to the region, granting broad slices of land to those. As for the great infra-structure investments, the most famous example was the construction of the enormous “transamazônica” road, which pursue more than 4 thousand kilometers of extension, and crosses the whole Brazilian Amazon territory horizontally, connecting the Brazilian Northeast macro-region with Peru and Ecuador.

In the beginning of the 1990 decade, however, the international economic crisis which affected heavily all Latin-American countries, the recent emerging environmental concerns about deforestation, and the end of Brazilian military dictatorship government have caused those explicit occupation policies to lose strength (Andersen et. al 2002). Still, even without

[21] The Rubber cycle is probably the most relevant example of these cyclical immigration waves towards Brazilian Amazon before 1960. It happened in the decade of 1870, and the migration flows were basically constituted of individuals coming from the northeast region of Brazil "pushed by the misery of the great droughts of the northeast" (Tom Amazon Project, 2011), and searching to work in the rubber extractive industry which emerged within the region after an international crisis of the rubber market.

those policies, the Brazilian Amazon population not only continued to expand, but urbanization and population growth have reached their peaks during the following period. As already mentioned, data from IBGE Census show that from 1970 to 2010, local urban population jumped from 42% to 71%. In the last decade, overall, the Brazilian Amazon population grew by 20%, whereas the rest of Brazil grew only by 10%. Ten out of the 19 cities that doubled their populations in the last ten years are located in the Brazilian Amazon, being Manaus, a city with 1.7 million inhabitants, one of them.

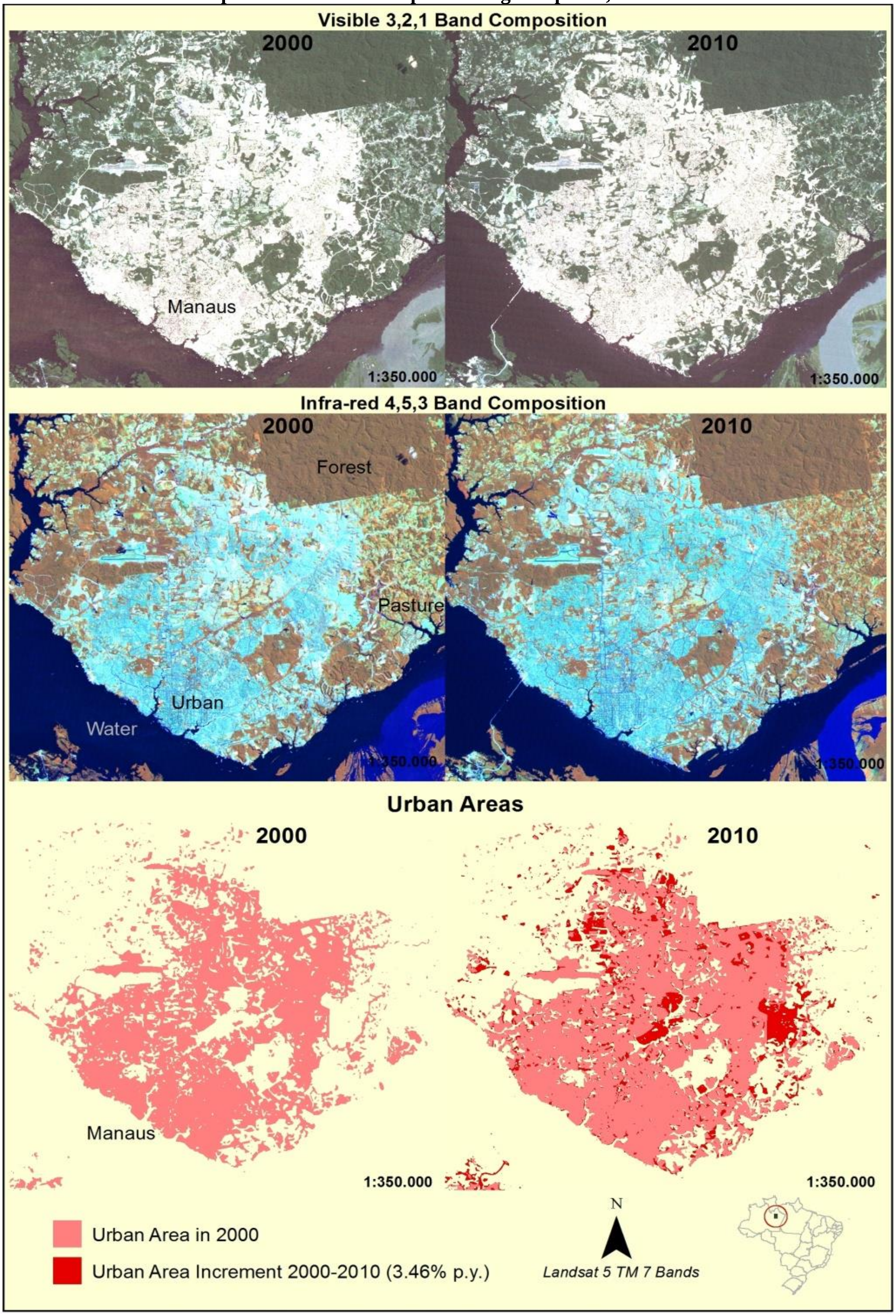
Specifically regarding the growth of Brazilian Amazon metropolitan regions in the last decade, in order to illustrate the rapid process of urban expansion within these, we have built maps of urban sprawl from 2000 to 2010 of Manaus, Belém and Cuiabá-Várzea Grande, which are the first, second and fourth major urban centers of Brazilian Amazon, respectively. For such, we have used data from satellite[22] images, and the results are shown in maps 3.1, 3.2 and 3.3.

As we can see from these 3 maps, urban sprawl in the last decade in these three metropolitan region can be considered extremely high: it presented a yearly growth rate of 3.46% 2.18% 3.89% in Manaus, Belém and Cuiabá-Várzea Grande respectively from 2000 to 2010, which are quite large rates for regions whose base population was over 500,000 inhabitants in 2000 (with Manaus and Belém surpassing 1 million inhabitants).

Such urbanization process, however is not taking exclusively in the region's metropolitan areas. In order to illustrate it, we have also built the map (Map 3.4) of urban sprawl for São Félix do Xingú, a medium size city within Amazon whose population was around 35,000 inhabitants in 2000, and jumped to more than 90,000 in 2010.

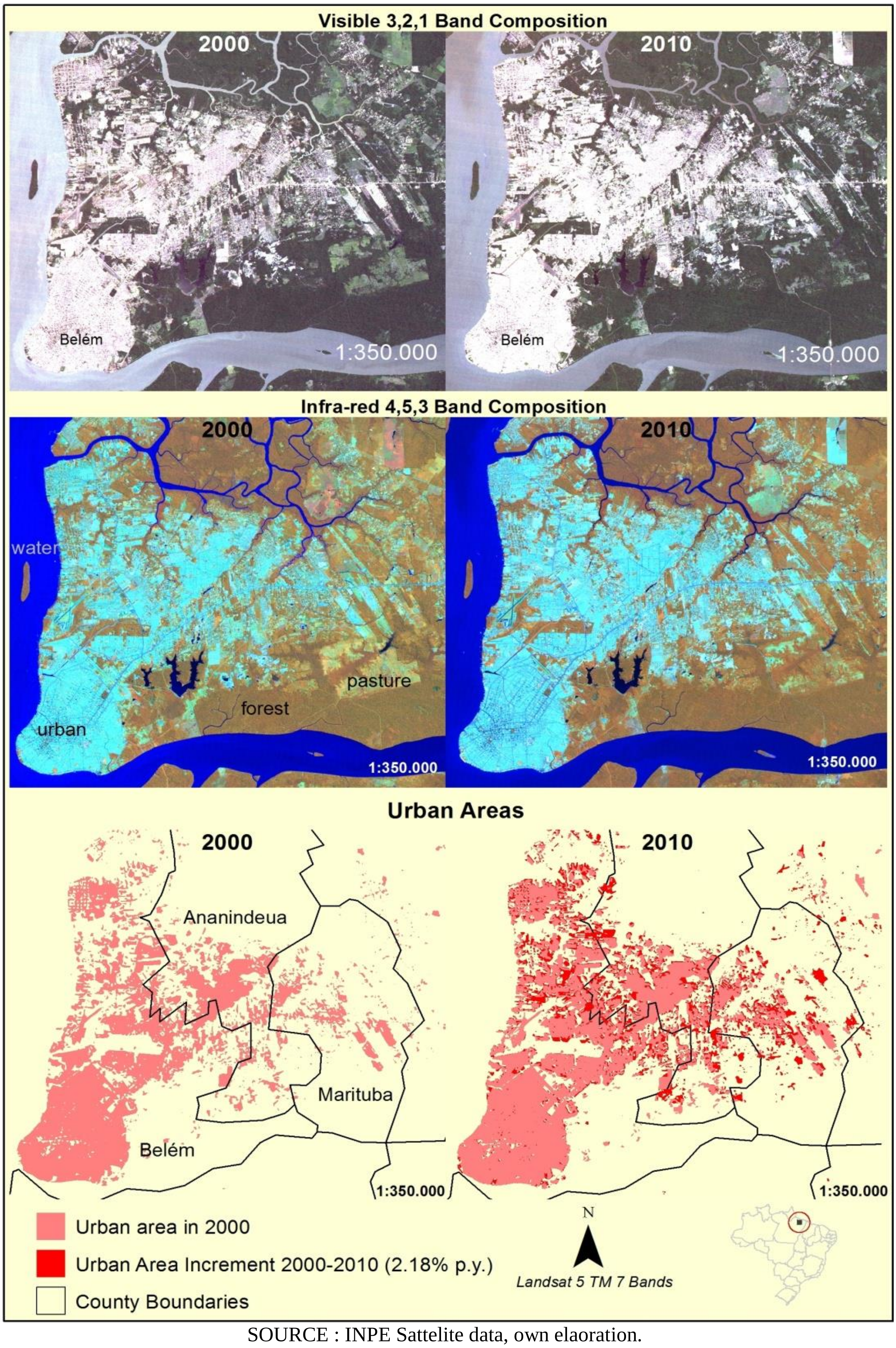
[22] Specifically, we have used sattelite (Landsat) images from INPE (National Institute for Spatial Research) database to build these maps.

Map 3. 1: Manaus Metropolitan Region Sprawl, 2000-2010



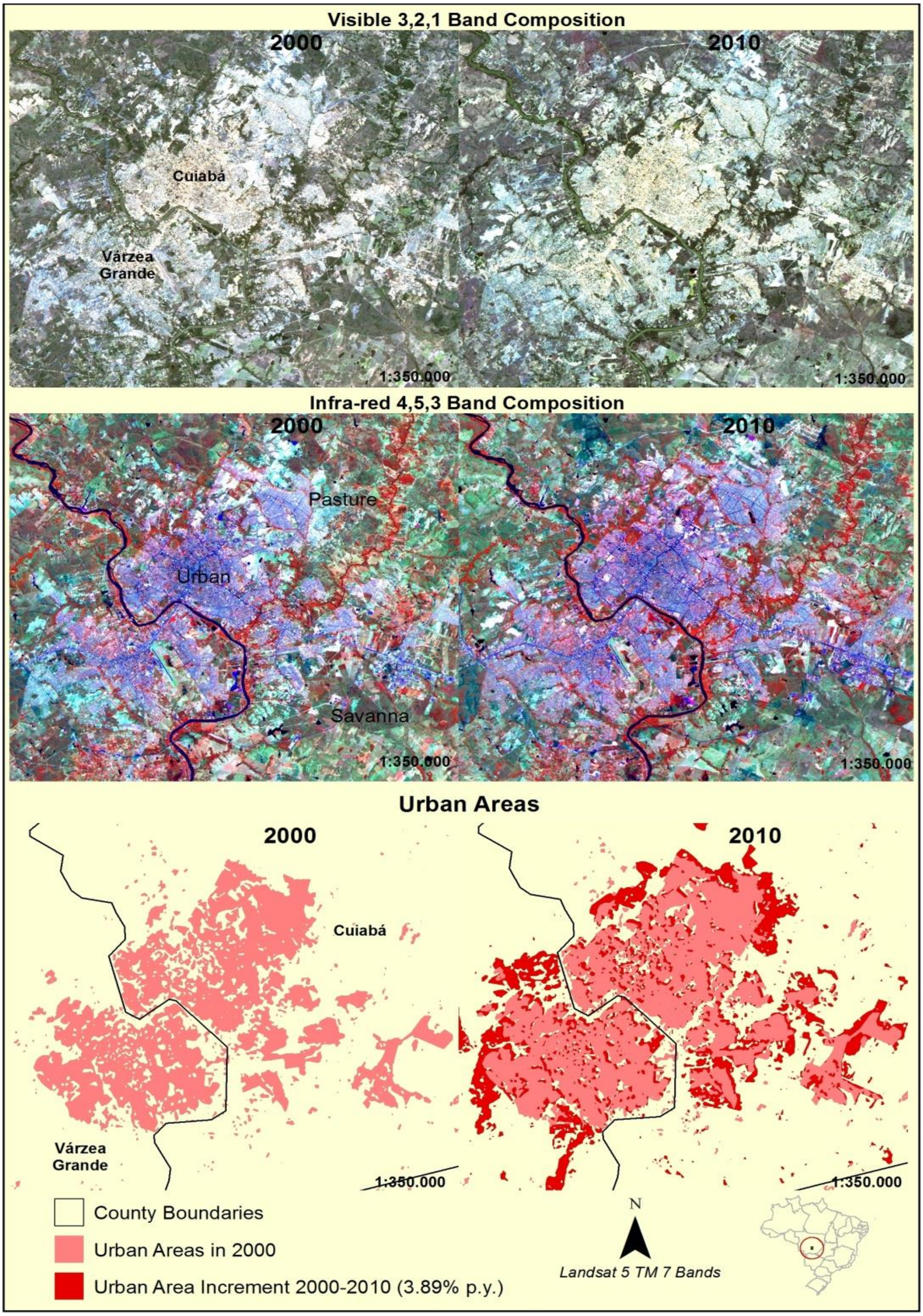
SOURCE : INPE Sattelite data, own elaoration.

Map 3. 2: Belém Metropolitan Region Sprawl, 2000-2010



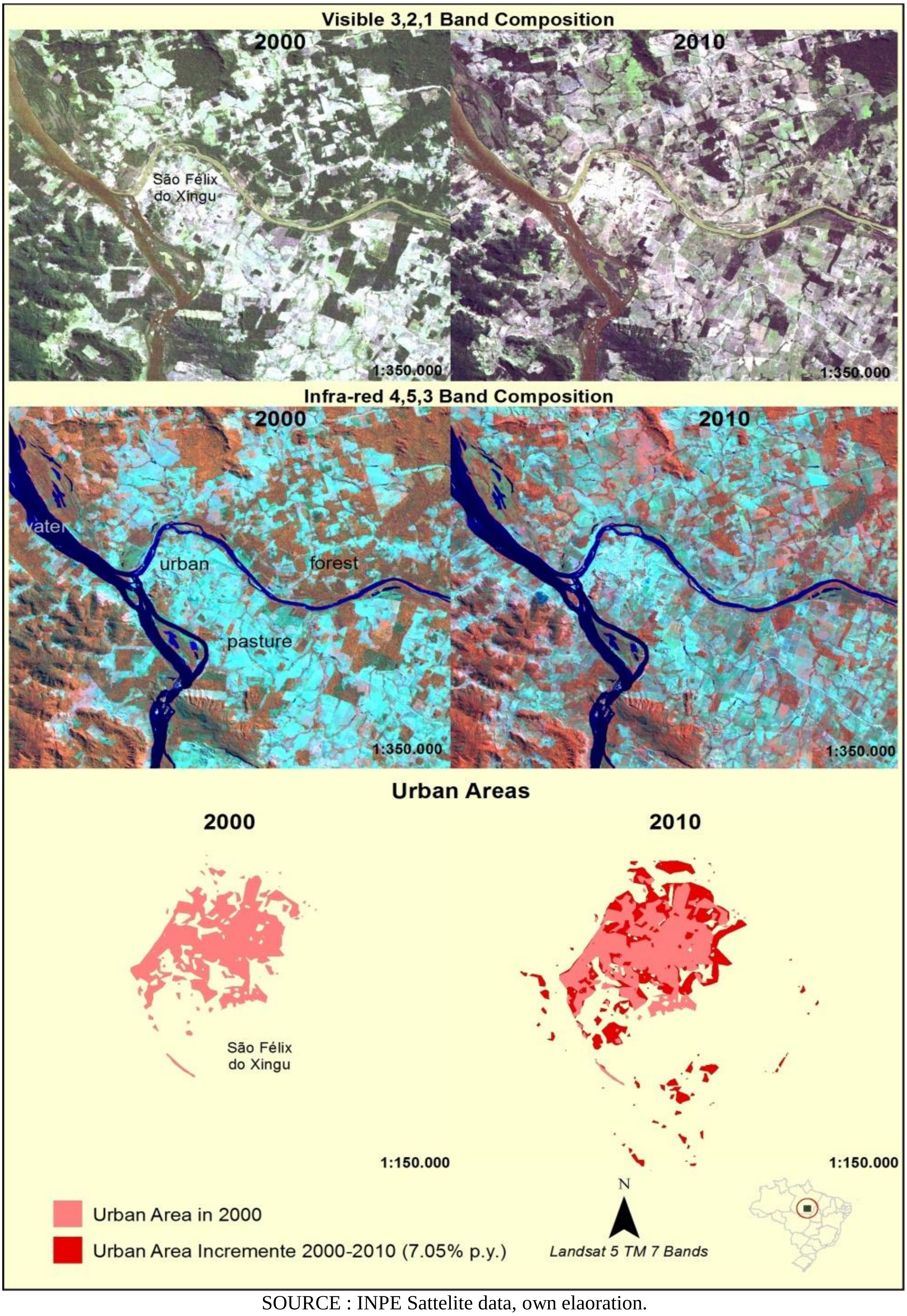
SOURCE : INPE Sattelite data, own elaoration.

Map 3. 3: Cuiabá-Várzea Grande Urban Conglomerate Sprawl, 2000-2010



SOURCE : INPE Sattelite data, own elaoration.

Map 3. 4: São Félix do Xingú Urban Sprawl, 2000-2010



SOURCE : INPE Sattelite data, own elaoration.

As we can see, this municipality grew even more than the metropolitan regions, by an outstanding yearly rate of 7.05%. This is only one example out of several ones showing the size of the urbanization process taking place among medium-sized cities within Brazilian Amazon recently.

In terms of the regional division we have implemented in this study, such historical and geographic background results, nowadays, in the current socioeconomic configuration described in table 4.1.

Table 3. 2: Current Socioeconomic Aspects of Brazilian Amazon and Brazil

Socio-economic Data	Amazon Metropolitan Regions		Rest of Amazon		Rest of Brazil	
	Value	% of Brazil	Value	% of Brazil	Value	% of Brazil
Population (2007)	6,294,629	3.32%	18,856,584	9.96%	164,224,392	86.72%
Agriculture and Livestock GDP (thousand 2008 R$)	1,326,061	0.88%	32,286,665	21.47%	116,776,385	77.65%
Share of Agriculture and Livestock GDP	1.40%	-	21.89%	-	4.20%	-
Industry GDP (thousand 2008 R$)	24,706,486	3.44%	29,702,860	4.13%	664,160,609	92.43%
Share of Industry GDP	26.03%	-	20.14%	-	23.90%	-
Services GDP (thousand 2008 R$)	51,886,648	3.05%	74,006,624	4.35%	1,575,958,292	92.60%
Share of Services GDP	54.67%	-	50.18%	-	56.70%	-
Government GDP (thousand 2008 R$)	16,994,976	3.77%	11,491,458	2.55%	422,486,918	93.68%
Share of Government GDP	17.91%	-	7.79%	-	15.20%	-
Total GDP (thousand 2008 R$)	94,914,171	3.14%	147,487,607	4.88%	2,779,382,204	91.98%
GDP per capita* (R$ per person)	15,079	-	7,822	-	16,924	-

*PIB de 2008 dividido pela população em 2007

SOURCE: Ipeadata. Own Elaboration.

From table 3.2, we note that 25% of the Brazilian Amazon population region live within metropolitan areas, a number far from negligible for a region still considered by many as rural. In terms of sectoral GDP participation, the Rest of the Amazon (apart from the metropolitan areas) exhibits a higher relative participation of Agriculture and Livestock on its GDP composition than the rest of Brazil or the Metropolitan regions of Amazon, which is expected given that, as argued, urbanization is still lower at this regions. Moreover, the GDP per capita from the Rest of the Amazon is about half of the one on Amazon metropolitan regions. Likewise, the value of this variable for the Rest of Brazil is the very close to the one from the Brazilian Amazon metropolitan areas. This confirms what many spatial economics theoretical models predict (see, for example, Fujita & Thisse, 1999): urbanization is usually accompanied by, or even the source of, economic growth and development. This particular aspect will be studied in the next chapter of this thesis.

In terms of size, the Metropolitan regions of the Brazilian Amazon represent only 3.32% of the total Brazilian population, which may be considered an apparently small number. Nevertheless, this number will prove to be very important in the results obtained in this chapter, as it serves as the basic comparative terms to the impacts measured here, as it will become clearer later on. Also, the total population of the Amazon region as a whole, also seem small if compared to the Brazilian total population, in a range of only 13.3%. However, it is important to keep in mind that the deforestation caused by this local population may not be as small as its relative size, especially due to the fact that this population lives much closer to the forest than the rest of Brazilian individuals. In fact, in the following sections we show that apparently relatively small population percentages represent a relatively high share of deforestation, meaning that both closeness of markets to the forest and the growing urbanization process seem to play an important role in terms of environmental impacts.

In respect to deforestation, table 4.2 brings the overall information about this variable within the two regions by which we divided the Brazilian Amazon.

Table 3. 3: Forest Cover and deforestation in Brazilian Amazon

Forest Data	Amazon Metropolitan		Rest of Amazon		Total Brazilian Amazon
	Km^2	% of tot. Br. Amazon	Km^2	% of tot. Br. Amazon	
Total Area	507,588	10.03%	4,551,353	89.97%	5,058,941
"Original"* Forest Area	371,005	9.85%	3,394,329	90.15%	3,765,334
% ("Original"* Forest Area/ Total Area)	73.09%	-	74.58%	-	74.43%
Remanescent Forest Area	354,917	11.69%	2,680,161	88.31%	3,035,078
% (Remanescent Forest Area/ "Original"* Forest Area)	95.66%	-	78.96%	-	80.61%
Acumulated Deforestation	16,088	2.20%	714,238	97.80%	730,327
% (Acumulated Deforestation/ "Original"* Forest Area)	4.34%	-	21.04%	-	19.40%
Increase do Deforestation in 2008/2007	210	1.57%	13,134	98.43%	13,343
% (Increase do Deforestation in 2008-2007 / Remanescent Forest)	0.06%	-	0.49%	-	0.44%

SOURCE: INPE, Own Elaboration.

In table 4.2, it is noteworthy that the majority (88%) of the rainforest is located outside the Metropolitan regions' borders. Seen from another perspective, this number shows that even though Metropolitan regions are typically urbanized areas, they still hold about 12% of the Amazon forest in their municipalities' area. Still, as we will see in the next section, these metropolitan regions can be held responsible for a major part of the deforestation of the land located in its neighbor areas from the Rest of Amazon.

3.5. The Input-Output Model Results

3.5.1. The 3 Regions Productive Structure

Before we continue to the main goal of this chapter, which is to measure the deforestation impacts of local demand vectors and urbanization in Brazilian Amazon, we initially describe the economic structure of the three regions, which by its turn will serve as a basis to analyze the results concerning these impacts.

Therefore, in order to characterize this productive structure, we have calculated, from the Input-Output matrices, the production multipliers, as well as the employment and the value added generators for the three regions chosen in this analysis, always considering the direct, indirect and induced multiplier effects of the inter-regional system (for more methodological details, see Miller & Blair, 2009). Figures 3.2 - 3.10 bring these statistics for each of the 32 sectors in each of the three region of this study.

Analyzing those figures, one first general result that calls attention is the fact that it is easy to note that the two Amazon regions present larger dependence from the Rest of Brazil, than the Rest of Brazil depends on Brazilian Amazon as a whole, economically. As the Amazon region represents only 13% of total Brazilian population, this was a somehow expected result, for in this sense, the Amazon region is a relatively small economy within a much larger one. In order to see that relative dependency pattern, it is important to comprehend how to interpret these figures.

Thus, we explain how to interpret them with an example for the multiplier effects, but a very similar interpretation is also given to the generators: in the case of the figures representing the multipliers, the height of each column *j*, with *j* being one of the 32 sectors of the economy, reflects how much one additional unit of family consumption (from any of the three regions) of goods produced by sector *j* located at the region *x* specified in the figure's name, generates in terms of total output in the economy of the three regions as a whole. Therefore, for example, we can see in Figure 3.2 that one aditional monetary unit (R$) of demand of the Public Services sector in the Metropolitan Regions of Amazon result in the generation of almost 4.5 monetary units in the in ecnomies of the three regions, being slightly more than 2 of these units generated within the Metropolitan regions themselves (shown by the blue colored part of the respective column), a little less than 2 units generated within the Rest of Brazil (shown by green colored part of the respective column), and about 0.5 units generated in the Rest of Amazon (shown by red colored part of the respective column).

One example on how to interpret the generators of employment and value added would be: the first column of Figure 3.5 indicates that for every thousand dollars of additional demand from families of the Metropolitan areas of the Amazon for Agriculture and Forestry, 120 jobs are generated in these Metropolitan Areas, 11 jobs are created in the Rest of the Amazon, and 48 jobs in the rest of Brazil.

Therefore, specifically in our figures, the height of each column represent the size of the multiplier (or generator) effect, with colors representing how this multiplier effects are spread across the three regions (each represented by a different color). In this sense, it is possible to observe that the demands of the rest of Brazil (in all sectors) always tend to generate less multiplier effects on the Amazon region than the demands of the Amazon region (for both metropolitan regions and the rest of Amazon) relatively generate on the economy of the Rest of Brazil.

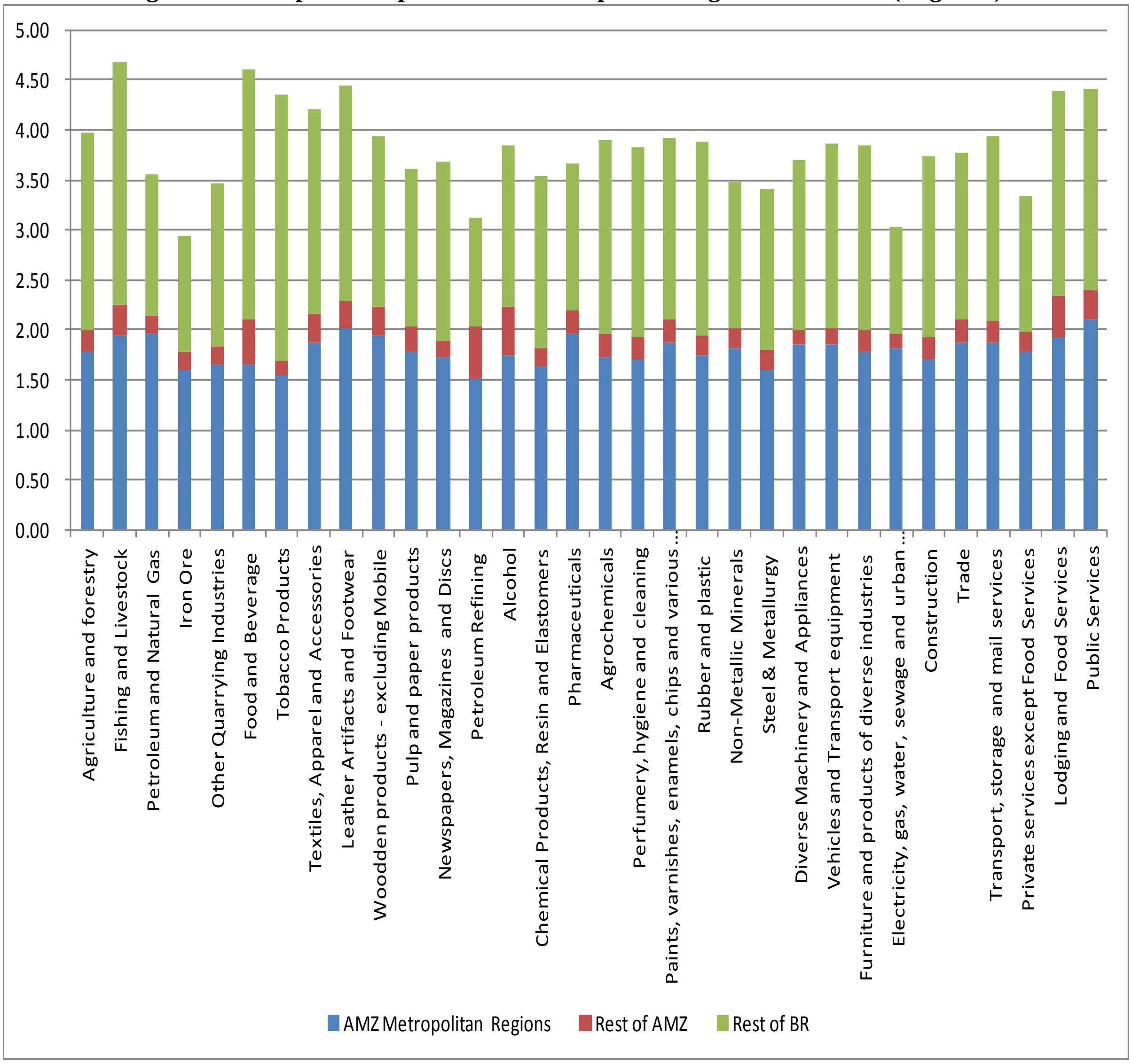
Also, it is important to enphasize that these multipliers and generators all incorporate three different, yet correlated, production phases induced by this additional unit of sector j final demand in region x, which are: the direct production of this unit in sector j itself (in region x); the output indirectly produced in all sectors and regions in order to supply the necessary inputs to industry j, so this sector can produce the aditional unit of demand (and also to supply the inputs for the industries which produced the inputs for sector j); and the additional outputs and inputs necessary to attend the new aditional consumption in all sectors and regions which was induced by the increase in the household income of families who worked on the whole production production of all these outputs. This final income is generated because, in the version of the Input-Output model chosen for this study, the households are the owners of labor and capital involved in the economy’s production chain.

Therefore, moving to the interpretation of the results brought by figures with the multipliers’ results (Figures 3.2 to 3.4), we note that the Metropolitan regions of Amazon generate spillovers in the Rest of the Amazon in the same proportion as the Rest of the Amazon generates spillovers in the metropolitan areas, in general, for all sectors. Bearing in mind that the population of MR's accounts for only 25% of the Amazon as a whole, this is a first evidence that the highly urbanized areas indeed tend to generate higher economic growth and development than rural areas. This result, theoretically, may be attributed to the major presence of increasing returns to scale, agglomeration economies, knowledge spillovers, and several other attributes of urbanization which we will study in deeper detail in the next chapter of this thesis (see, for example, Fujita & Krugman, 1999; Krugman, 1991; Gleaser, 2008; Igliori 2009). Sectorally, it is easy to notice that in both Amazon Regions of our model, Livestock and Fishery and Food and Beverages are the sectors which present the largest output multipliers. As literature revised in the introduction of this chapter, and also table 3.1 show, Livestock are held responsible for most of Amazon deforestation process, from a supply side of the economy perspective. Moreover, Food and Beverage sector is closely connected to Livestock and Agriculture as an intermediate buyer. These evidence altogether lead us to interpret this multiplier effects' result as a first evidence that in terms of deforestation, local demand from the Amazon region as a whole (Regions 1 and 2) might present considerable impacts in terms of deforestation.

In terms of employment generation, Figures 3.5, 3.6, and 3.7 go in the very same direction of the results from the output multipliers: Sectors highly linked to Agriculture, Livestock, and Food tend to be the ones that generate more jobs in all Amazon regions. Furthermore, Figure 3.5 shows that increases in consumption of Food and Beverages, and also of Fuels (Ethanol and Petroleum Refining) in the metropolitan areas of the Amazon region tend to cause an increase in employment in the Rest of the Amazon larger than the average increase caused by other sectors. This evidence is similar to the one discussed regarding the output multipliers, and therefore already supports the argument being created in this section, in which metropolitan areas of Amazon seem to exert a significant deforestation impacts related to their demands for agricultural and livestock products.

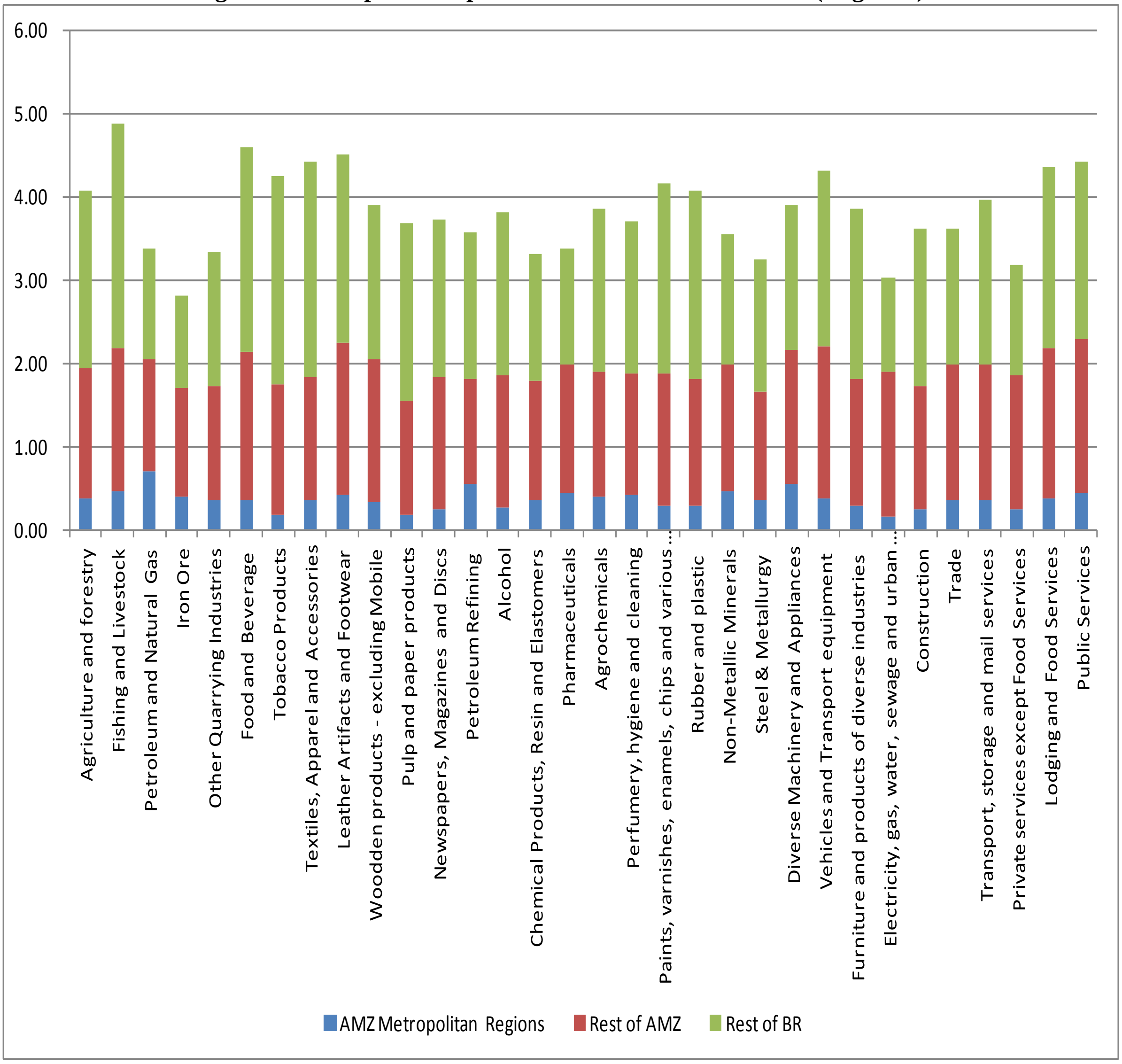
Regarding the generation of added value, once again a similar picture is observed, but with a slight difference: For all regions, in general, even though Agriculture and Livestock are among the most relevant sectors in terms of creation of added value throughout the production chain, this time the Services sectors tend to be the most important ones, according to this measure. This confirms the tendency of urbanization that Brazil and the Amazon region is going through in the last decades, because the development of the Services sectors is, in general, closely connected to urbanization processes, since Urban areas tend to present a higher participation of services in the GDP composition. To which concerns our results, however, we note that once again Agriculture and Livestock sector display an important role in terms of value added, as well as the Food and beverage sector, which corroborates the arguments brought by evidence of the output multipliers and employment generator effects.

Figure 3. 2: Output Multipliers of the Metropolitan Regions of Amazon (Region 1)



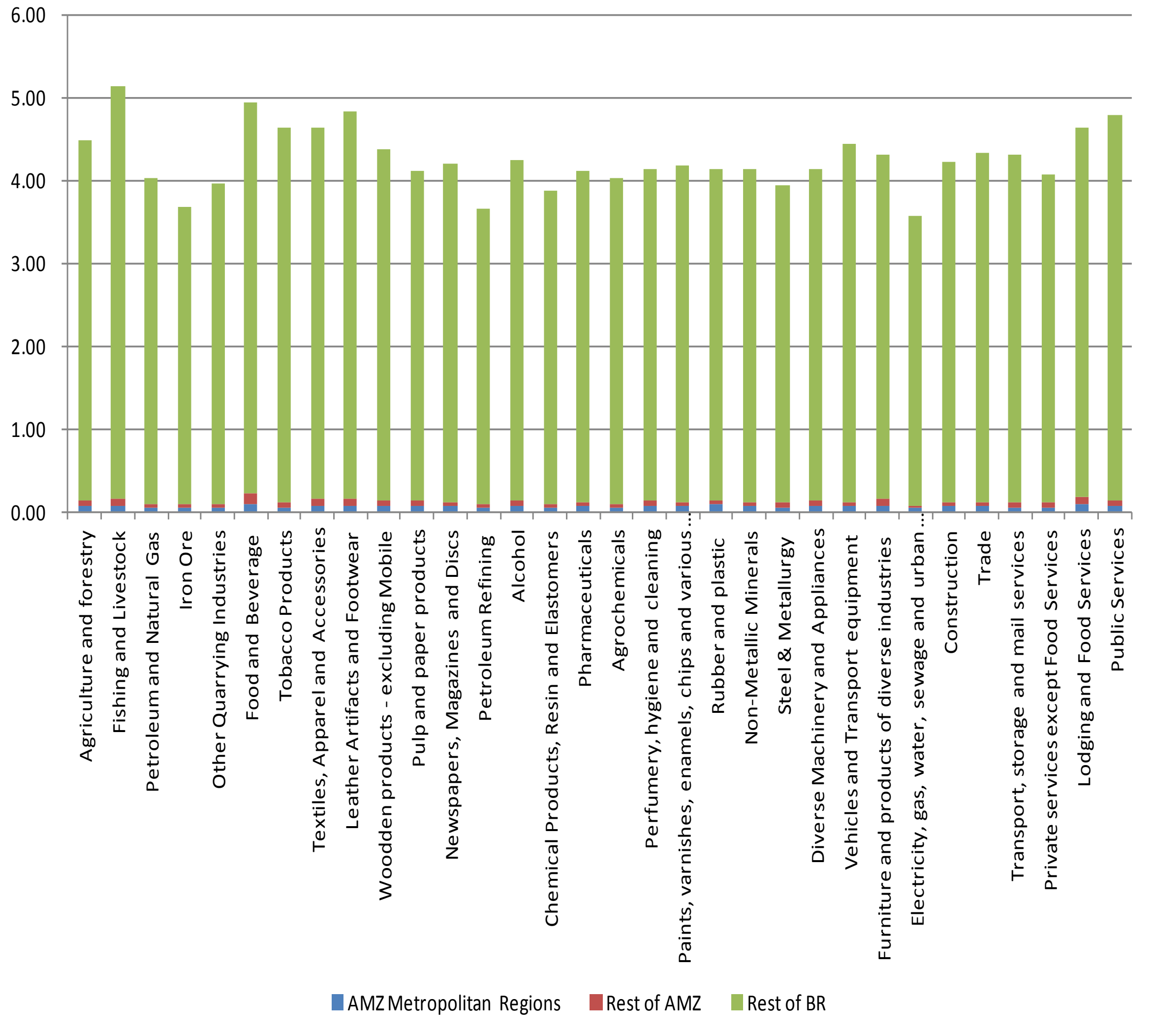
SOURCE: Own Elaboration.

Figure 3. 3: Output Multipliers of the Rest of the Amazon (Region 2)



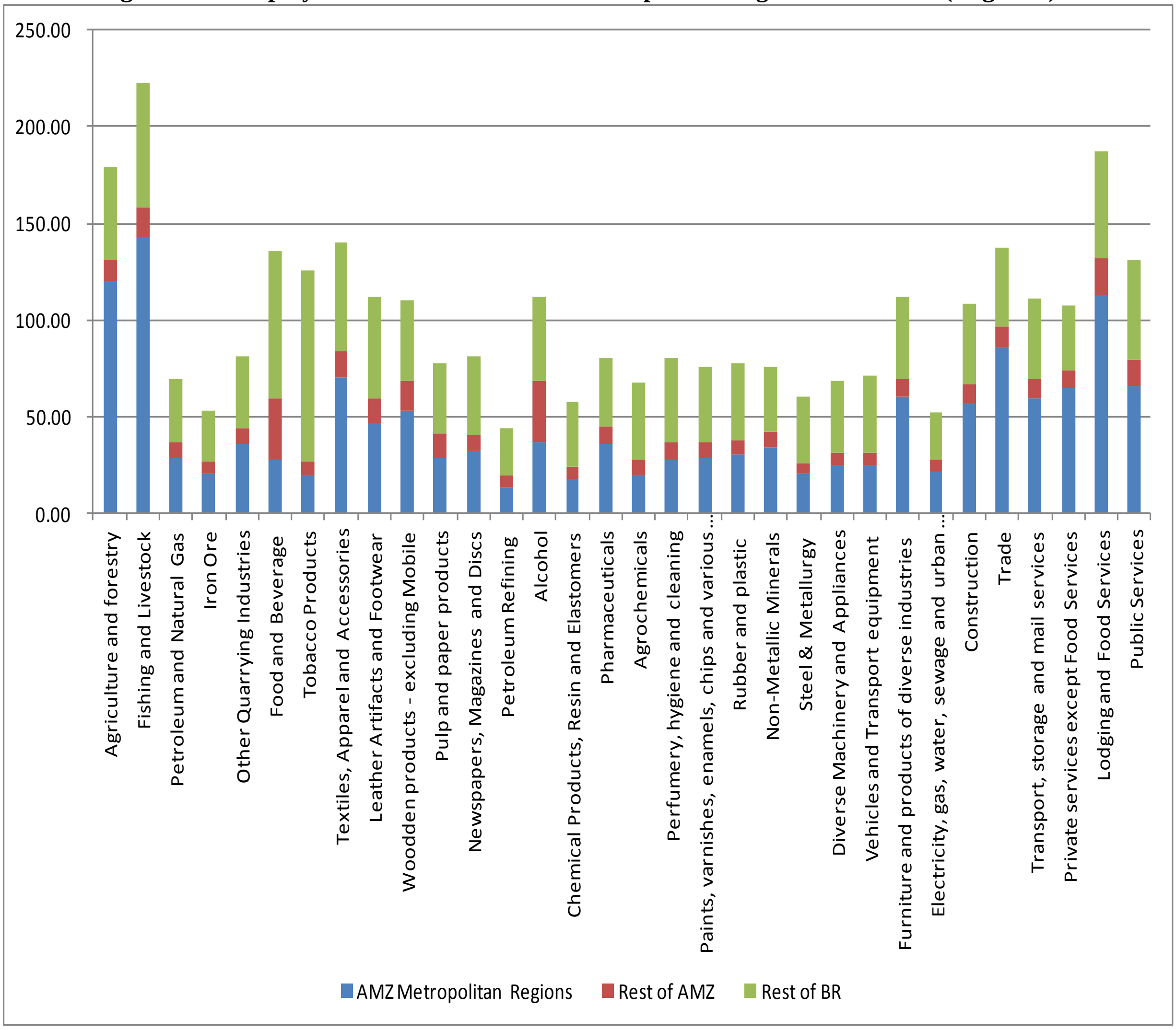
SOURCE: Own Elaboration.

Figure 3. 4: Output Multipliers of the Rest of Brazil (Region 3)



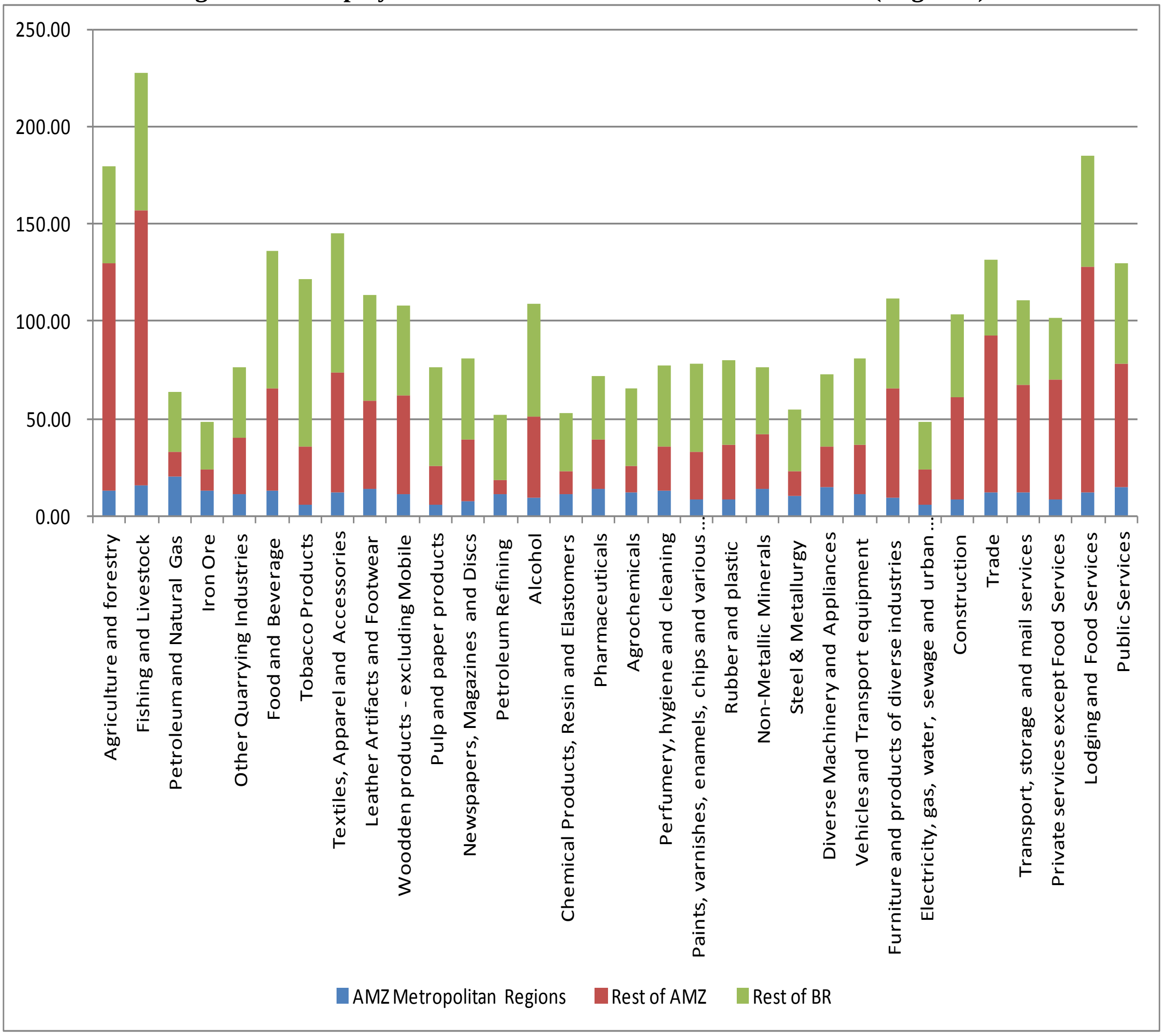
SOURCE: Own Elaboration.

Figure 3. 5: Employment Generators of the Metropolitan Regions of Amazon (Region 1)



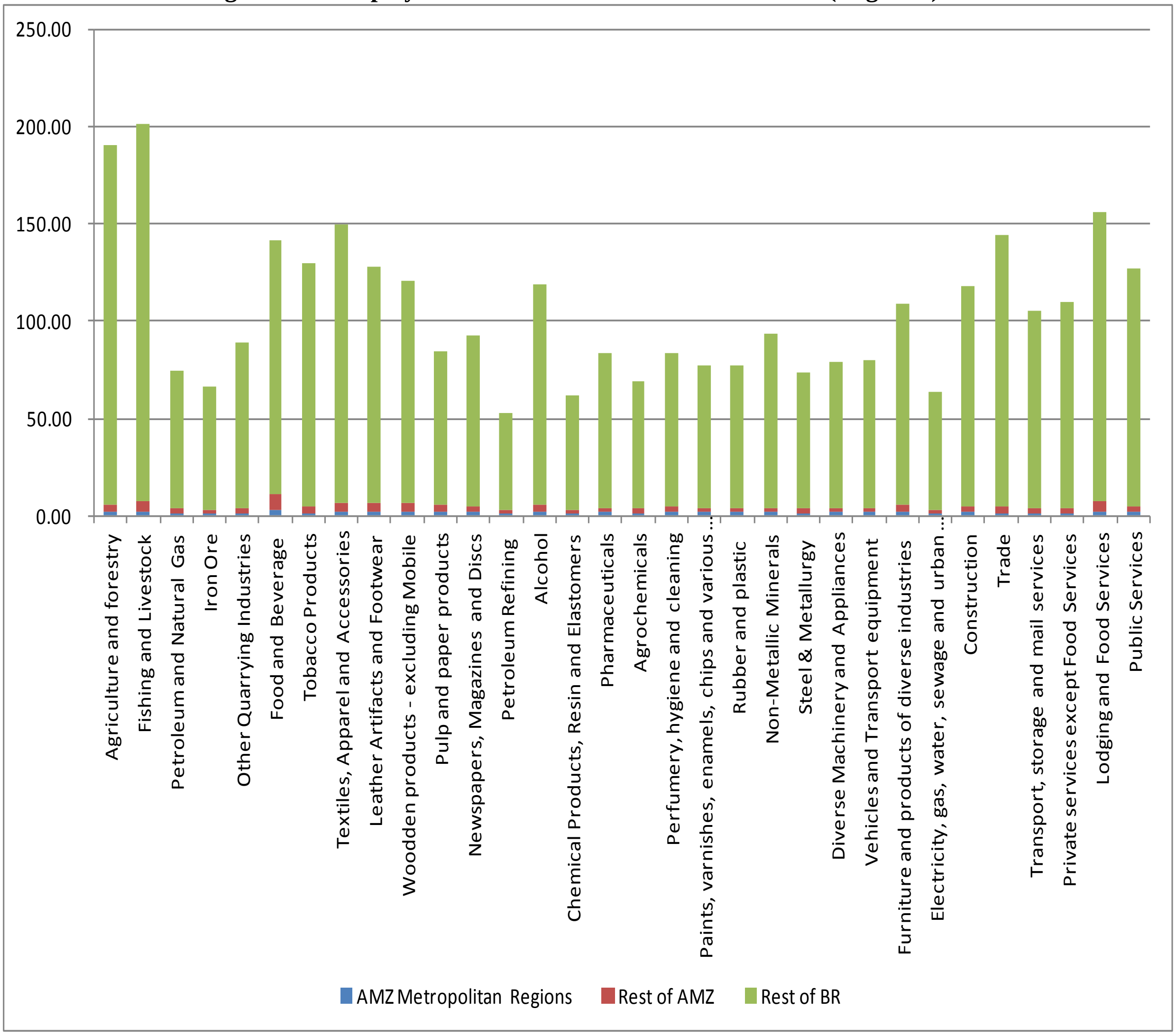
SOURCE: Own Elaboration.

Figure 3. 6: Employment Generators of the Rest of the Amazon (Region 2)



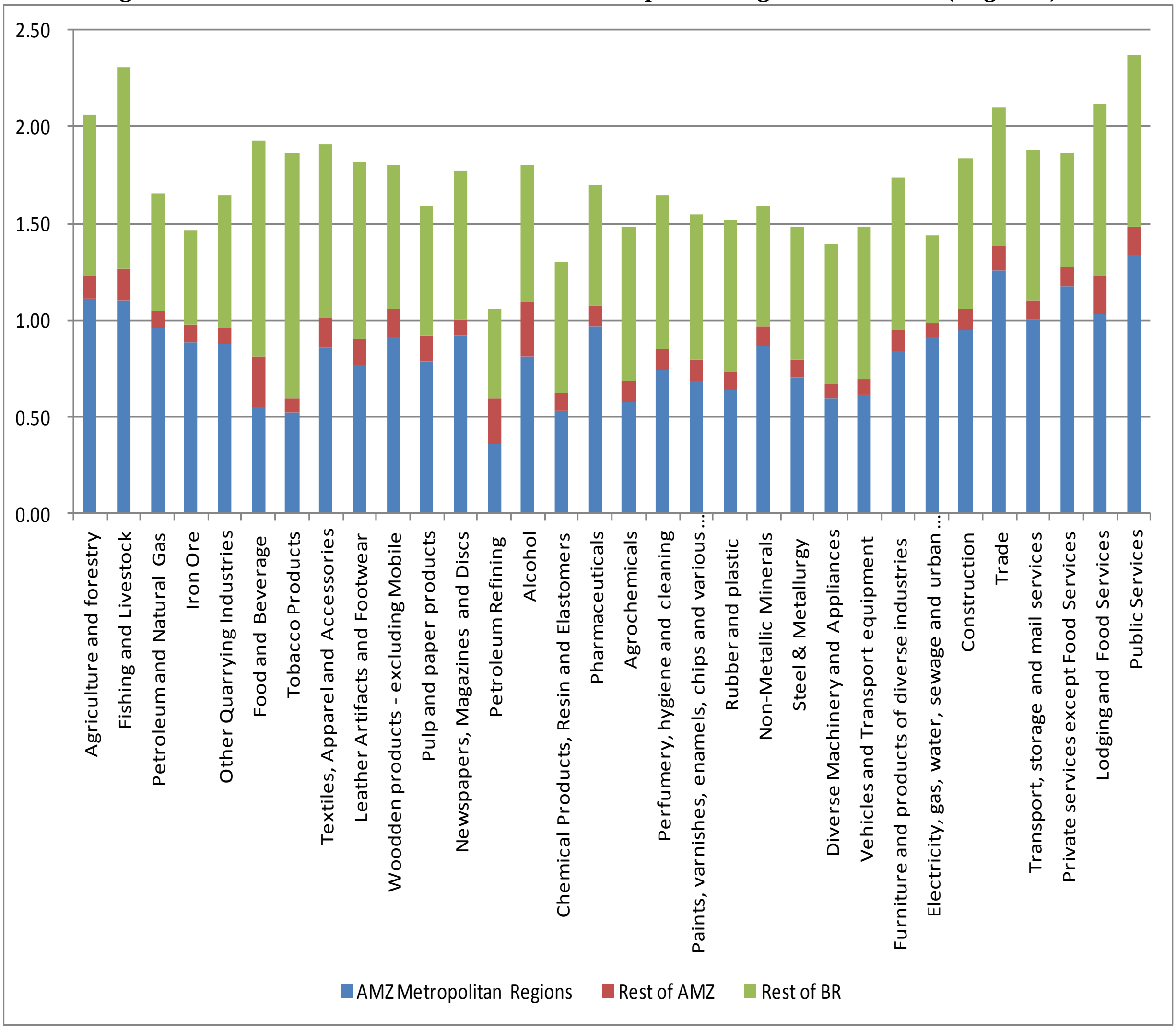
SOURCE: Own Elaboration.

Figure 3. 7: Employment Generators of the Rest of Brazil (Region 3)



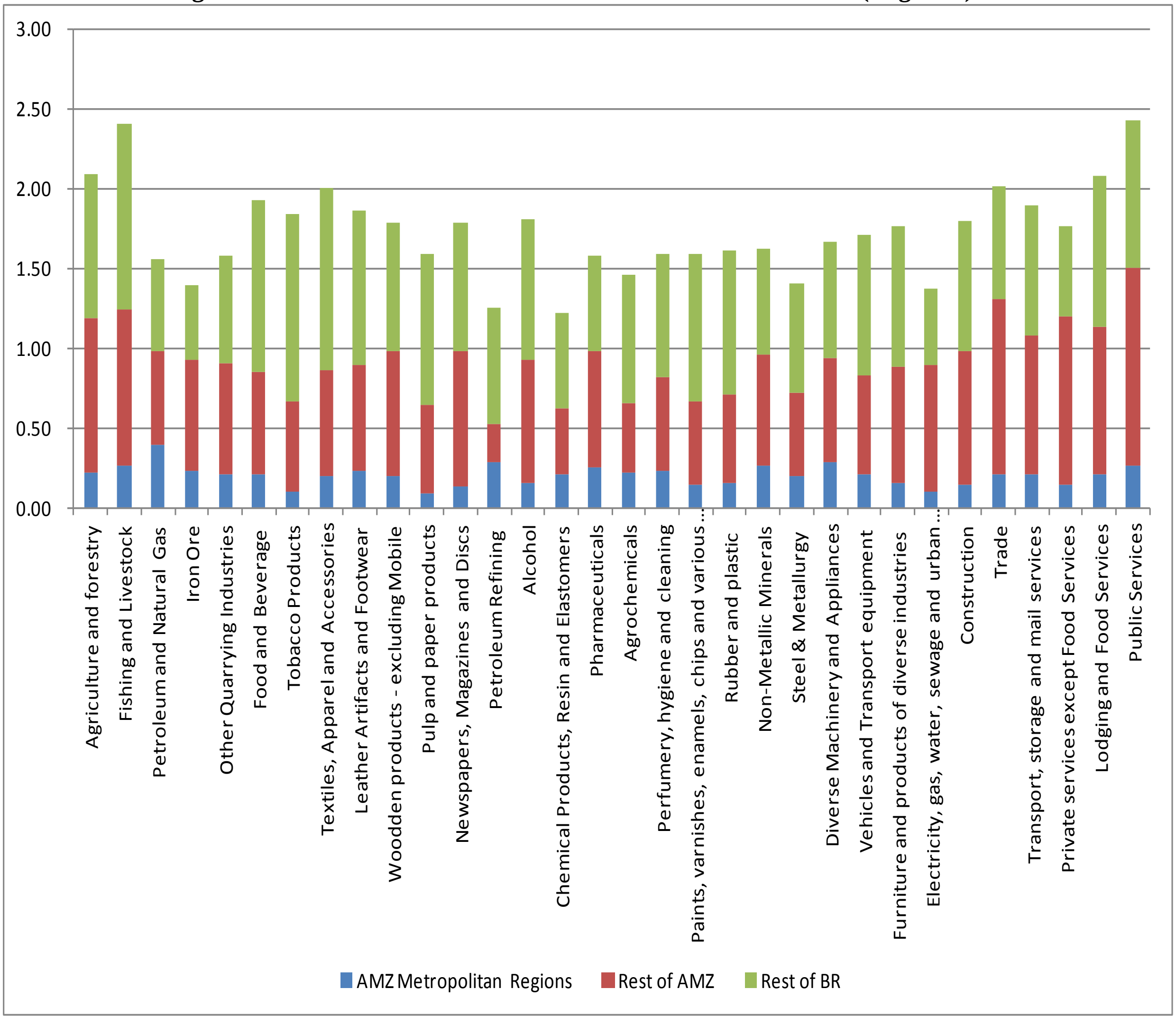
SOURCE: Own Elaboration.

Figure 3. 8: Value Added Generators of the Metropolitan Regions of Amazon (Region 1)



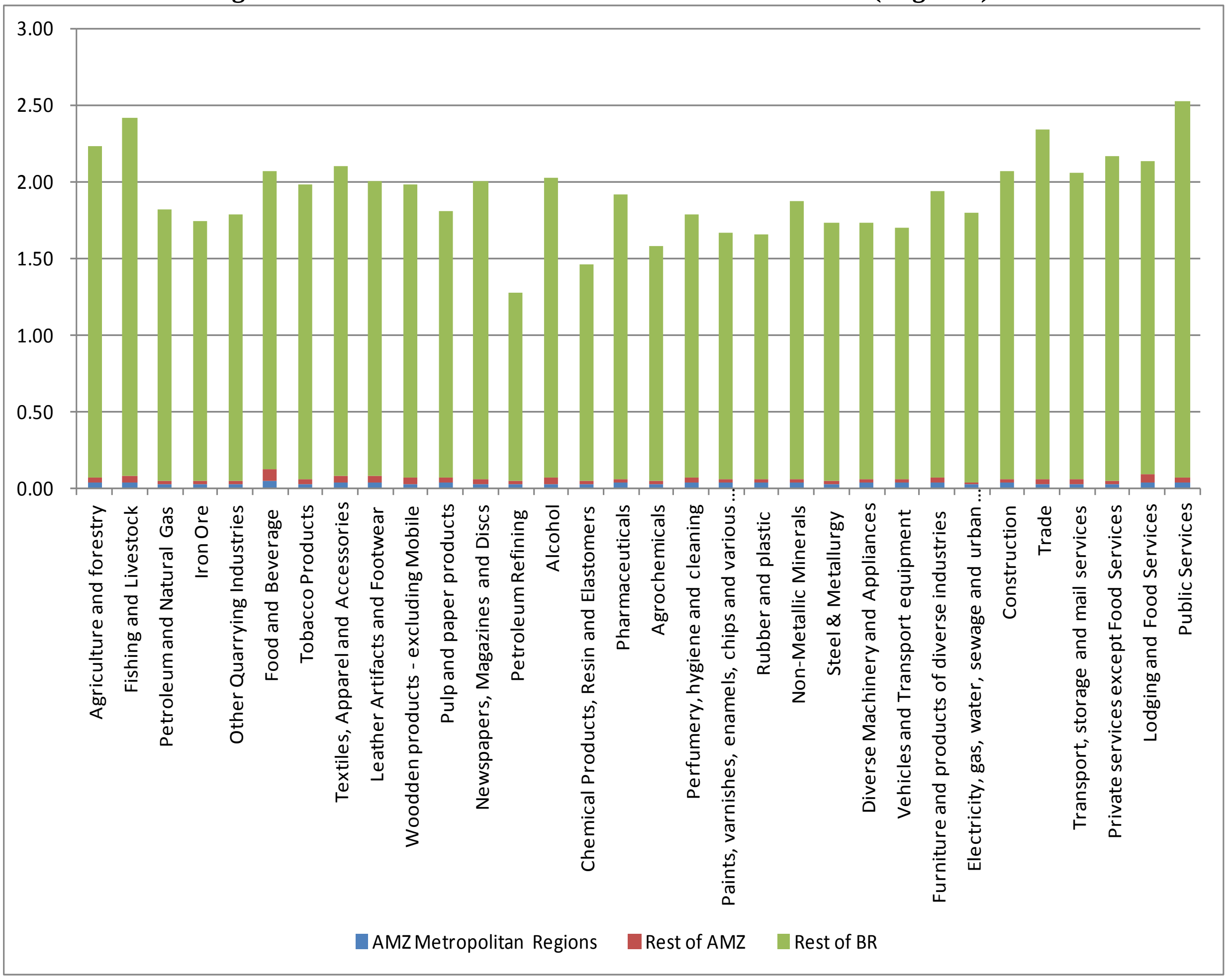
SOURCE: Own Elaboration.

Figure 3. 9: Value Added Generators of the Rest of the Amazon (Region 2)



SOURCE: Own Elaboration.

Figure 3. 10: Value Added Generators of the rest of Brazil (Region 3)



SOURCE: Own Elaboration

3.5.2. Impacts of local demand vectors of Brazilian Amazon

Economic Impacts

Given the description of the productive structures of the 3 Regions, we now continue to the central question of this study, which is try to measure the impacts of the local demand vectors over deforestation in Brazilian Amazon. In order to do so, we start analyzing the economic impacts of these local demand vectors, without losing sight of some "key" sectors more closely related to deforestation. These key sectors were defined accordingly to their connections to Livestock and Agriculture, which as we have seen, are the sectors directly related to deforestation through competition for land use.

Firstly, Table 3.4 shows how the output produced in Region 2 (Amazon except for its Metropolitan regions) in these 10 key sectors was distributed regionally throughout the 3 regions of Brazil, in 2004, and also to which type of consumption these were intended to (household consumption, exports or intermediate consumption). We have chosen to focus on the output destination of Region 2 because, as we have seen in the previous sections, this is the region where the forest is mostly concentrated (88%), which means that its production is closely interconnected with deforestation itself.

Table 3. 4: Output produced in each "key" of the Rest of the Amazon (Region 2), according to its consumption destination

		Household Consumption						Exportations		Intermediate consumption					
		Metropolitan Regions of Amazon		Rest of Amazon		Rest of Brazil				Metropolitan Regions of Amazon		Rest of Amazon		Rest of Brazil	
		Millions of R$	% of Output produced by the sector in the Rest of Amazon	Millions of R$	% of Output produced by the sector in the Rest of Amazon	Millions of R$	% of Output produced by the sector in the Rest of Amazon	Millions of R$	% of Output produced by the sector in the Rest of Amazon	Millions of R$	% of Output produced by the sector in the Rest of Amazon	Millions of R$	% of Output produced by the sector in the Rest of Amazon	Millions of R$	% of Output produced by the sector in the Rest of Amazon
Sectors in the Rest of the Amazon	Agriculture and forestry	1,033	4.1%	794	3.2%	999	4.0%	4,744	19.0%	1,481	5.9%	3,718	14.9%	7,596	30.3%
	Fishing and Livestock	587	4.8%	499	4.1%	682	5.6%	248	2.0%	969	8.0%	2,169	17.9%	3,084	25.4%
	Food and Beverage	2,047	14.2%	2,192	15.2%	2,914	20.2%	2,296	15.9%	430	3.0%	1,422	9.9%	1,609	11.2%
	Leather Artifacts and Footwear	6	2.0%	26	8.3%	17	5.6%	23	7.4%	3	0.8%	49	15.9%	34	11.0%
	Woodden products -	13	0.4%	18	0.5%	31	0.9%	2,071	60.8%	221	6.5%	802	23.5%	712	20.9%
	Pulp and paper products	4	2.0%	3	1.5%	18	9.0%	162	80.0%	12	5.7%	25	12.4%	83	41.0%
	Newspapers, Magazines and	5	9.3%	7	12.6%	6	11.4%	0	0.0%	9	17.8%	22	42.4%	11	20.9%
	Alcohol	20	2.9%	18	2.5%	94	13.2%	0	0.0%	72	10.1%	37	5.2%	184	26.0%
	Furniture and products of	56	24.0%	121	51.3%	37	15.6%	64	27.1%	16	6.7%	25	10.4%	21	8.7%
	Construction	244	4.0%	2,330	38.1%	98	1.6%	580	9.5%	67	1.4%	353	7.7%	30	0.6%

SOURCE: Own Elaboration.

We can observe in table 3.4 that, in terms of household consumption, the output produced to attend the direct vector of final demand from families living within the Amazonian Metropolitan Regions is, in general, very close to, or even greater than the output designated to attend the direct demand from households living in the whole Rest of Brazil. For example, out of the total output produced in the Rest of the Amazon by the sector of Agriculture and Forestry, 4.13% is produced to attend the final consumption of the Amazon Metropolitan Regions households, while 3.99% is designated to attend direct consumption of households in the rest of Brazil. This result is very similar for the industries of Livestock and Fisheries, Leather Artifacts and Footwear, and also Newspapers and Magazines. As for the sectors of Construction and Furniture and Products of diverse industries, these shares of output to attend the Metropolitan regions of Amazon are even more bigger than the share designated to attend households in the Rest of Brazil. Furthermore, if we add up the output designated to attend consumption from families of both Metropolitan and Non Metropolitan Regions of the Amazon (Regions 1 and 2), and compare it to the output directly designated to attend the consumption of families from the Rest of Brazil, then the final consumption of families within Brazilian Amazon generate more direct production in Amazon local industries than do the households from other areas of Brazil, even though the Rest of Brazil's population is far bigger than the population of Brazilian Amazon.

Concerning the Food and Beverage sector, which is one of the most important sectors in terms of deforestation impacts due to its high level of intermediate consumption of inputs produced by the Livestock sector, we can see in table 3.4 that families from the Metropolitan Regions of Amazon consume almost 15% of the total output of this sector produced by this sector in the Rest of the Amazon. Furthermore, if we also take into account the consumption of families from Region 2 along with the consumption from families of the Metropolitan regions, this percentage adds up to 30%, while families from the rest of Brazil account for the consumption of only 20% of the total output from this sector in region 2. Even if we add up exports to the consumption of families from the rest of Brazil, still, we would have 35% of Food and Beverage output being sold outside the Amazon region, against 30% being sold within it. Bearing in mind that the population in the Brazilian Amazon accounts for only 13% of total Brazilian population, is seems clear that in terms of production directly (without taking into account intermediate consumption from other sectors) designated to attend families consumption, households living within the Amazon regions weighs much more than households from abroad or even from the rest of Brazil.

Moreover, our results from table 3.4 suggest that this weight is even heavier for the highly urbanized metropolitan regions of Amazon, as economic theory predicts, due to the higher consumption standards that people from more urbanized areas tend to present. The only exceptions to this direct output destination towards final consumption argument are the sectors of Pulp and Paper Products and Wooden Products, since production in these sectors in the Rest of Amazon is mostly exported to other countries, still according to table 3.4. In terms of deforestation, though, these sectors are not among the main responsible for deforestation, as pointed by literature reviews previously. In this sense, thus, the argument stands.

However, even though this already may be considered as a first evidence supporting that local demand vectors are relatively important as drivers of output and deforestation, it is important to point out that intermediate consumption has not been taken into account yet, therefore, the whole picture is still unfinished. As we are able to notice in Table 3.4, for most of the key sectors chosen, destination of total output produced each sector of the Rest of the Amazon region is rarely mostly consumed directly by households, especially for the cases of Agriculture and Forestry and Livestock and Fishing, which are exactly the ones directly connected to deforestation through land use competition for land use. This happens because these sectors are at the base of the production chain, thus, they serve as suppliers of inputs for most of the other industries.

In order to incorporate these intermediate consumption effects, i.e., the production of inputs needed to produce the output to attend the final demand, we used the Input-Output Inter-regional Leontief matrix as given by equation 3.14, in order to implement in a strategy explained below, which allows us to calculate the output in each sector of each region that will have to be produced to attend the direct consumption of families of each of the 3 different regions, but also accounting all production of inputs in all sectors of all regions that are necessary to produce such output.

This strategy consists of the following. First, we isolate the 4 regional final demand vectors of the system: 1) The vector CF^{RMAM} of consumption from the households of the Amazon Metropolitan Regions, which shows how much the households from the Amazon Metropolitan Regions consume directly (disregarding intermediate consumption) from each sector in each of the three regions; 2) The vector CF^{REAM} of consumption from the households in the Rest of the Amazon, which shows the direct consumption of the families from the Rest of The Amazon (disregarding intermediate consumption) from each sector in each region; 3) The vector CF^{RBR} of consumption from the households in the Rest of Brazil, which shows how much the families from the Rest of Brazil consume directly (disregarding intermediate consumption) from each sector in each region; 4) The vector exportations, which shows how much the rest of the World consume directly from each sector of each region (*EXP*). Each of these four is a (96 x 1) vector, accounting for the 32 sectors in each of the 3 regions.

Then, we pre-multiplied each of these four vectors by the Inverse Leontief Matrix of the Inter-regional system (of dimension 96 x 96), described by the 3.14 equation, as follows:

$$PTS^{RMAM} = L * CF^{RMAM} \tag{3.15}$$

$$PTS^{REAM} = L * CF^{REAM} \tag{3.16}$$

$$PTS^{RBR} = L * CF^{RBR} \tag{3.17}$$

$$PTS^{EXP} = L * EXP \tag{3.18}$$

where PTS^{W} is a 96 x 1 vector in which each element represents the total output that will have to be produced, in each sector of each region, in order to attend the final demand vector (i.e. the consumption) from families of region W, already taking into account all direct and indirect effects (i.e. all output attending both final demand and intermediate consumption) involved in that production, and with $W \in (RMAM; REAM; RBR; EXP)$.

Each of these four PTS^{W} vectors can be split regionally into three vectors, as follows:

$$PTS^{RMAM} = \begin{bmatrix} PTS^{RMAM}_{RMAM} \\ PTS^{RMAM}_{REAM} \\ PTS^{RMAM}_{RBR} \end{bmatrix} \tag{3.19}$$

$$PTS^{REAM} = \begin{bmatrix} PTS^{REAM}_{RMAM} \\ PTS^{REAM}_{REAM} \\ PTS^{REAM}_{RBR} \end{bmatrix} \tag{3.20}$$

$$PTS^{RBR} = \begin{bmatrix} PTS^{RBR}_{RMAM} \\ PTS^{RBR}_{REAM} \\ PTS^{RBR}_{RBR} \end{bmatrix} \tag{3.21}$$

$$PTS^{EXP} = \begin{bmatrix} PTS^{EXP}_{RMAM} \\ PTS^{EXP}_{REAM} \\ PTS^{EXP}_{RBR} \end{bmatrix} \tag{3.22}$$

where each PTS^{Y}_{Z} is a (32 X 1) vector in which each element represents the output produced in each sector of region *Z*, in order to attend the final demand vector (i.e. the consumption) from families from region *Y*, already taking into account all direct and indirect effects throughout all the regions, with $Y \in (RMAM; REAM; RBR; EXP)$ and $Z \in (RMAM; REAM; RBR)$.

It is important to clarify that, differently from the generators and multipliers from the previous section, we are not accounting for the induced effects coming from the income generated in the production chain that families receive in these simulations. The reason for that is because in these calculations, we are treating the consumption from families of each region as exogenous, since we are aiming to see what is the resulting output in each sector in the Rest of Amazon region which results exactly from these consumption vectors. Put in other terms, these demand vectors are the exogenous shocks we are implementing into the Input-Output interregional system, and if we treated them as endogenous, as it is necessary to calculate the induced effects (Miller & Blair, 2009), then we would be double-counting the these shocks.

The general idea behind this strategy is that in the calculations of each of the four PTS_Z^Y we assume that the only source of final demand (i.e. consumption) in the Brazilian economy is the vector of consumption of households in the region *Y*, and then calculate the resulting output in region *Z*. Thus, comparing among each PTS_Z^Y permits us to visualize how much each demand from each region contribute for the production of each sector in all regions.

As we are interested to correlate our results with deforestation, and the Amazon rainforest is mainly concentrated in Region 2 (Rest of the Amazon), in table 3.5 we show the resulting vectors PTS_{REAM}^{RMAM} , PTS_{REAM}^{REAM} , PTS_{REAM}^{RBR} , PTS_{REAM}^{EXP} estimated in our calculations, which refer to the output generated in each sector of the Rest of the Amazon, including direct and indirect production, intended to attend, respectively, the consumption from families of the Amazon Metropolitan Regions (CF^{RMAM}), families from the Rest of the Amazon (CF^{REAM}), families from the Rest of Brazil (CF^{RBR}), and Exportations (*EXP*).

We explain how to interpret the results from Table 3.5 with an example, as follows: assuming that the only source of final demand in the Brazilian economy is the consumption vector, in the year of 2004, of households who live within the Metropolitan areas of the Amazon, then the resulting output that would be produced in the Rest of the Amazon region, considering both direct and indirect production, would be R$ 2,056 million in the sector of Agriculture and Forestry; R$ 1,251 million in the sector of Livestock and Fisheries, R$ 2,457 million in the sector of Food and Beverage, and so on. Moreover, we also present on table 3.5 four columns of shares which represent how much each of these values represent in terms of the total output given by the sum of the 4 outputs which result from each of the 4 consumption vectors. These columns are introduced in order to compare how much each regional vector of final demand can be held “responsible” for the production in each sector in the Rest of the Amazon.

Table 3. 5: Output generated in each sector of the Rest of the Amazon, considering both direct and indirect production, produced to attend the consumption from families of the Amazon Metropolitan Regions, families of the Rest of the Amazon, families of the Rest of Brazil, and Exportations

.	PTS^{RMAM}_{REAM}		PTS^{REAM}_{REAM}		PTS^{RBR}_{REAM}		PTS^{EXP}_{REAM}	
	Millions of R$ (A)	*% of total Demand driven Output [A/(A+B+C+D)]*	*Millions of R$ (A)*	*% of total Demand driven Output [B/(A+B+C+D)]*	*Millions of R$ (A)*	*% of total Demand driven Output*	*Millions of R$ (A)*	*% of total Demand driven Output [D/(A+B+C+D)]*
Agriculture and forestry	2,056	10.88%	1,447	7.66%	7,542	39.90%	7,856	41.56%
Fishing and Livestock	1,251	16.53%	940	12.42%	3,770	49.83%	1,605	21.21%
Food and Beverage	2,457	19.86%	2,542	20.55%	4,381	35.42%	2,989	24.17%
Leather Artifacts and Footwear	9	6.95%	31	24.98%	42	33.66%	43	34.41%
Woodden products - excluding Mobile	49	1.52%	50	1.55%	367	11.34%	2,766	85.59%
Pulp and paper products	10	3.53%	8	2.56%	76	25.82%	200	68.08%
Newspapers, Magazines and Discs	11	19.23%	26	44.34%	15	25.87%	6	10.57%
Alcohol	48	13.84%	33	9.46%	209	59.77%	59	16.93%
Furniture and diverse industries	64	19.81%	137	42.40%	50	15.32%	73	22.48%
Construction	23	11.80%	85	43.18%	26	13.27%	63	31.76%

SOURCE: Own Elaboration.

Analyzing the results from table 3.5, we find more evidence about the relatively high importance that local demand vectors from the Amazon region have, in terms of the productive impacts they exert over the sectors more closely connected to deforestation, in the regions where the forest is mainly located. For the Agriculture and Forestry and for Livestock and Fishing sectors of the Rest of the Amazon region, consumption from families living within the Brazilian Amazon (the sum of Region 1 with region 2) can be held responsible for approximately 20% of the total output needed to attend all 4 demand vectors from the system. For the Food and Beverage sector, this percentage reaches an even higher value: approximately 40%. With exception of the sectors of Wooden products except Mobile and Pulp and Paper Products, which are not very expressive in terms of size, percentages similar to, or even higher than the ones observed for the sectors above. In the Construction sector, for example, more than 50% of total production in the Rest of the Amazon is due to the local demand vectors within the Amazon region.

When compared to how much the demand vector from the Rest of Brazil can be considered responsible for production in these same sectors of the Rest of Amazon, these results show another evidence in favor of the high importance that local (from Amazon itself) demand vectors exert over production in the Amazon rainforest area, and consequently, over deforestation. These local impacts become especially relevant if we bear in mind that population of the Amazon region represents only 13.3% of total Brazilian population, and the shares of output produced I the Rest of Amazon due to the local Amazon demands (from regions 1 and 2) are above 30% (thus, more than twice as big) in 7 out of 10 of the key sectors.

This, as we have justified in the previous section of this work, is a result which goes along with the lessons from spatial and urban economics models, due to the fact that the forest is located within the two Amazon regions, and thus, proximity to local markets makes a smaller population weight relatively more than the much larger market from population in the Rest of Brazil.

Furthermore, these results also show that proximity is not the only factor matters in this sense. Comparing the outputs generated by each of the 2 demand vectors from Brazilian Amazon, it is easy to notice that the Metropolitan Regions demand vector, for most of the sectors, induces a greater share of output production than do the demand vectors from households of the Rest of the Amazon, even though population from these metropolitan areas represents a lower share of the total Amazon population. For the Livestock and fishery, for example, Metropolitan Regions are responsible for 16.53% of the total resulting production in the Rest of the Amazon, while the consumption vector of households from the Rest of the Amazon itself is responsible for only 12%. Similar “within Amazon” results are observed for the other selected sectors in table 3.5. This result suggests that the demand vector of each individual from the Amazon Metropolitan Regions weights more than the demand from an individual living within the Amazon, but outside the Metropolitan areas, in terms of the output that these consumption vectors generate, both directly and indirectly.

As can be noticed, a comparison of the output generated by each demand vector with the size of the population that this vector represents is extremely relevant to achieve our goals here. Therefore, in order to ease these comparisons, we calculated the results from table 3.5 in per capita terms, in which we divided each resulting output in table 3.5 by the respective population whose demand vector induced the production. Table 3.6 brings the results[23].

[23] This *per capita* indicator was not calculated for the exportation vector, for that would imply dividing the output generated by the exports vector in the Rest of the Amazon by the whole population of the World, which would underestimate the results from this vector, because not all countries are consumers of goods and services produced within the Amazon Region, which means that we would have to trace the destination of the exports from the rest of the Amazon, not feasible with the database used in this paper.

Table 3. 6: Per capita output generated in each sector of the Rest of the Amazon, considering both direct and indirect production, produced to attend the vectors of consumption from families of the Amazon Metropolitan Regions, families of the Rest of the Amazon, families of the Rest of Brazil and Exportations

Sector in the Rest of Amazon	PTS_{REAM}^{RMAM}	PTS_{REAM}^{REAM}	PTS_{REAM}^{RBR}
	R$ per capita	***R$ per capita***	***R$ per capita***
Agriculture and forestry	326.6	76.8	45.9
Fishing and Livestock	198.7	49.8	23.0
Food and Beverage	390.3	134.8	26.7
Leather Artifacts and Footwear	1.4	1.7	0.3
Woodden products - excluding Mobile	7.8	2.7	2.2
Pulp and paper products	1.6	0.4	0.5
Newspapers, Magazines and Discs	1.8	1.4	0.1
Alcohol	7.7	1.8	1.3
Furniture and products of diverse industries	10.2	7.3	0.3
Construction	3.7	4.5	0.2

SOURCE: Own Elaboration.

As it is possible to notice in table 3.6, consumption from each individual living in different regions of Brazil and Amazon results in different scales of production in the Rest of the Amazon region. It is also clear that consumption from each individual living within the Amazon Region results in more production in all sectors (including those more directly related to deforestation), than does the consumption from each individual living in the Rest of Brazil. Furthermore, consumption from individuals who live within the Metropolitan regions of the Amazon drives a greater output in the Rest of Amazon, than does the consumption from individuals who live in the Rest of Amazon itself, even though the latter live closer to the forest. As an example, the table shows that the demand from each individual living within the Metropolitan regions of the Amazon results in a total output of R$198.70 in the Livestock sector of the Rest of the Amazon, while the consumption from each individual within the Rest of the Amazon results in R$49.80, and consumption from each individual living in the Rest of Brazil results in only R$23.00 of output in the same region (Rest of the Amazon) and in the same sector (Livestock and Fishery).

Once again, this result is exactly the one that spatial economics models would predict: consumption from population located closer to the forest tend to weight more than consumption from each individual living in farther regions, especially due to lower transportation costs. It is important to notice that distance is not the only issue in this sense. In fact, despite the fact that the metropolitan regions from the Amazon are located farther from the forest than the Rest of the Amazon, the output driven by each family living within these huge urban conglomerates is greater, once urbanization is accompanied by development and economic growth, which tends to lead to higher consumption standards of each individual living in more densely urbanized areas.

Deforestation Impacts from local demand vectors and urbanization in Brazilian Amazon

We now turn specifically to the main goal of this study, which is to measure how much deforestation is caused by the local demand vectors of Brazilian Amazon, in terms of the land use change driven by these in order to make it possible for sectors such as Livestock and Agriculture to attend these demands. In order to do so, we adopted a specific strategy which we describe in the following paragraphs.

Based on the data from the Second Brazilian Inventory of Emissions and Anthropogenic Removals of Greenhouse Gases, as described in the previous sections of this chapter, we were able to obtain the area of forest covered land, measured in hectares, which has turned into pastures for Livestock and cleared lands for Agriculture from 1994 to 2002. In parallel, the model of Input–Output which we use in this study contain the data concerning the total output in the Amazon region (for both Metropolitan and non-Metropolitan areas) that these two sectors produced, as well as the destination (regional and sectoral) of such output, all for the year of 2004. With these in hands, we were able to estimate the deforestation coefficient of these two sectors for the two Amazon regions considered here, following the same concept used to calculate the coefficients of employment and value added:

$$DC^{AGR} = \frac{A_{FOR}^{AGR}}{X^{AGR}} \tag{3.23}$$

$$DC^{LIV} = \frac{A_{FOR}^{LIV}}{X^{LIV}} \tag{3.24}$$

In which DC^{AGR} is the deforestation coefficient of the Agriculture sector in both Amazon regions; A_{FOR}^{AGR} expresses the average yearly area of original forest covered land which has turned into Agriculture land in the Amazon region, whose calculation was made based on the annual rate of land transition between forest and Agriculture in Amazon between the years of 1994 and 2002; X^{AGR} stands for total output, measured in millions of brazilian reais, which the Agriculture sector of Amazon has produced in 2004; DC^{LIV} is the deforestation coefficient of the Livestock sector in both Amazon regions; A_{FOR}^{LIV} represents the average yearly area of original forest covered land which has turned into pastures for Livestock production, whose calculation was made based on the annual rate of land transition between forest and pastures for Livestock in Amazon between the years of 1994 and 2002; and X^{LIV} stands for the total output, measured in millions of brazilian reais which the Livestock sector of Amazon produced in 2004. Thus, the deforestation coefficient of the Livestock sector measures how much one additional monetary unit of production in the Amazon Livestock sector results, on average, in deforestation within the Amazon region, due to the increase in land transition from forest to pasture lands which is necessary to expand the production of this sector in one unit. The interpretation of the deforestation coefficient of Agriculture in Amazon is very similar.

By implementing the calculation of equations 3.23 and 3.24, we have obtained the DC^{AGR} and the DC^{LIV} coefficients for our data These are, respectively, 8.45 and 127.25 hectares per million Reais produced in Agriculture and Livestock. This means that for each monetary unit of output produced by the Livestock sector of the Amazon region, 127.25 hectares may be deforested within the region, based on the annual rate of land transition between forest covered areas and pastures in Amazon, from 1994-2002, applied for the output of the year 2004. This first result complies with literature, which points out that in terms of land use, from a supply side of the economy perspective, livestock is the kind of land use that is mostly responsible for deforestation in the Amazon region.

Given these coefficients, in order to find how much deforestation is due to the demand vector from households of a Region *i*, $\forall i \in [1,3]$, we implement the following procedure: firstly, we multiply DC^{AGR} by the total output of the Agriculture sector in Amazon that must be produced to attend this demand vector, taking in consideration both direct and indirect necessary production of this sector to meet this demand. Them, we do the same calculation for the livestock sector, that is, as we multiply DC^{LIV} by the total output of the Livestock sector in Amazon that must be produced to attend the demand vector of region *i*, taking in consideration both direct and indirect necessary production of this sector to meet this demand. Then, we add up these two values, and then obtain the total deforestation driven by the demand vector from Region *i*.

Here, we present an example of such calculation, using the notation described so far in this study: suppose that we want to measure the total amount of deforestation which is being driven by the consumption of households who live within the Amazon Metropolitan Regions. Then, we must multiply the Livestock deforestation coefficient in the Amazon region (DC^{LIV}), as calculated by equation 3.24, by the elements in PTS_{RMAM}^{RMAM} and PTS_{REAM}^{RMAM} that represent the output of the Livestock sector in the two regions of Amazon (Regions 1 and 2), which had to be produced to attend the consumption vector from households of the Amazon Metropolitan regions, including both direct production of the output designated for the final demand, and also the indirect production of the inputs needed to make such output. The result gives us the total (direct and indirect) deforestation caused by the Livestock sector to attend the consumption of households from the Amazon Metropolitan Regions. Then, we repeat this procedure for the Agriculture sector, obtaining the deforestation caused by the Agriculture sector to attend the consumption of households from the Amazon Metropolitan Regions. Finally, we add up these two deforestation values, then obtaining the total deforestation caused by the demand vector of households in the Amazon Metropolitan Regions.

Reproducing this procedure for all 4 regional demand vectors of our analysis (the demand vector of the households from the Amazon Metropolitan Regions; the demand vector of the households from the Rest of the Amazon; the demand vector of the households from the Rest of Brazil; the demand vector of Exportations), we obtained the results shown in Table 3.7.

Table 3. 7: Deforestation on Brazilian Amazon caused by the demand vectors of each Brazilian region

Regional Demand Vectors	Annual Deforestation (ha)	Population	Annual Deforestation per capita (ha / 100 inhabitants)	Relative per capita Deforestation
Household consumption from families within Metropolitan Regions in Amazon (A)	191,513(16%)	6,747,872	2.8	7.7
Household consumption from families in the Rest of the Amazon (B)	134,110(11%)	16,729,306	0.8	2.2
Household consumption from families within Amazon (Total: A + B)	325,624 (27%)	23,477,178	1.4	3.7
Household consumption from Families in the Rest of Brazil	590,451 (49%)	159,442,364	0.4	1.0
Exportations	283,335 (24%)	NA	NA	NA
Total	**1,199,411 (100%)**	**NA**	**NA**	**NA**

SOURCE: Own Elaboration

As we can see, in absolute terms, about 73% of the Brazilian Amazon deforestation is due to the demand vectors from regions outside the Brazilian Amazon, being 49% attributed to the household consumption of households of the Rest of Brazil, and 24% attributed to exportations for other countries. Thus, household consumption from within the Amazon region, in absolute terms, is responsible for 27% of the yearly Amazon deforestation. At a first glance, such Amazon demand vectors' share may be considered small, as it is lower than the share attributed to the Rest of Brazil. However, once again considering that the Amazon population represents only 13% of total Brazilian population, it becomes clear that each individual within the Amazon region weights much more, in terms of the deforestation impact he causes, than the individuals from outside this area. This is confirmed when we measure the deforestation caused by each regional demand vector in per capita terms[24], as we have made in the last two columns of table 3.7. Analyzing these specific results, by analyzing column 3 of table 3.7 ("Deforestation per capita"), we notice that one hundred individuals living within the Amazon region (Metropolitan regions or the Rest of Amazon) caused, on average, a deforestation impact of 1.4 hectares in 2004, due to their consumption vector, whereas one hundred individuals living in the Rest of Brazil caused a deforestation impact of only 0.4 hectare in Brazilian Amazon forest in this same year, according to our calculations. Seen from another perspective (as in column 4 of table 3.7), this means that one individual who lives in the Amazon region exert a deforestation impact over the Amazon forest 3.7 times higher than one individual from the Rest of Brazil.

A similar comparison concerning the Brazilian Amazon divided into the two regions proposed here shows that one individual living within the Metropolitan regions of the Amazon exert a deforestation impact 7.7 times higher than one average individual living in the Rest of Brazil, which means that individuals living within the Amazon Metropolitan regions exert a deforestation impact two times higher than individuals living within Amazon, but outside its Metropolitan regions, even though the forest is mostly located outside these metropolitan centers. Therefore, this may be considered as an evidence which confirms that the urbanization process happening in Brazilian Amazon may be resulting in higher deforestation levels, since the individuals who live in more urbanized areas are also the ones who exert the highest deforestation impacts, even being farther from the core of forest than other Amazon residents. Furthermore, this result may be considered in accordance to spatial and urban economics theories which defend that urbanization comes along with economic growth and development (see Gleaser, 2008, and Fujita & Thisse, 1999), since it seems likely that the higher deforestation impact from the consumption vector of individuals living within Amazon metropolitan regions are probably due to the greater consumption standards which

[24] We did not calculated the per capita results for the exportations vector, since we would be underestimating this result, because we would have to divide the resulting deforestation caused by exportations by the population of all countries, however, not every country imports from the Amazon region.

urbanization might have provided them recently. This specific point is deeply addressed in the next chapter of this thesis.

It is important to remind, at this point, that even though this result may be considered highly expected by economists in general, for some reason it is an aspect of Amazon deforestation which is still being overlooked by empirical literature on the subject, in spite of being essential in terms of policy prescription for forest preservation.

Finally, one last consideration we make is that these results refer to the year of 2004. As the urbanization process taking place in Brazilian Amazon is still an ongoing and increasing process within the region, then it is most likely that the deforestation impacts of local demand vectors and local urbanization which have been presented here are considerably larger nowadays.

3.5.3. Concluding remarks

Brazilian Amazon has been going through an evident process of population boost and growing urbanization in the last two decades. Spatial economics models clearly point out that such process may bring relevant impacts over local land use and deforestation, due to development and growth associated with such urbanization, and also because of the lower transportation costs working as incentives to sell locally the output produced in previously forest covered areas. However, somehow, this process is still being overlooked by literature when it comes to the matter of investigating the main causes of the Amazon rainforest deforestation.

That being put, this chapter provided an attempt to fill part of this gap, by trying to measure how much of this deforestation may be attributed to the consumption of goods and services by households who live within the Brazilian Amazon region, also comparing it to deforestation driven by consumption of individuals who live outside the Amazon region. Moreover, we have also attempted to incorporate the referred urbanization process as an important aspect of these local demands drivers. For such, we have isolated the deforestation impacts attributed specifically to the five Amazon Metropolitan Regions, which present the highest urbanization rates of the region, and also compared these impacts to the ones caused by the demand vectors of households living in the Rest of Brazil, as well as to the ones caused by the consumption of families who live within the Amazon, but outside those Metropolitan regions.

Using an Inter-regional Input-Output model with socioeconomic data, and crossing this database with information on land use transition from forest areas to agricultural and livestock land use, we found robust evidence that these local demand vectors play an important role in terms of driving deforestation within Brazilian Amazon. Results show that even though local population from the Amazon region represents only 13% of total Brazilian population, it drives around 27% of the total yearly deforestation within the region. Also, the demand vector from families who live within the Amazonian Metropolitan Regions is responsible for more than a half of this 30% rate, even though only 25% of the Amazon population live within these areas. In per capita terms, results show that the demand vector from one individual living within the Amazon region, but outside the Metropolitan areas, generates 2.2 more deforestation than the consumption vector of one individual living outside the Amazon region. Moreover, the consumption vector of one individual living within the Amazonian Metropolitan Regions causes a deforestation impact 7.7 times higher than the impact of the demand vector from one individual living outside the Brazilian Amazon.

Furthermore, the results from the economic multipliers, generators, and sectoral analysis calculated by the Input-Output methodology implemented in this study all corroborate this importance of urbanization and local demand as drivers of deforestation in Brazilian Amazon.

Finally, bearing in mind that all results presented here refer to the year of 2004, and Amazon local urbanization process and population growth still have been increasing rapidly ever since, then it is most likely that the results presented here are probably underestimated if extrapolated to more recent years. This means that local demands and urbanization probably exert a deforestation impact even higher than the ones estimated here nowadays, even though our results already show that these impacts are of extreme relevance to forest conservation.

Therefore, we conclude by affirming that these local demand and urbanization impacts cannot remain being overlooked by literature when it comes to try to determine which are the main causes of deforestation in Brazilian Amazon, especially concerning the elaboration and implementation of policies to prevent such deforestation in the future.

4. COMERS AND GOERS: DISENTANGLING MIGRATION FLOWS IN BRAZILIAN AMAZON

4.1. Introduction, Motivation and Objectives

In the previous chapters of this study, we have shown that Brazilian Amazon is going through a process of growing urbanization and population growth in recent decades, and have found evidence that such process seems to be causing the emergence of a trade-off between economic development and deforestation within the region, as immediate impacts of it.

Given this scenario, one complementary topic which emerges from this discussion regards the drivers of such recent population increase and high urbanization rates. Specifically, in order to complement the understanding of the full picture related to these processes, it seems necessary to investigate what might be causing these changes in Amazon occupation, and how local population might be changing in terms of its main characteristics. Comprehension of these aspects is especially important to future policy designs of local occupation, deforestation and economic growth.

Essentially, urbanization and population growth are associated with two possible direct and non-mutual causes: migration and vegetative growth. Historically, Brazilian Amazon has been occupied mainly by immigration flows from other countries and other parts of Brazil. Recently, on the other hand, this picture seems to be changing, as we further investigate ahead: in the last decade, census data indicates that the population growth addressed in this study is mostly based on local vegetative growth. Nevertheless, migration still may be considered as an important aspect of the demographic changes going on the region, especially because it may drive shifts in the local population socio-economic characteristics, which by its turn may cause different environmental and economic impacts, similarly to the ones addressed in the previous chapters of this thesis.

4.1.1. Brazilian Amazon: Historic occupation based on immigration

Historically, it is a well-known fact that the occupation of Brazilian Amazon by non-indigenous population, who molded the capitalist structure of its current regional economy, occurred lately when compared to the rest of Brazil, and was mostly based on immigration from southern areas of the country and abroad populations. Until the decade of 1960, the region as a whole experienced a very slow and irregular growth, both economically and in terms of population. Becker (2013) points out that in 1777 only three hundred people lived in Manaus (the current largest city of Brazilian Amazon, with over 4 million inhabitants), whereas in Rio de Janeiro population had already surpassed 35 thousand inhabitants in the same year. Urbanization and demographic growth, in this period, was mostly based on a cyclical structure, with short periods of expansion followed by long periods of stagnation. These short-term expansion periods were generally based on immigration flows intended to exploit recently found (at the time) natural resources, with the long stagnation periods following after international crisis in these products markets. The most famous example among these short-term expansion periods was the Amazon's rubber cycle, between 1879 e 1912, where latex extraction created an impulse of migration flows towards the region, with migrants coming mostly from the Northeast macro-region of Brazil, promoting a reasonable growth in Manaus, Belém and Porto Velho, three of the major cities within brazilian Amazon nowadays. However, these outbreaks were very small when compared to the colonization level of other areas in Brazil (Becker, 2013).

In the decade of 1960, this picture has changed drastically. With the new (at the time) Brazilian federal government dictatorial regime, a huge change on the Amazon geopolitical concerns has emerged, and policies such as an the National Integration Plan, in which definitive occupation of Amazon was intended, started to be implemented. This Plan was based on infrastructure investments, fiscal incentives and low interest rates loans to entrepreneurs who moved towards Brazilian Amazon. As intended, this resulted in an expressive regional population and urban growth in the following three decades (Becker 2013), mainly due to liquid[25] immigration flows.

These explicit policies ceased in the 1980 decade. However, their legacy resulted in a significant the development of the region's internal dynamics of occupation: an analysis of Census data from the following decades shows that both population growth and economic development, as well as urbanization, have continued to increase in the following decades. As previously discussed, population grew by 29% from 1991 to 2010; urban population grew from 56% to 71%. Therefore, even though the majority of the government incentives towards occupation have ceased more than 2 decades ago, positive shifts in local urbanization and population continued to exist within the region.

[25] By liquid, we mean gross immigration to Brazilian Amazon minus gross emigration from this region.

However, a brief analysis of migration data from Brazilian demographic census shows that such recent population growth was no longer caused by large migration inflows towards the region, which means that population of Brazilian Amazon grew mostly due to vegetative growth in the last two decades, oppositely from previous decades. We further detail this point next.

4.1.2. Brazilian Amazon recent migration flows: population increase based on vegetative growth

The migration flows which are studied throughout this chapter are the so called "fixed-period" immigration flows towards Brazilian Amazon from 2005 to 2010 (departing from the Rest of Brazil), and the "fixed-period" emigration flows departing from Brazilian Amazon towards The Rest of Brazil from 2005 to 2010. Besides, specifically in this section, where we describe and compare the size of these flows, we also calculate such "fixed-periods" migration flows for the period 1995-2000.

This "fixed-period" type of migration is built as the following: the database containing the migration data which we work with is the IBGE[26] Brazilian Census for the year of 2010, which contains information about the municipality each individual used to live in 2005, and in which city this same was living in the year of 2010. According to the criterion which we have adopted, if these two municipalities were different, it is considered that the individual has migrated. If these are the same, then we assume that no migration has occurred for this individual[27]. Thus, we are considering as migration only the flows that somehow endured along this fixed period, which is exactly the definition of "fixed-period migration". This means that any individual who left one municipality after 2005, and then returned to this same origin before 2010 is considered a non-migrant, according to our perspective. The choice of such criterion was mainly due to database restrictions, as information on this kind of migration is the most well-detailed and consistent (in terms of fewer missing data) in Brazilian 2010 Census database. We are aware that this type of migration flow present important limitations, such as not considering the period between 2000 and 2005, or discarding migrants who have migrated and returned to their origins between 2005 and 2010. However, from the perspective of trying to include all types of individuals from the Brazilian population in our sample, this database might be considered satisfactory, as it is built over

[26] Brazilian Institute of Geography and Statistics ("*Instituto Basileiro de Gerografia e Estatística"*).
[27] The same method was applied for the 2000 Census database, in order to calculate the fixed-period migration flows in the period 1995-2000, as described in this section.

Census data, which means that it includes all kinds of Brazilian individuals[28] in the sample. Also, not taking in consideration migration flows which have last for less than 5 years may not represent a major problem in terms of our goals, since we are mostly interested in analyzing long run changes in the population composition, and migrants who returned to their shortly after their departure may not be considered as drivers of long run changes.

Tables 4.1 to 4.4 bring a brief geographic description of these migration flows regarding Brazilian Amazon, according to their origins and destinations classified by States and Macroregions of Brazil. Figures 4.1 and 4.2 bring this same information visually.

Table 4. 1: Amazon Immigration Flows, by Origin Macroregion, 2005-2010

By region		
Macroregion	**Flow (migrants)**	**%**
Northeast	69,067	31%
Southeast	64,724	29%
Central	55,535	25%
South	35,121	16%

SOURCE: IBGE 2010 Census data, Own Elaboration

Table 4. 2: : Amazon Emigration Flows, by Destination Macroregion, 2005-2010

By region		
Macroregion	**Flow (migrants)**	**%**
Central	89,832	40%
Southeast	75,072	33%
Northeast	44,868	20%
South	30,898	14%

SOURCE: IBGE 2010 Census data, Own Elaboration

From these tables, we can see that immigration flows towards Amazon (departing from the Rest of Brazil) between 1995 and 2000 were slightly superior than the size of these immigration flows between 2005 and 2010 (see table 4.3), and the size of emigration flows from Brazilian Amazon (towards the Rest of Brazil) from 1995 to 2000 was slightly lower than the size of these flows in the period of 2005 to 2010 (see table 4.4). Furthermore, the overall Brazilian Amazon emigration from 2005 to 2010 was also higher than the overall Brazilian Amazon immigration in the same period. Still, as previously showed, population and urbanization continued to grow within the region during the 2000 decade. This evidence

[28] Differently from other databases, such as the one used by Freguglia (2007) which consists of formal workers only, although the author's objectives were not to specifically study determinants of migration, since he only uses migration as one important explanatory variable to investigate the causes of income differentials among Brazilian states.

suggests an interesting pattern of the population increase happening within Brazilian Amazon: such growth, in the last decade, was mostly based on a self-reproduction dynamic of local population, which was already living within Brazilian Amazon borders, than it was on liquid immigration movements. Or, put in other terms, local population grew by vegetative growth throughout the last decade, in opposition to its historic tendency of occupation via immigration flows. This, per se, shows that the recent pattern of population growth and urbanization taking place within the region might be considered as a structural change in terms of local occupation, and evidences a relative endogeneisation of this process, in terms of the its local population (and internal markets) being less dependent on external inflows to grow.

Table 4. 3: Amazon Immigration flows, by Origin State

AMZ Imigration flow, by Origin State		
State	**Region**	**Imigration (Nº of migrants)**
Goiás/GO	Center-West	33,598
São Paulo/SP	Southeast	32,265
Paraná/PR	South	20,908
Minas Gerais/MG	Southeast	16,734
Maranhão/MA	Northeast	13,364
Rio de Janeiro/RJ	Southeast	12,879
Piauí/PI	Northeast	12,331
Mato Grosso do Sul/MS	Center-West	12,031
Ceará/CE	Northeast	12,009
Bahia/BA	Northeast	10,028
Distrito Federal/DF	Center-West	9,907
Pernambuco/PE	Northeast	8,892
Rio Grande do Sul/RS	South	8,543
Santa Catarina/SC	South	5,669
Alagoas/AL	Northeast	5,173
Paraíba/PB	Northeast	3,590
Espírito Santo/ES	Southeast	2,846
Rio Grande do Norte/RN	Northeast	2,570
Sergipe/SE	Northeast	1,109
AMZ Immigration Flow (2005-2010)		**224,447**
AMZ Previous Immigration Flow (1995-2000)		**248,640**

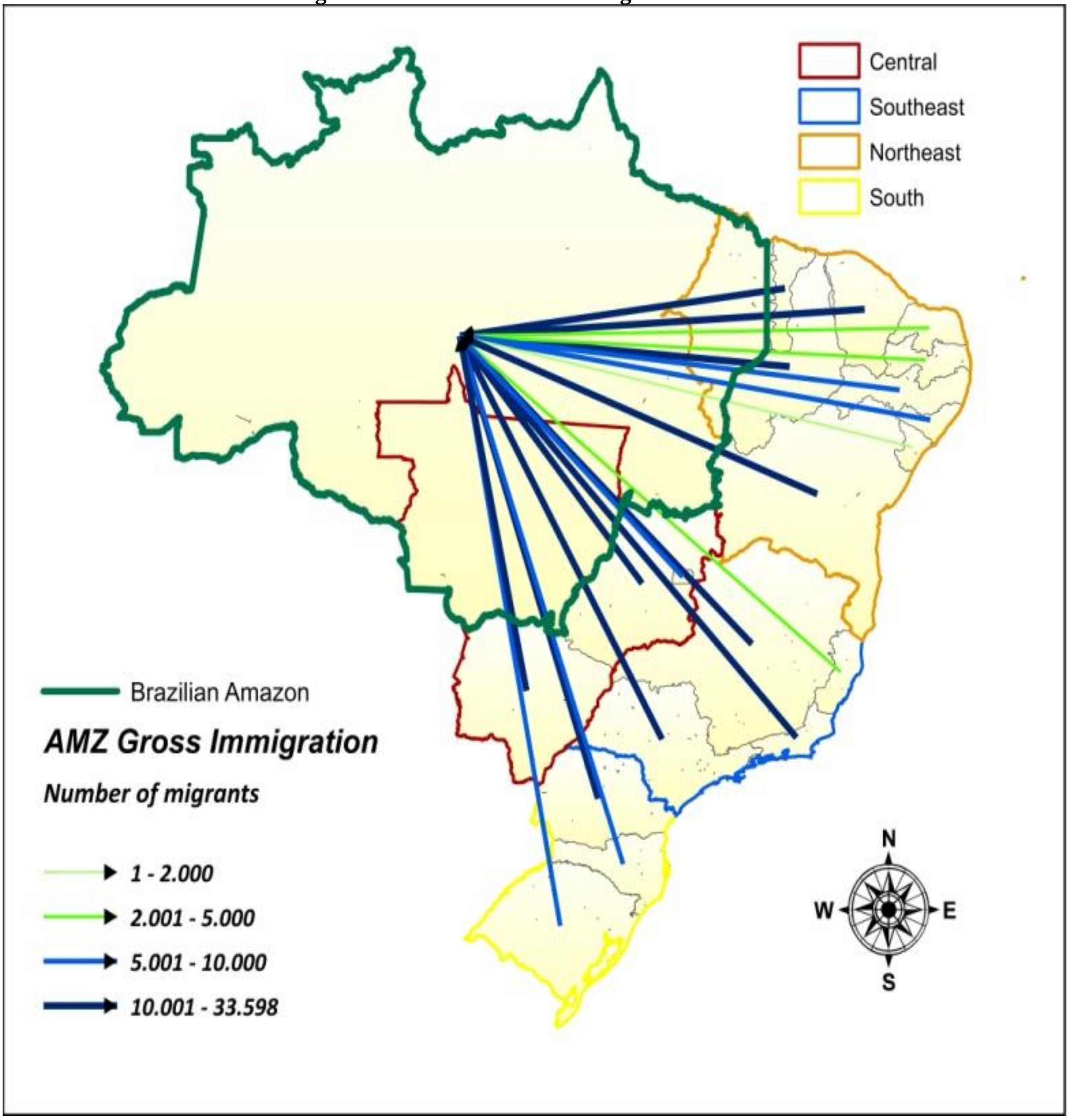
SOURCE: IBGE 2010 Census data, Own Elaboration

Table 4. 4: Amazon Emigration Flows, by Destination State

AMZ Emigration flow, by Destination State		
State	**Region**	**Emigration (Nº of migrants)**
Goiás/GO	Center-West	61,278
São Paulo/SP	Southeast	41,374
Distrito Federal/DF	Center-West	17,155
Paraná/PR	South	17,085
Minas Gerais/MG	Southeast	15,020
Rio de Janeiro/RJ	Southeast	14,753
Mato Grosso do Sul/MS	Center-West	11,399
Ceará/CE	Northeast	9,672
Maranhão/MA	Northeast	8,493
Santa Catarina/SC	South	8,340
Piauí/PI	Northeast	7,822
Bahia/BA	Northeast	6,829
Rio Grande do Sul/RS	South	5,472
Pernambuco/PE	Northeast	4,910
Espírito Santo/ES	Southeast	3,926
Paraíba/PB	Northeast	2,465
Rio Grande do Norte/RN	Northeast	2,417
Alagoas/AL	Northeast	1,365
Sergipe/SE	Northeast	894
AMZ Emigration Flow (2005-2010)		**240,670**
AMZ Previous Emigration Flow (1995-2000)		**235,032**

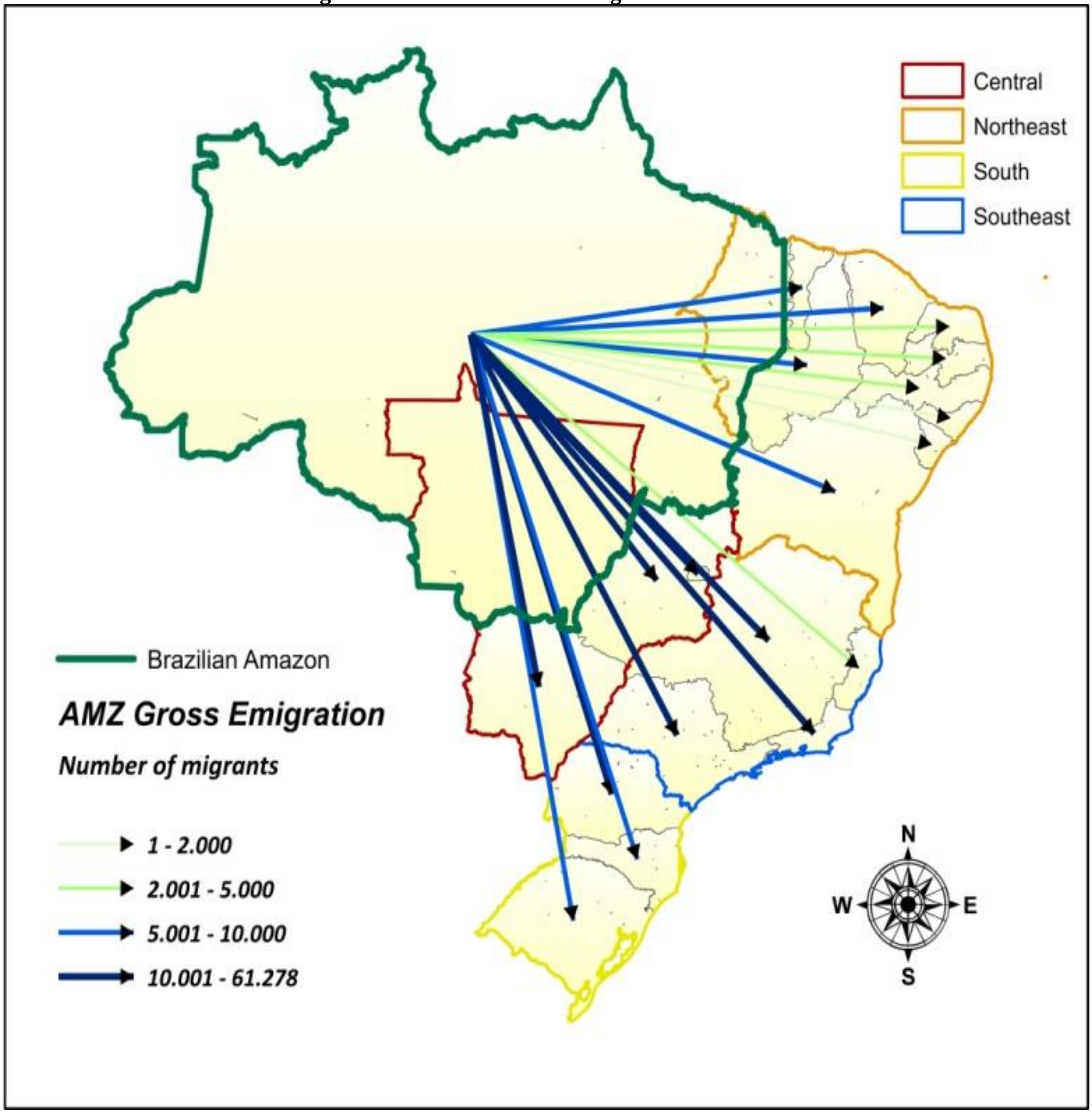
SOURCE: IBGE 2010 Census data, Own Elaboration

Figure 4. 1: Amazon Gross Immigration Flows



SOURCE: IBGE Census data, Own Elaboration

Figure 4. 2: Amazon Gross Emigration Flows



SOURCE: IBGE Census data, Own Elaboration

Nevertheless, this does not mean that migration flows do not play an important role regarding the process of urbanization, or even in demographic and economic terms. This is so especially because of two main reasons: first, migration flows usually tend to change the average population characteristics in the migrants' origins and destinations, due to social, cultural, economic and other differences which they may pursue. The clearest example would be the argument commonly found in economic literature[29], by which migrants tend to be more

[29] Which we further discuss and review in the next sections.

highly skilled than the average local populations of non-migrants. Second, it is possible (and even likely) that immigrants and emigrants pursue different goals and preferences when it comes to the main drivers of their migration decisions (as our results will confirm for the case of Brazilian Amazon, in the following sections of this chapter). In this case, and if we are able to identify those differences, then it is possible to build perspectives regarding the future relevance of migration flows as a part of the whole demographic changes occurring within the region, and also to build perspectives regarding the future trend of the migration flows themselves, along with and their role in terms of local occupation.

Further analysis of the Brazilian Amazon migration flows bring a few more interesting insights, which highlights the importance of studying such flows due to the demographic changes that those might bring to the region. For instance, it is possible to notice from tables 4.1 to 4.4, ad also from figures 4.1 and 4.2, that Amazon immigration flows are quite different from its' emigration flows, in terms of the main origins and destinations of the migrant population. In this sense, the main difference regards the Northeast and the Central macro regions of Brazil. While the Southeast and the South regions are responsible for about 30% and 15% of both immigration and emigration flows of Brazilian Amazon, respectively, The Northeast region presents itself as the main supplier of immigrants to Brazilian Amazon, providing 31% of these, even though it is the destination of only 20% of emigrants who leave the region. Economically, this brings an important message: the Northeast region is currently the poorest macroregion of the country, even poorer than Amazon itself, which comes in the second position in such negative aspect. Thus, this means that a great part of Amazon immigrants still come from the poorest region of Brazil, on one hand, whereas the emigrants who leave Amazon also tend to move towards other richer areas. This aspect, by itself, may be interpreted as first evidence that the search for higher levels of real income, as predicted by theoretical models (which we further describe in the following sections), apparently may apply for Brazilian Amazon immigrants, on average.

Oppositely to this context concerning the Northeast region as a supplier of immigrants, the main destination of Amazon emigrants is the Center-West region of Brazil, where the rich soybean belt is located at, hosting a share of 40% of these individuals' destinations, whereas population immigrating towards Brazilian Amazon coming from this macroregion responds for 25% of total immigration flows towards Brazilian Amazon, coming in third place among all Brazilian macro regions. As the Center-West region of Brazil pursue higher development indicators than does the Northeast region or the Brazilian Amazon itself, this corroborates the argument that even though liquid immigration to Brazilian Amazon seem to be diminishing, migration flows related to this region might play an important role in terms of impacting the socio-economic characteristics of its population.

Table 4.5 brings a comparison between some of the individual characteristics of Brazilian Amazon immigrants, its emigrants, and the average migrant of Brazil as a whole.

Table 4. 5: Migrants' profile of individual characteristics

	AMZ Imigrants	AMZ Emigrants	Brazil
Average* Age in 2005	33 (13)	32 (13)	34 (14)
Caucasian	46%	41%	50%
Black	7%	8%	8%
Oriental	2%	1%	1%
Mullatto	45%	49%	40%
Indian	0%	1%	0%
Women	46%	50%	50%
Men	54%	50%	50%
Born in Destination city	7%	13%	11%

*Standard Deviations in Brackets

SOURCE: IBGE 2010 Census data, Own Elaboration

As we can see, a few differences appear between these three kinds of migrants, regarding their average individual characteristics. The first one is that Amazon immigrants differ slightly in terms of gender, when compared to both Brazilian average migrant and the Amazon emigrant. The majority of Brazilian Amazon immigrants is composed by men, while for the Brazilian Amazon emigrants and for the average migrant in Brazil, the shares of men and women are rigorously equal.

A second interesting fact evidenced in table 4.5 is that return migration seems to be less frequent between Brazilian Amazon immigrants than between Brazilian Amazon emigrants: on average, Amazon's emigrants who were born at their destination municipalities are almost two times more frequent than Amazon's immigrants. This may be interpreted as a sign that Amazon population might be changing more than the population in the rest of the country, in terms of its characteristics, since its new comers more often come from different cities. In turn, this may also be interpreted as evidence in favor of the argument that even tough migration flows shall not be considered the main driver of the recent growth in local population, on the one hand, it is probably contributing to change the set of average characteristics of the region's residents.

4.1.3 Specific objectives and Structure

Given this current context of rapid local population increase based on vegetative growth, along with urbanization of the Brazilian Amazon, the specific goal of this chapter is to find evidence on which are the main determinants of immigration flows from other parts of Brazil towards Brazilian Amazon, comparing those to the causes of emigration flows from Brazilian Amazon towards the rest of Brazil, as these flows might play an important role in terms of determining future trends of economic, social and cultural characteristics of local population. In order to do so, we divide this chapter in 6 parts, with the first one being this introduction. In the next section, we review a few theoretical models regarding migration. In section 3 we present the methodology implemented. Section 4 brings a brief description of the database, section 5 contains the descriptive analysis along with the econometric results. Finally, section 6 concludes.

4.2. Literature review and discussion

In this section we review the literature regarding migration, focusing on determining which are the main variables we must try to include in our empirical modeling of migration determinants, and what are the main empirical problems we have to deal with in order to make our estimations robust.

4.2.1 Theoretical models on migration determinants

Migration is an issue debated from different perspectives, configuring the subject as an interdisciplinary matter. Geographers, economists, historians, sociologists and even biologists developed different approaches to try to develop answers to the same question: what determines migration flows? In the following, we sum up the main groups of models which support the evidence found in the empirical sections of this chapter.

Neoclassic Economic models: migration as a response to real income and employment differentials

The Neoclassical economic models treat migration through the same kind of approach from neoclassical models which do not involve space: from an atomistic perspective. To them, migration is a result of individual choices based on preferences, in which decision is made through a cost-benefit analysis of destinations' and origins' characteristics, and their socio-economic differences. In other words, each individual chooses whether to migrate or not, and where to migrate to, trying to maximize his own well-being. Examples of this kind of modeling are found in Todaro (1969), and Sjastaad (1962). The majority of these models consider the real income differential between origins and destinations as the main interest variable to be analyzed, due to one reason: like in most neoclassical economic modeling, the individual's well being is maximized through consumption of goods and services, which by their turn are acquired through income. Thus, migrants tend to maximize their expected real income choosing the location where they are going to live and work. As dos Santos et al. (2005) point out, the individual's skills and the city's capacity of offering jobs which match the need of these are also relevant, but still through an income maximization perspective: individuals will seek to live in places which offer conditions for them to maximize their productivity, which in equilibrium, will lead them to the maximum real income available given their skills' endowment. In empirical terms, this justifies the addition of other municipalities' characteristics (besides real income) as explanatory variables of migration, such as the city's infrastructure level, for example.

Other models, as in Harris & Todaro (1970) consider the comparison of the employment level in origins and destinations as important as real income to the individual, considering that even if the income difference does not compensate, it is possible that the migrant might take into account the probability of finding a new job in his destination as the main decision factor.

Another group of models with this neoclassical perspective, but with slightly different assumptions, is the group of "micro founded" macroeconomic growth models (Solow, 1956; Barro & Sala-i-Martin, 1995). In these models, individuals are treated as the labor force, and thus, migration represents the mobility of the labor factor in the economy production function, and is not decided at the individual level, but occurs through a simple functional form which responds to real income as the control variable. Still, this functional form is also based on the real wage differential between regions, with migrations flows going towards places and firms which offer higher real income to workers. More recent models of this school (Romer, 1990) have built a new set of hypothesis, and have introduced new variables such as technology and learning as determinants of income and migration. Still, the main logic of migration as a result of income and employment differentials remains.

New Economic Geography and Urban Economic Models: migration based on real wages and employment, with the addition of other regional differentials

The group of models from the New Economic Geography along with Urban Economics authors are also based on the neoclassical tradition, that is, their basis rely on individual choices of maximization which result in migration patterns and other equilibrium results. The main characteristic they share, to what concerns our goals here, is the fact that migration is once again mainly determined by real income and employment differentials. However, these models also include the importance of many other regional aspects into the analysis, such as urbanization, transportation costs, increasing returns to scale, and geographic differentials, justifying the inclusion of other variables in empirical analysis.

Krugman (1990) and Fujita et al. (1999), for example, in their center-periphery models, treat migration as crucial for understanding the dynamics of urban growth. In their models, migration flows are determined by the mobility of labor, by an equation in which workers migrate from one city to another in a myopic[30] search for higher real wages. These wages, in turn, respond to this dynamics, rising and falling according to the labor supply, which in turn is precisely determined by the arrival and departure of employees. The cities, in turn, tend to grow or decrease according to these migration flows, which clearly are involved with the local economic dynamics.

Similarly, Gleaser et al. (1995) construct (and test empirically) a model in which cities are centers of workers and free capital mobility, and that the growth of these cities is explained by migration of workers seeking two goals: higher real wages, better quality of life. This second goal is one different aspect from other models, and is justified by the authors' empirical findings, which supports the inclusion of variables which try to capture these aspects in any empirical analysis.

Following similar strategies, Urban Economics models add another aspect of migration determinants: they include special movements of companies, which in turn affect the dynamics of urban growth by determining the equilibrium of wages, which ends to attract or disperse workers, thus also affecting migration. Still under the neoclassical tradition, these firms migrate in search of opportunities to maximize their profit, either by reducing

[30] These searches are considered myopic because the worker does not take into consideration, in these models, the fact that his migration decisions affect the labor market, and thus, wages are going to respond to his migration decision, making wages observed in equilibrium to differ from the income that the worker was expecting at the time he decided to migrate.

transportation costs, or by the possibility of incorporating new technologies available in the urban destination (Duranton & Puga, 2001). In this sense, labor markets, transportation costs, and the possibility of innovations jointly determine agglomerations and urbanization, which in turn attract or disperse waves of migration to cities, in a complex feedback dynamic between all these variables (see Duranton, 2007; Findeisen & Sudekum 2008; Igliori, Abramovay & Castelani, 2012).

Human Capital theory: the role of education and training in migration

The Human Capital Theory disposes of the same methodological structure from the previous theories presented so far: migration flows are explained by individual behavior oriented by cost-benefit analysis. However, differently from the neoclassical and spatial economics theories, the individual's choice of migration is now considered an inter-temporal matter, in which personal the level of education and training influences heavily on this decision. Becker (1993) builds a model in which individuals decide whether to migrate or not, and also where to move to, according to the level of education that this individual pursue compared with the average level of training and education that each of his possible destinations require. More specifically, the model is based on a matching criterion, in which the migrant rationally tries to match his own level of accumulated human capital with the level offered by the available destinations, in order to choose the destination which may provide him the higher present value of future benefits.

Therefore, these models may also be interpreted as based on individuals' choices to maximize real income, but in an inter-temporal dimension. Nevertheless, they new important elements to the analysis: regional differences in labor markets regarding their demands for different levels of skills and training; and the migrants' decision and capacity to accumulate human capital are also introduced with a very important role. Empirically, these models imply that along with the current level (at the migration date) of real income in destinations and origins, the educational level of these places must also be included as explanatory variables in regressions explaining migration flows. Borjas (1989) gives a clear example of this importance. According to the author, given the level of human capital of the migrant and the labor market demand for this capital at destination, it is likely that this individual will not necessarily choose the destination which provides him the highest immediate level of income: in many situations (depending on the individuals' characteristics and risk aversion) individuals prefer to choose destinations which may provide him with good future prospects of income, even in cases which such destination is not the one that offers him the highest real income at the present date. Of course, such future perspectives also depend on the level of human capital that this individual pursue, once this level affects heavily his capacity of earning higher levels of future income in destinations which attend his matching criterion.

It is important to notice one subtlety of this argument. In these models, the decision of migration made by individuals with a higher level of education should not be interpreted as evidence that these migrants are seeking to raise their own education levels. Instead, it represents that these individuals exhibit a higher propensity to search for more specialized labor markets in their destinations, the ones which may fit their already accumulated human capital level. In this sense, destinations with a relatively high share of population with college degree, for example, might be seen by migrants as regions with a relatively high specialization level of their labor markets, and thus, as destinations where a highly specialized worker might fit his greater human capital level more easily. Gary Becker (1993) highlights this argument: the author argues that in these studies, the immigrant's income, at the moment of migration, will be probably lower than local average income within the destination's labor market, as the migrant still does not pursue some of the skills needed to fit himself at his new labor environment, such as a good mastery of local culture or language. As time passes, the share of immigrants who have invested in a high level of human capital tend to work harder than average in order to increase their salaries, motivated by the need to achieve a level of income return consistent with its past investment in education. In turn, this makes them overcome the average wage of local population at destination in the long term.

Moreover, one corollary of this theory is that population with higher degree of education tend to be more mobile than the average person. The reason is that once an individual made his investment in education, he tends to be more leaned towards migrating, if such movement is necessary to pay off this investment. As technological centers tend to be agglomerated in a limited number of counties and, mainly due to the Marshallian agglomeration externalities (as see Fingleton et al. 2005), there is a tendency that these individuals will migrate towards these centers at some point.

The New Economist of Labor Migration: migration as a group (family) decision of risk minimization.

The school of New Economists of labor migration change the basis of migration analysis from an individual decision to a group decision (Taylor, 1986; Mincer, 1978). According to the authors, in many cases, migration decision is taken by a group of individuals who share a connection between themselves. A family deciding whether or not to move to a new city would be the most obvious example, in which its members have strong laces of connection, and thus tend to discuss important long run decisions such as migration (Harbinson, 1981). They defend that such decision is usually taken not necessarily in order to maximize present and future income, but in order to minimize costs and risks. Taylor (1986), for example, argues that the existence of asymmetric information and uncertainty related to labor market drive individuals to make the migration decision along with other members of their domiciles and family, thus minimizing the risks of a drop in their quality of life through sharing the risk that, for example, one isolated member does not find a new job at the destination, or even that one member of the family finds a job which pays less than the wage he used receive in their origin.

To what concerns this work, we consider that this school of modeling complements the analysis based only on individual decisions: it brings two new elements into the analysis, which we take into consideration in the empirical session. The first one is the fact that any empirical analysis which is made only at the individual level, that is, with each individual being the unit of observation, probably will not be able to capture these effects of group decision, since individuals in the same family or domicile may differ in their characteristics and preferences. In this sense, working with migration flows along with individual decisions, which is our strategy here, somehow captures this group effect, as we will further discuss later.

The second element is that risk mitigation must be added to the explanatory variables, as this is the crucial economic point of these models. In order to do so, in our econometric models, we have included variables representing the income distribution level of origins and destinations, based on the fact that better income distribution in one city is probably positively correlated with lower risk of earning lower levels of income. Complementarily, we have included other non-monetary variables, such as access to sanitary treatment, which somehow represents a minimum level of quality of life for individuals which may find themselves temporarily unemployed or earning an extremely low income. These “distribution” variables endue a few econometric problems which we discuss later. However, given the theoretical importance of including risk-variables described so far, we have chosen to include those in our empirical models, even with those possible econometric issues.

Structuralist models: reinforcing the role of urbanization and economy dynamics

The approach of structuralist models regarding the determinants of migration is basically given by an effort to understand the complex interdependence between several variables that may affect migration decisions in the period considered. Unlike most schools guided by the neoclassical logic, these models do not try to isolate the effects of each variable, but rather to understand how the interdependencies between each of them affects migration flows as a whole.

Among these complexities, authors of this school as Singer (1976) emphasize that, besides the main push-pull factors traditionally analyzed separately, such as income and employment levels at origins and destinations, empirical modeling should also take into account historical and structural characteristics of these origins and destinations, such as their degree of industrialization and urbanization, which may also act as push-pull factors. In this sense, such theories are consistent with models of Urban Economics and New Economic Geography, highlighting the relevance of real wages as a determinant of migration, but at the same time conditioning the future dynamic growth of these wages to the level of economic development of the municipalities, which in turn is linked to their urban, industrial and technological dynamism (Krugman, 1991; Fujita et al. 1999; Duranton, 2007; Findeisen & Sudekum, 2008; Duranton & Puga, 2001; Gleaser et al, 1995).

An example of a model that can be fit as a mix of a “macro neoclassical” approach with a NEG model, thus resulting in a somewhat complex structuralist model, is the Matsuyama & Takahashi (1998) model, in which migration flows are determined by the choice of individuals (seen as workers) among cities according to their preferences towards the several aspects which each city has to offer. More specifically, in this model, individuals choose between municipalities according to a quality of life index, which by its turn is influenced by the city’s economical dynamism in terms of its number of specialized firms, its capacity to trade goods with other cities, the share of participation of the services sector, and the productivity of its workers (which determines real wages). Thus, the model results in several different equilibriums of demographic (population) distribution, with each equilibrium being determined by a complex relationship between several variables whose importance were already highlighted individually in most of the theories described so far.

Therefore, generally guided by these models, we have chosen to include in our econometric analysis some explanatory variables which capture the degree of urbanization and the population's mobility dynamics at origins and destinations, such as the share of employment in the tertiary sector, the percentage of urban population, the level of aging of the city's population, among others.

4.2.2 A brief literature review on migration empirical modeling

Given the theoretical background of the determinants of migration which will guide our analysis, we now turn our focus to literature contributions in terms of the problems that arise when dealing with migration databases. Furthermore, we review a few studies which demonstrate the typical approach taken empirically when it comes to the study of migration through econometrical analysis.

The sample selection problem: a stylized fact for empirical studies of migration (Chiswick, 1999 & 1978; Borjas, 1994; dos Santos, Menezes & Ferreira, 2005; Sjastaad, 1962)

According to literature, the main problem concerning empirical modeling of migration involves the econometrical issue of sample selection. This point was firstly raised and well-detailed by Chiswick in 1978, and modeled in 1999 by the same author. In his model, Chiswick (1999) follows Sjaastad (1962) central point in which migration is regarded as an investment which may raise labor productivity, in which the individual decides whether or not to migrate based its costs and benefits. Benefits are measured in terms of the real wage differential that the migrant may obtain by migrating, along with other non monetary benefits regarding his preferences, such as preferring specific weather conditions at the destination instead of the average origin's climate. Costs also regard all monetary and non-monetary transactions involved in migration: opportunity costs, the costs of leaving family and friends, costs of adaptation, costs of finding a new job, etc. In this sense, the liquid return of migration (*r*) for one average individual is given by equation 4.1.

$$r = \frac{B_d - B_o}{C_{nm} + C_m} \tag{4.1}$$

where B_d are the benefits which the migrant may obtain at his possible destinations; B_o are the benefits which the migrant obtain by staying in his origin; C_{nm} are the non-monetary opportunity costs of migration; and C_m are the monetary costs of migration. In this model, it is assumed that there are two types of individuals in the economy: the high skilled individuals (*h*), and the ones with low skills (*l*). By skill, the author refers to many unobservable characteristics, such as willingness to work, intelligence, capacity of adaptation to new situations, etc. It is assumed that each high skilled individual receives k times more benefits than the low skilled ones in both origins and destinations, which means that $B_{d,h} = (1+k)B_{d,l}$ and $B_{o,h} = (1+k)B_{o,l}$, with $k \geq 0$. It is also assumed that the monetary costs of migrating are the same for both kind of individuals, that is $C_{m,h} = C_{m,l}$. On the other hand, as the high skilled worker tend to earn a higher income than the low skilled worker, independently of where he lives, then the opportunity cost of leaving the origin is also higher for the high skilled worker. For simplicity reasons, without loss of generality, it is assumed that this opportunity costs difference for the two types of individuals is also given by the rate *k*. Thus, $C_{nm,h} = (1+k)C_{nm,l}$. Therefore, the liquid return of migration for the high and the low skilled workers are, respectively, given by equations 4.2 and 4.3.

$$r_h = \frac{(1+k)B_{d,l} - (1+k)B_{o,l}}{(1+k)C_{nm,l} + C_m} = \frac{B_{d,l} - B_{o,l}}{C_{nm,l} + \dfrac{C_m}{(1+k)}} \tag{4.2}$$

$$r_l = \frac{B_{d,l} - B_{o,l}}{C_{nm,l} + C_m} \tag{4.3}$$

Under these conditions, $r_h \geq r_l$ because $k \geq 0$. Thus, as the liquid rate of return of migration to high skilled workers is higher than the liquid rate of return to the low skilled one, this model specifications imply that the high skilled worker has a greater incentive to migrate than the low skilled do. More specifically, two hypothesis are necessary for the conditions of this model to hold, therefore, for this implication of higher incentives to the high skilled worker to be true: 1) it is needed that labor markets indeed tend to pay more workers with higher skills, and 2) monetary costs of migration exist, and are positive. These assumptions are considered to be realistic and are mostly also assumed in any neoclassical which assumes that higher real income is due to higher labor productivity, and also that transportation costs are positive[31].

[31] This is true in almost any model which considers location as one of the decision variables, that is, when distance is taken into consideration, even under neoclassic conditions, transportation costs are considered to exist and be positive.

Thus, this model implies that there is a positive selection when it comes to studies of migration flows: migration flows are composed only by migrants, which in turn tend to be more highly skilled than the average population, since non-migrants are mostly composed by low skilled workers. If the researcher is interested in investigating the migration decision of any individual among a population, migrant or not, this turns out to be a problem, because traditional estimation methods which do not take this selection problem into consideration will produce biased results. In other words, if what interests the researcher is to study the effects of different variables on the migration decision of any random individual of the population, being this individual a future migrant or not, then this positive selection matter must be taken into consideration, otherwise estimators will be consistent only for the share of the population composed by only potential migrants. In simple terms, this means that if this sample selection problem is not considered by estimation procedures, results will be valid only for the share of population which already pursue a higher propensity towards migrating, due to unobservable intrinsic characteristics. And as we further discuss in the methodology section, this is not the case in our study, since we wish to study what drives migration flows, and for such it is important to consider both migrants and non-migrants decisions, as we further discuss in the next sections.

Other theoretical models add different aspects to this one, always confirming this sample selection result. Borjas (1987) adds uncertainty by considering the expected return instead of a simple current liquid return, and Katz and Stark (1986) add asymmetric information to the model hypothesis. Both of them result in the same positive selection problem, along with the possibility of even increasing the bias caused by such econometric problem.

Sample selection in Brazilian migration flows

Since this theoretical problem emerged in literature, many empirical studies were produced to try to test the existence of such positive bias, in practice. Chiswick (1978), in the work that has inspired the model presented above, showed that immigrants in the US were positively selected when compared to the average American population. For Brazil, many studies corroborate this positive selection result. Among them, dos Santos et al. (2005) find evidence that within the decade of 1990, Brazilian workers who have migrated were also positively selected when compared to the local average workers. Freguglia (2007) confirms this result for the beginning of the 2000 decade in Brazil, using panel data estimation for formal workers.

Still regarding sample selection in Brazil, one important counterpoint must be raised. Even though results from recent literature (above) tend to confirm that brazilian migrants may compose a highly skilled population, some historical patterns of Brazilian migration might put such evidence in check. One may argue, for example, that the great flows of migration from the Northeast region to the Southeast, which happened in Brazil from the decades of 1970 to 1990, were mostly composed by low skilled workers who were seeking to escaping from declining economies of this region, looking for new job opportunities in more developed markets inside the country. Such argument would be based in the logic by which only the more skilled workers would manage to endure in their jobs in a lower and declining economic environment, which means that the low skilled ones would end up in fleeing towards other markets.

Nevertheless, even if such argument is true for our case, and no sample selection is present in our sample, to what concerns the methods and goals of this study, this does not imply in estimation problems or inconsistent econometric results. The reason for that is because in the methodology implemented here (Heckit estimators), which we further discuss and present in the next sections, the possible sample selection problem is treated as in Heckman (1979), that is, with the inclusion of one additional explanatory variable representing the average probability of migration. In turn, this means that if sample selection indeed occur for Brazilian migration flows, it is being taken in consideration and controlled, whereas if no sample selection is present, results are still unbiased, since we would be just adding one additional irrelevant explanatory variable, which implies in efficiency loss, but no bias or inconsistency in the parameter estimations (see Wooldridge, 2002, pp. 49-76) in robust Ordinary Least Squares (OLS) estimators, which are the ones we use in the second stage of our Heckit models.

General empirical results on the determinants of migration

Many studies on the determinants of migration flows exist on literature (see, for example, Crozet, 2004; LeSage and Pace, 2005 and 2008; DaMata, 2007), especially for the developed countries. Even though only some of them control for the positive selection problem explored above, results tend to converge towards the corroboration of the majority of the theoretical models revised in this chapter. Crozet (2004), for example, study migration flows within a few European countries, such as Germany, Italy and Spain. In general, they find that cities where the share of employment in the tertiary sector is higher tend to attract more migrants, which they interpret as higher levels of urbanization in destinations working as a push factor for migration flows. Moreover, they argue that this result may also be interpreted as a Jacobs externality evidence: as more urbanized centers are also the wealthiest, this may be considered an evidence that migrants in these countries tend to seek higher levels of real income when they decide to migrate.

LeSage and Pace (2005 and 2008) develop a new estimation procedure, in which they incorporate new spatial elements in regression analysis of migration flows. Specifically, the authors argue that migration flows must be treated as an spatial variable, meaning that not only variables within origins and destinations may interfere on the migrant's decision, but also these variables at the neighbors of these origins and destinations may affect the migration flows between them. The authors argue that these neighbor effects may be relevant because municipalities tend to be a part of an economic inter-connected system of cities, in which, for example, the economy of one municipality might be dependent on the economy of one of his bigger neighbor municipality. This would be the case, for example, when individuals use neighbor cities as dormitory municipalities, in which they work in one city, and live within one of its neighbors. Another example would be the neighbor effects of a capital city: migrants who decide to migrate towards one capital's neighbor usually consider the capital's economic structure and conjecture as important elements to be considered. Bearing these arguments in mind, the authors develop a new spatial econometric procedure in order to incorporate both origins and destinations neighbors effects. The development of such new procedure was necessary because the spatial neighbors' weights matrix needed to accomplish two types of neighboring effects (in destinations and in origins) is much larger than in standard spatial econometric models.

Further explaining this point: by "standard", we mean that usually the municipality itself is the observation unit of a study in spatial models whose variables are observed by municipality. In this case, the size of the spatial neighbors matrix is *nxn*, being *n* the number of municipalities in which these variables are observed. This is what we call a "traditional" approach in Spatial Econometrics, as most of the studies in this field follow this structure (see Elhorst, 2003). On the other hand, in the methodology developed by LeSage and Pace (2005 and 2008), as they consider both origin's and destination's neighbors, their unit of observation is now the migration flow between two cities, even though explanatory variables are observed at the municipality level, and afterwards divided between destinations and origins. As each origin city present positive flows for more than one destination, then their matrix size tend to be much larger than *nxn*, with n still being the number of cities in the sample. For example, if we imagine that each municipality share migration flows with every other city in the sample, then we will have n^2 observations within the sample, since the migration flows are the units of observation. If we assume that any flow between two regions may be a neighbor to any other flow between any other regions, this would require a n^2xn^2 spatial weights matrix, in order to implement spatial econometric models such as a SAR (Spatial Autoregressive Model) or a SEM (Spatial Error Model)[32]. However, it is important to note that the explanatory variables are still observed only once per municipality, although they appear two times in each observation unit: one for the destination, and one for the origin. Given this dimensional issue, the authors (LeSage and Pace, 2005 and 2008) develop a specific method in order to build this such "augmented" matrix, and also in order to implement SAR, SEM and SAC (Spatial Autocorrelation Model) models consistently under this complex sample size matter. Further details of this specific methodology is discussed in the next sections of this chapter, in which we implement part of this method in one of our estimation approaches.

In order to test their new methodology, the authors implement an exercise of applying their new estimator to US county data, including explanatory variables such as real income and the city's educational level in both origin and destination counties of migrants, as well as in the origin's and destination's neighbors. Also, these models include the migration flows from the origin's neighbors and towards the destination neighbors as explanatory variables. Their results also corroborate most of the results from the theoretical models of migration: migration flows are higher towards cities with lower unemployment and costs of living[33], and higher per capita income. Moreover, the authors find that cities whose population exhibit higher degrees of education are also the ones whose population pursue greater mobility (in terms of greater migration flows), as expected by human capital theory. Also, younger people tend to migrate more frequently, according to their findings, as they find that higher the share of population with 22-29 years in origins and destinations, the larger the migration flows between them. This result, argue the authors, is also expected, as young people tend to hold a greater incentive to migrate. Also, they find that individuals near retirement are usually less leaned towards migrating, which is also considered expected by the authors, due to the

[32] For the specific definitions of these models, see LeSage, 2008.
[33] Which, as we will see later on, are represented by the average rent within each municipality.

argument that this group of people tend to be less willing to change their jobs and lifestyle right before they reach retirement, according to a risk minimization logic.

Specifically for Brazil, DaMata (2007), who try to investigate the determinants of migration flows of individuals with high levels of education between 1995 and 2000 all over Brazil, using census data. The author finds evidence that this subpopulation is statistically attracted to municipalities with higher real income, higher levels of education, lower inequality, and other urban amenities. Thus, most of the theoretical models presented in this section are corroborated by their evidence. It is important to notice that the authors do not try to correct for positive selection bias in this work. However, on the one hand, if their interest population is specifically the group of individuals with higher levels of education, selection is not a problem in their empirical strategy, as non-migrants behavior is not encompassed in their focus. In this sense, their result may be interpreted as: once one highly educated individual has already decided to migrate, he tend to choose destination cities with higher levels of income, education, urban amenities and equality. On the other hand, if they were interested in studying all highly educated individuals' (non-migrants and migrants) decision whether or not to migrate, then these results may be considered biased due to the presence of selection bias, as previously discussed. Moreover, these authors also include spatial dependence in their explanatory variables' vector. However, differently from LeSage and Pace (2005 and 2008), only the "traditional" spatial econometrics models are approach is implemented, as only explanatory variables of the destination are included in the model, which allow them to use a "traditional" neighbors' weights matrix regarding only the destination municipalities as neighbors.

4.3. Empirical modeling: methodology

In this section we describe the methodology which we have chosen to implement in order to achieve the goals of this study. We begin by describing and discussing the general approach which permeates all methodological strategies adopted here. Next, we describe how we have addressed the sample selection problem which may arise in our econometric estimations, as previously discussed. Then, we describe in detail the specific econometric strategy which we have implemented, and then conclude this section describing the database sources and resuming a few other advantages and limitations from our approach.

4.3.1. General approach: the push-pull analysis

In this study, we adopt the general strategy which is referred as the "push-pull analysis" by empirical literature on migration. As Golgher, Rosa and Araújo (2005) describe, the push-pull method consists of interpreting migration flows and decisions as a result of many different aspects and characteristics regarding both the origins and destinations of these migrants. In this sense, the aim of this approach is to find evidence on which are the average socio-economic, demographic, political, and other characteristics of the migrants' origins which disperse (push) them away, and what are the average characteristics of the migrants' destinations which attract (pull) them to their destinations.

According to this method, if we split different groups of migrants coming from different places, or moving towards different locations, it is possible to find evidence that these groups may behave equally regarding the level and performance of some explanatory variables in their origins and destinations, and differently regarding other variables. It is possible, for example, that individuals moving from A to B exhibit an average preference of heading towards places which offer higher real wages, independently of the educational level of any of these places, whereas individuals moving from B to A exhibit opposite average preferences, not caring about current real income differentials, but seeking markets whose educational levels may match their previous human capital investments.

In other words, this strategy consists on dividing migrants into different groups according to their destination and origins, and then try to measure how each of the explanatory variables on these origins and destinations may affect each group's migration flows, on average, always comparing these differences between the explanatory variables' effects for each group.

In our case, specifically, we have divided migration flows between Brazilian Amazon and the Rest of Brazil in two groups: the group of Brazilian Amazon immigrants, composed by all individuals with more than 18 years in 2005 whose origin municipality was located outside Brazilian Amazon borders (but within Brazilian frontiers), and whose destination was located within Brazilian Amazon; and the group of Brazilian Amazon emigrants, composed by all individuals with more than 18 years whose origin municipality was located inside Brazilian Amazon, and whose destination was located outside Brazilian Amazon borders (but within Brazil).

The main reason why we have chosen such division is that by dividing our sample in this way, it is possible to investigate empirically which variables that influence individuals to move from the Rest of Brazil to Brazilian Amazon, and what variables may influence individuals to leave Brazilian Amazon, moving towards the Rest of Brazil, to then compare these results in order to draw a scenario of what characteristics attract (or expel) individuals to (from) Brazilian Amazon recently. Such comparison is essential to understand part of the demographic transition which Brazilian Amazon has been passing through recently, since such migration movements tend to be closely related to urbanization and demographic growth within one region.

A few important considerations must be made concerning such general strategy, especially regarding the database of migration flows which have been used. The first one concerns the type of migration which we work with in all regressions: the fixed date migration. Details regarding this point was already discussed in the first section of this chapter. Still, given that the migration flows which compose our sample all refer to the period between 2005 and 2010, another important remark is that, in order to try avoid simultaneity bias, all exogenous right-hand side variables in our regressions refer to years prior to 2005. Further explaining this point, migration flows may affect several economic variables of both origin and destination municipalities. When individuals move from one city to another, they cause downward shifts in the labor supply curve of their origin, and an upward shift in their destination's labor supply curve. They also affect the overall rate of labor productivity of both origin and destination, since according to the literature reviewed here, migrants tend to be more highly skilled than the average workers who simply remained in their cities. Moreover, these migrants are new inhabitants in their destination cities, therefore, they increase the demand for goods and services as a whole. These are only examples of how migration may cause shifts in economic variables. As our main goal is to determinate the exact opposite, that is, how these economic variables determine migration flows, we wish to avoid this reverse causality matter.

In this sense, we chose to use only explanatory variables which refer to periods prior to those concerning migration flows, in an inter-temporal attempt to reduce this kind of bias, since migration flows may affect economic variables only afterwards they have occurred. That is, by the fact that the set of explanatory variables have occurred prior to the migration flows considered as dependent variables, it is less likely that such flows might cause these explanatory variables. We are aware that such strategy does not account for all possible sources of reverse causality, nevertheless, it is one method commonly found in economic literature (see, for example, Fingleton 2003, Gleaser et al. 1995 and Barro & Sala-i-Martin 1992).

Thus, the majority of the explanatory variables included in the regression refer to the year 2000, mostly obtained in the IBGE Census database, with only 5 exceptions: costs of living refer mostly to the year of 2005 (as we further explain in the following sections); the number of academics in each municipality and the IFDM (Firjian Index of Municipal Development), which also refer to the year of 2005; and the immigration flows and the income growth rate between 1991 and 2000 which were collected in the 1991 and 2000 IBGE demographic Census.

Another important consideration we must make is the choice of excluding from the sample any individual with less than 18 years in 2005. The reason for such was in order to minimize the number of individuals who do not endow the power to decide whether to migrate or not, but instead must follow their family decision. Our criterion for the age of 18 is based on the fact that this is the legal majority age in Brazil, and thus, at least institutionally, individuals older than that are free to decide where to live by themselves. Additionally, we also excluded from the sample any migrants who came from, or moved to another country, for three reasons: 1) data availability of these individuals' origins and destination municipalities (which is the geographic level we work with) would be scarce and difficult to find; 2) it is reasonable to believe that individuals who migrate towards other countries exhibit persistently different preferences and goals than intra-country migrants, especially concerning their motivations to migrate, and also the information they pursue about their destinations; 3) These type of individuals represent less than 5% of total migration related to the Amazon region in our database, which means that they may be considered as a residual part of all the region's migration process.

Finally, the most important restriction that Census migration database imposed on our general methodology is that it contains no information whether the individual came from an urban or a rural area of his origin municipality. This, in turn, prevents us from implementing an analysis of rural exodus within Brazilian Amazon, which would be essential in order to comprehend in deeper detail a different aspect of the urbanization process itself. Nonetheless, in order to capture at least a partial effect of the effects of urbanization on migration, as well as to follow the theoretical literature reviews in prior sections, we have included variables which represent such urban sprawl among the explanatory variables, and as we will see later, results proved to be quite significant.

4.3.2. Dealing with the sample selection problem

As seen in the literature review section, sample selection is a matter commonly found in regressions where migration flows are treated as the dependent variable. Thus, in this section we describe the strategy by which we have dealt with this issue, in our regressions.

The general structural form of the regressions we want to estimate assume the form of equation 4.4.

$$y_{od} = \alpha + \beta_o X_o + \beta_d X_d + \gamma D + \varepsilon \qquad (4.4)$$

where y_{od} represent the *nx1* vector of migration flows (with *n* being the number of flows) between origin and destination cities of the migrants in our sample, X_o represents the matrix of characteristics of the origin city; X_d represents the matrix of characteristics of the destination city; and D represents the matrix of specific exogenous characteristics of the flow between the origin and destination (such as distance between these two cities; and ε is the error term, for which we assume (for now) that $\varepsilon \sim N(0, \sigma^2)$.

Given such structural form, there is an important subtlety we must comprehend regarding which is the population of interest that our empirical methodology addresses: in order to achieve the specific goals of our study, the population of interest which we have to study may not be composed only by migrants, but instead must be composed by every individual, migrant or not, that could possibly migrate from one municipality to another, even if this individual decides not to migrate. That is, the population of interest is composed by both migrants or non-migrants.

To see that, it is important to keep in mind that we wish to study what are the determinants of migration flows (our endogenous variable), and migration flows are influenced by individuals' decisions to migrate, as well as individuals' decisions of not to migrate. An example helps to illustrate this point: suppose that every individual in one municipality decides not to migrate during the period examined. In this case, migration flows departing from this city would be zero towards any destination. In this sense, the non-migrant population affects directly migration flows by deciding not to migrate. Thus, in order to study migration flows as a dependent variable, we must consider both migrants and non-migrants' decisions as determinants of these flows, which is the same that considering both migrants and non-migrants as the population of interest.

However, as already mentioned, migration flows are, by definition, composed only by migrants, which means that they only represent one part of the population of interest., Or, in technical terms, our endogenous variable *y* (the flows of migrants) is composed by migrants only, therefore, it does not account for a part of population who also affect the size of the migration flow (thus affecting the interest population), which is the non-migrants population. If migrants and non-migrants were statistically similar enough in terms of their intrinsic characteristics related to their migration decisions, then one could argue that migrants represent a random subsample of the whole population of interest, and no sample selection problem would arise. But as seen in the literature review section, both theory and empirical evidence found in studies regarding Brazil and/or other countries' migration point out that migrants and non-migrants are mostly considered to be different in terms of such individual characteristics. As already described, migrants are considered to be more highly skilled than non-migrants, and such difference in skills is what determines the migration decision itself. Thus, if this theory applies to the case of Amazon migration, then the migrants who compose our migration flows exhibit characteristics different from the population of interest of our analysis (composed by migrants and non-migrants, who determine altogether the migration flows).

In fact, Wooldridge (2002, pp. 552) gives an example of the sample selection problem as the one considered by Heckman (1979) which is very similar to the case of migration: in trying to estimate the wage offer equation for people at the working age, the author argues that there is a sample selection problem that must be considered, because it is it is impossible to observe the wages (the dependent variable) of the share of the working age population who is not actually working, and these populations differ in terms of their individual characteristics. In parallel, in our case, we want to study what determines decisions whether to migrate or not, but migration flows (our dependent variable), by being compose only be migrants, do not include the decision of not to migrate (taken by non-migrants), which is the same as not observing such decision. As migrants and non-migrants may differ in terms of their characteristics, a sample selection problem arises, as in the example considered by Wooldridge (2002, pp. 552).

At a first glimpse, one tempting strategy to deal with this problem would be to simply include the total population of both origin and destination as two of the explanatory variables in the regressions, since these populations include both migrants and non-migrants population for each pair of cities. However, this strategy still does not account for the sample selection problem as a whole, because despite partially including non-migrants within the analysis, it does not capture the differences in preferences and skills between the migrant and the non-migrant population, which is the core of the sample selection problem in this case.

All these things considered, in order to deal with this sample selection problem, we chose to implement an estimator based on Heckman (1979), which literature mostly calls “Heckit” (see Wooldridge, 2002). The traditional procedure concerning such method, consists of firstly estimate a “selection equation”, which would be to estimate the probability of the selection problem to occur for each observation, according to a set of explanatory variables that might at least partially determine such selection process. In our case, this means to estimate the probability of each individual to migrate, according to his individual characteristics, because the selection problem is based on different individual characteristics between migrants and non-migrants. Once this selection equation is estimated, we use its estimated coefficients to calculate the estimated inverse Mills ratios of each observation, which consists of their predicted normalized probabilities by the estimated selection equation (for details, see Heckman 1979, and Wooldridge, 2002). Then, these inverse Mills ratios are inserted as an explanatory variable in the main regression to be estimated, which in our case is the migration flows regression, as in equation 4.4.

In the specific case of our study, however, two slight modifications had to be implemented in this traditional procedure, in order to adjust it to our database and to the specific case of migration flows. These adjustments are necessary to deal with the fact that Heckit estimator which we implement must consider that in each stage of the procedure, the units of observations are given in different scales: in the first stage, the regression used to estimate the probability of migration must be taken at the individual’s scale, as this probability depends on individual’s characteristics; whereas at the second stage, regressions use the migration flows between cities as the dependent variable, which means that the scale of the observations is given at the municipalities’ pair level.

In this sense, the adjustments which we have implemented to the Heckit procedure in this study were the following: first, we have estimated the inverse Mills ratios representing the probability of migration per individual, that is, we ran the Heckit first stage at the individual scale. Then, using these individual inverse Mills ratios, we have calculated the average inverse Mills ratio for each brazilian municipality, to then finally insert both Average inverse Mills ratios in the origin and in the destination municipality as explanatory variables of each flow of migration between a pair of cities within the Heckit second stage.

More specifically, we firstly estimated the Heckit first stage according to equation 4.5.

$$Pm_i = \gamma + \delta_i x_i + \delta_i^o x_i^o + \mu \qquad (4.5)$$

where Pm_i is a binary variable which equals 1 if the individual has migrated, and 0 if he has lived in the same municipality from 2005 to 2010. As in any probability models, the normalized predicted values of this variable for each individual may be interpreted as the probability of the individual *i* to be a migrant (see Wooldridge, 2002). γ is a constant term, x_i is a vector of individual characteristics, with δ_i being its respective vector of partial correlation coefficients between each individual characteristic and the probability of migration of the individual *i*. x_i^o is a vector of characteristics of the origin municipality of the individual *i*, with δ_i being its respective vector of partial correlation coefficients between each origin's characteristic and the probability of migration of individual *i*. Finally, μ is the error term, which is assumed to be independent of the explanatory variables.

Four important regards must be made in relation to the Heckit first stage given by equation 4.5. The first one is that this equation served as the first stage of all Heckit models concerning migration flows as the dependent variable estimated in this study. Second, this equation was estimated through a probit estimator, as recommended by Heckman (1979) and Wooldridge (2002). Third, the sample used to estimate equation 4.5 includes all brazilian population over 18 years in 2005, and no specific division regarding Brazilian Amazon was made at this point. The reason for that is because at the next step of our general approach, the results from this stage will be aggregated at the municipality level of all Brazilian cities, and specific inter-municipal divisions will only be made at the second stages of our modified Heckit procedures.

Fourth, the most delicate point of equation 4.5: one might notice that the characteristics of the individual's origin municipality were included in the equation as regressors, whereas the characteristics of the destination cities were not. The reason for that is: as this stage is estimated at the individual scale's level, and both migrants and non-migrants are included, it becomes impossible to include destination's variables for non-migrants, since they did not migrate, which means that the characteristics of his destination would be exactly the characteristics of his origin. This would cause perfect colinearity among regressors, and these variables would have to be dropped from estimations. Thus, it is only possible to include variables representing the characteristics of the origin municipalities. And this was done in order to control for fixed effects of these places, so that we could obtain consistent estimators.

The most important point of this first stage concerns the inclusion of the individual characteristics. Specifically, we have included the individual's age in 2005, his gender, the ethnic group to which he belongs, and a dummy representing the aspect of being born at the destination[34] city or not (in which the variable assumes the value 1 if the individual has born in the destination municipality, and zero otherwise). The role of these variables is to try to capture the effects, at least partially, of the possible differences in skills and preferences between migrants and non-migrants, or in other terms, to capture the drivers of the sample selection matter, to then include it on the second stage (in the form of the estimated probability to migrate), thus correcting the sample selection bias. As for the inclusion of the characteristics of the origin municipalities, the objective was simply to capture fixed effects from those cities that might influence the migrants decision, so that the coefficients of the individual characteristics might capture only the intrinsic unobservable characteristics related to the differences in skills and preferences of the migrant, which are the cause the sample selection problem.

One important limitation of the Heckit first stage estimated in this study is the fact that we could not include the educational level of each individual in 2005 as regressors, as this information were not available at census data for this year[35]. This is so because literature on the subject indicates that one individual's education may be highly correlated with his skills, in the sense that highly skilled individuals tend to have a greater incentive of paying the costs

[34] In the case of non-migrants, this dummy variable assumes the value 1 if the individual was born in the origin city, as the origin and destination municipalities are the same in this case.

[35] Even though the information of the individual's educational level was available for the year of 2010. However, we have chosen not to include the individual's educational level in 2010 as a regressor, because this variable refers to a period subsequent to the migration itself, which means that such educational level might have been caused by migration (for example, if the individual migrated to a certain destination in order to study at some University). In this sense, including the level of education of the individual in 2010 could have caused reverse causality endogeneity, which would generate biased estimators.

of acquiring a higher educational level, as these costs tend to be relatively lower to them (exactly due to their higher ability to learn. For theoretical literature on the subject, see the classic signaling model from Spence, 1973, and for empirical analysis on the subject, see Ashenfelter & Krueger, 1992; Bounjour et al, 2003; Miller et al, 1993). In order to try to compensate the absence of these variables, we have included at the x_i^o vector measures of the educational level of the individual's origin city, represented by its share of population with College degree, and also its share of illiterates among the population. Even though these are not ideal, they still might capture, at least partially the individual's educational level: individuals with high educational level are more likely to be found at cities with a higher share of population with College degree, and with a lower level of Illiteracy. Moreover, an extenuatory to this problem is the fact that the main goal of estimating this equation is not exactly to consistently obtain the coefficients of correlation of explanatory variables (of equation 4.5) with the probabilities of migration: in fact, the main goal of the first stage regression of the Heckit procedure is to at least partially capture intrinsic unobserved differences between migrants and non-migrants through the correlation of these differences (measured by their characteristics) with their willingness to migrate. This means, as Heckman (1979) and Wooldridge (2002) argue, that the only necessary condition of the first stage regression is that at least one among the variables within the first stage equation 4.5 must be correlated with the sample selection criterion (in our case, the individual's unobservable skills reflected by the individual's probability to migrate), with such variable not being included at the Heckit second stage. This, in turn, is a much weaker condition to attend to, since all individual characteristics were included only in the first stage of all our Heckit estimations.

Additionally to attending such important criterion, two other measures cited by Wooldridge (2002) shows that the first stage of our Heckit procedures was estimated robustly: as we will see in the next sections, our estimations of the individual's probability to migrate seem to produce a good fit on actual data, as the predicted probabilities of migration to individuals who actually migrated are statistically higher than the predicted probabilities to those who did not migrate, as it is to be expected. Moreover, these predicted probabilities measured as the average inverse Mills ratios at the second stage always proved to be significant in our second stage regressions, which implies that the sample selection problem was indeed an issue to be dealt with in our estimations, and it is also (at least partially) being corrected by the inclusion of these average inverse Mills ratios (Heckman, 1979). Also, the individual's characteristics included in equation 4.5 all proved to be statistically significant, indicating that probability of migration is at least partially explained by our first stage regression.

Continuing the detailing of the methodological strategy to deal with the sample selection problem, we further explain how the average Inverse Mills ratios were calculated. The next step towards such calculations, after equation 4.5 was estimated through a probit model (see Wooldridge, 2002), was to compute the predicted values for each individual given by the estimated coefficients of this model, to then calculate the inverse Mills ratio per individual. This is done so by firstly substituting the coefficients of equation 4.5 by the estimated coefficients of the probit equation, as in equation 4.6[36].

$$\hat{Pm}_i = \hat{\gamma} + \hat{\delta}_i x_i + \hat{\delta}_i^o x_i^o \tag{4.6}$$

Then, we input the actual values of x_i^o and x_i for each of the *i* individuals into this equation 4.6, thus generating the non-normalized predicted values, which we call (as in Wooldridge, 2002) $x\hat{\delta}$. Finally, we normalize each of these predicted values by applying the transformation $\lambda(x\hat{\delta}) = \phi(x\hat{\delta}) / \Phi(x\hat{\delta})$, where $\Phi(.)$ is the cumulative distribution function of the error term μ (which also corresponds to the cumulative distribution function of Pm), and $\phi(.)$ is the probability density function of μ (which also corresponds to the probability density function of Pm). $\lambda(x_i \hat{\delta})$ corresponds exactly to the estimated inverse Mills ratio of individual *i*.

Then, the next step was to calculate the averages of these individuals' inverse Mills ratios per municipality, weighting such averages by the number of individuals in each of these cities. These calculations have resulted in what we refer as the average inverse Mills ratio of each city, and as these are calculated at the municipal level, it is possible for us to include them for both origin and destination cities of each flow of migrants in the second stage regressions.

Given these average inverse Mills ratios per municipality, we then proceed to the second stage of the Hierarchical Heckit method, which in terms of implementing this method means to estimate the determinants of migration flows between pairs of municipalities including the

---

[36] By definition of the probit procedure, in which coefficients represent average estimated correlations, since $\mu \sim N(0, \sigma^2)$, the mean value of μ is zero, and therefore the error term becomes zero in the prediction equation 4.6.

average inverse Mills in destinations and origins among the explanatory variables (following Heckman, 1979). The reason why we have included both average inverse Mills ratios in destinations and origins, even though we have only used the origin's characteristics when determining the individual probability of migration (as explained above), is because the sample selection issue tends to occur in both destinations and origins, as it is given at the individual level. Thus, as migration flows occur between two cities, and the Heckit procedure is based on including an estimated probability of the selection to occur, it seems important to control for the average probability of migration (which is the probability of selection) in both locations which are related to each observation. Moreover, as Wooldridge (2002) makes clear, adding a possibly irrelevant variable among the regressors may only cause inefficiency, while leaving a relevant variable outside the pool of explanatory variables causes inconsistecy. Thus, it is preferable to include the average inverse Mills ratio in both origin and destination, even if one of them may be irrelevant within the analysis.

Therefore, by including the estimated average Mills ratios in equation 4.4, we finally reach the basic form of the second stage Hierarchical Heckit equations of migration flows that we wished to estimate, and which is given by equation 4.7:

$$y = \alpha + \beta_o X_o + \beta_d X_d + \gamma D + \rho_o \lambda_o + \rho_d \lambda_d + \varepsilon \qquad (4.7)$$

where λ_o is the average inverse Mills ratio in the origin municipalities, and λ_d is the average inverse Mills ratio in the destination municipalities. We refer to equation 4.7 as a "basic" form because, in fact, the regressions we have actually estimated in the second stage of our strategy use this equation as a starting point. But before we describe to these specific increments, one final key point concerning the basic form of the migration flows' equation 4.7 must be made, which also regard the following modified versions of it. This key point concerns the values that y assume while representing flows of migrants between two cities, especially for the cases in which there were no flows between a pair of municipalities. Specifically, as both characteristics of the origins and destinations are to be included in the regressions' structural form 4.7 in order to capture both push and pull factors, it becomes difficult to argue that the explanatory variables of both origin and destination are actually relevant whenever y_i equals zero, that is, whenever a particular flow *i* between two cities is zero. This is so because of two main reasons: First, the "destination" and "origin" concepts become somewhat inaccurate in this case, as no individuals have actually moved from one city to another, which means that no real migration has occurred between these two cities. If no migration has occurred, a zero flow would be represented only by the non-migrant population, which by its turn may be composed by a majority of individuals who were not even considering the possibility of migrating to any destinations. In these cases, including specific destination characteristics, or even origin-destination differentials, seems to be unrealistic and inaccurate, and therefore may not be the correct approach.

Secondly (and complementarily), zero flows between two cities are by far the majority of the cases for such a large country as Brazil, which actually pursue over 5500 municipalities. Specifically in our case, if each of these 5500 municipalities shared positive migration flows with every other 5499 municipalities, this would result in more than 30 million positive flows between cities' pairs. However, only 253,038 positive flows are actually observed in 2010 Census' data, which represents only about 1% of the total possible flows. This is so because it is much more likely that each migrant tend to choose his destination from a much smaller poll of previously known municipalities, according to his knowledge about the existence of these destinations and their characteristics. Statistically, including all these zero flows on the sample would mean that only 1% of the sample would show some degree of variance in the dependent variable, even though all other variables would vary among observations. This, in turn, would cause bias to the results, with at least two related problems: first, it would certainly diminish the estimated average importance of destinations and origins characteristics towards migration, when compared to their actual importance, simply because of an asymmetric information matter concerning migrants' knowledge, and not due to these characteristics themselves. This may be interpreted as a bias in the estimation of the influence that the municipalities' characteristics may exert on migration. Second, this concentration of the dependent variable observation over a single value would represent a case of truncation of the regressions' data, which would require additional effort of bias corrections within the estimators, also requiring additional hypothesis and complications related to mixing several different methodologies[37].

Given these reasons, we chose to exclude from our sample all flows that equal zero between any two cities of Brazil. Therefore, the dependent variable of the second stage regressions from our Hierarchical Heckit approach were composed only by positive migration flows. However, one essential remark must be made at this point. If a migration flow between two

---

[37] Such as mixing a Tobit truncated regression method in order to correct the truncation bias (see Wooldridge, 2002) with the Heckit procedure which we already implement in this study.

locations equals zero, this is so because of the non-migrant population, which is also part of our interest population, as we have already discussed. In this sense, this exclusion could lead to a nonrandom sample of our interest population, as we are excluding observations composed only by non-migrants. However, this sample selection problem would be exactly the same that we are already account for by implementing the Hierarchical Heckit procedure: by estimating the propensity of every individual of Brazil to migrate, and including such probability at the migration flows regressions, we are controlling for all non-migrants decisions (of not migrating) in the whole country, including those who live within the cities whose flows equal zero, since the first stage (probability regression) is estimated with every brazilian individual[38] as an unit of observation.

All things considered, we now describe the specific changes we have made to equation 4.7, by which we finally reach the final regressions that we have estimated in this chapter.

Estimation strategy: the migrant's decision whether to migrate to (from) Brazilian Amazon or to (from) the Rest of Brazil

Given the inverse Mills ratios estimated as described above, the main strategy which we have adopted in order to estimate the second stage of the Hierarchical Heckit models[39] is given as follows: as we wish that our results from this second stage regressions reveal the main determinants of the migration flows related specifically to Brazilian Amazon, we have estimated migration flows regressions based in equation 4.7 for all positive flows between pairs of municipalities in Brazil as a whole, but also including the set of explanatory variables multiplied by a dummy variable representing the Amazon region among the regressors.

More specifically, in this first estimation strategy of this hierarchical Heckit second stage regression, we have changed equation 4.7 by including interaction dummies of the Brazilian Amazon with the set of explanatory variables. The specific way by which we have included

[38] Older than 18 years old.

[39] In fact, this strategy was not the only one implemented in this study. Actually, in order to include a more detailed spatial analysis of the Brazilian Amazon migration flows, since such spatial feature might be considered to be a relevant aspect of migration, as pointed out in the literature review section. Thus, we have also tried to implement a spatial econometric regression, following LeSage & Pace 2005 and 2008 methodology. However, some computational and other estimation problems have to be dealt with in order to specifically implement such method in the Census database. Nevertheless, after finding possible solutions to these problems, results proved to be quite similar to those obtained by the main strategy described in this section (4.3). Therefore, given that solving these problems brought the need for additional hypotheses to our estimations, and the main results did not change significantly, we have chosen to include these spatial regressions' methods description, the discussion of these referred problems and their respective implemented solutions, as well as the spatial regression results in the Appendix A.5. of this thesis.

these interaction dummies was by generating two separate equations, 4.8 and 4.9, which we have called, respectively, “the regression of migrants’ decision whether o migrate to Brazilian Amazon or to other part of Brazil” (equation 4.8); and “the regression of migrants’ decision whether o migrate from Brazilian Amazon or from other parts of Brazil” (equation 4.9).

$$y = \alpha + \partial_d^A \mathrm{A} + \beta_o X_o + \beta_d X_d + \partial_d \mathrm{A} X_d + \gamma D + \rho_o \lambda_o + \rho_d \lambda_d + \varepsilon \qquad (4.8)$$

$$y = \alpha + \partial_o^A \mathrm{A} + \beta_o X_o + \beta_d X_d + \partial_o \mathrm{A} X_o + \gamma D + \rho_o \lambda_o + \rho_d \lambda_d + \varepsilon \qquad (4.9)$$

where A represents the dummy vector variable of Brazilian Amazon, with $\mathrm{A} = (a_1, a_2, ..., a_n)$, with n being the number of flows, and with $a_j = 1$ in equation 4.8 if the destination municipality is located within Brazilian Amazon and the origin municipality is located outside Brazilian Amazon, and $a_j = 0$ otherwise (in equation 4.8), for any $j = 1,2,...,n$; and with $a_j = 1$ in equation 4.9 if the origin municipality is located within Brazilian Amazon and the destination municipality is located outside Brazilian Amazon, and $a_j = 0$ otherwise (in equation 4.9), for any $j = 1,2,...,n$. Also, ∂_d^A represents the coefficients of the Amazon dummy at the destination, which captures possible fixed effects of attraction (pull factor) of migration flows towards Amazon that are not specifically related to any explanatory variables; and ∂_o^A represents the coefficients of the Amazon dummy at the origin, which captures possible fixed effects of expulsion (push factors) of migration flows towards Amazon that are not related to any explanatory variables. Moreover, ∂_d represent the partial correlations’ coefficient vector of the interaction between the Amazon dummy vector and the set of explanatory variables at the destination (pull factors), and ∂_o represent the partial correlations’ coefficient vector of the interaction between the Amazon dummy vector and the set of explanatory variables at the origin (push factors).

In our estimations of equations 4.8 and 4.9, the sample is composed by all positive migration flows between any pair of brazilian municipalities. therefore, ∂_d and ∂_o, as coefficients of interaction dummies, may be interpreted as the following: how much each explanatory variable affects migration flows specifically related to Brazilian Amazon, relatively to the effect of that these same variables exert on migration flows related to other locations in Brazil, which by their turn are already measured by β_d and β_o. Therefore, if for example,

β_{dri} is the coefficient of partial correlation between the real income variable at the destination with the migration flows between two any cities located in Brazil (with the destination municipality located outside Brazilian Amazon), then the effect of real income at the destination city over migration flows towards destinations located within the Brazilian Amazon is given by $\beta_{dri} + \partial_{dri}$, with ∂_{dri} being the coefficient of the Amazon interaction dummy of the real income at the destination. A similar interpretation is given for the other Amazon interaction dummies' coefficients in ∂_d in equation 4.8, and in ∂_o in equation 4.9.

And it is due to that specific interpretation that we refer to these two equations as "the migrants' decision whether to migrate to (from) Amazon or other part of Brazil". Further explaining, in equations 4.8 and 4.9 we measure how much each explanatory variable at destinations and origins influences migration flows in Brazil as a whole, on average, but at the same time we also measure how much more (or less) each of these variables specifically influences the migration flows related to the Brazilian Amazon. By doing so, we are able to compare how these three kinds of migration flows (in Brazil's origins and destinations as a whole; flows whose destinations are located within Amazon and origins are located within the Rest of Brazil; and flows whose origins are located within Brazilian Amazon and destinations are located in the Rest of Brazil) differ in terms of their response to each of the push-pull factors among the explanatory variables. In this sense, we are able capture which are the average pull factors at destinations which may drive an average brazilian migrant towards destination municipalities, at the same time comparing how much these factors differ (in terms of size and significance) by being such destination located within the Brazilian Amazon or within the Rest of Brazil. Similarly, we are able to capture which are specifically the push factors at the origin municipalities which may drive the average emigrant away from one brazilian city, comparing how much these factors differ (in terms size and significance) by being this origin located within Amazon or within the Rest of Brazil.

Thus, this econometric strategy means exactly to compare three types of migrants (immigrants of Amazon, emigrants of Amazon and the average brazilian migrant) in terms of their preferences regarding each of the push-pull factors included in the explanatory variables vector. In terms of this study's goals, this comparison makes us able to identify the specific drivers of Brazilian Amazon immigration and emigration, compare them, and also analyze how they differ from migration flows all over Brazil as a whole, which in turn permits us to better understand different aspects of the population movements regarding the Brazilian Amazon.

As a final remark concerning this strategy, it is important to notice that the criterion which we have adopted to separate the three different types of migration flows (Amazon immigration, Amazon emigration and Brazil migration) does not account for intra-regional flows within Brazilian Amazon[40].

4.4. Database Sources and Other empirical limitations and advantages

In this section we briefly describe the sources of the databases used in this analysis, discuss a few limitations imposed by data availability. Further descriptive analysis of the other explanatory variables is made specifically in the following section of this chapter. The reason for that is because such analysis is closely connected to the regressions' results, therefore, interpreting the regressions results and the variables' descriptive analysis altogether ease the comprehension of all evidence found in this study.

4.4.1. Database sources

Most of the data used in our estimations and descriptive analysis was obtained in the Brazilian Census of 2000 and 2010 microdata, whose source is the IBGE[41] (Brazilian Institute of Geography and Statistics). The only exceptions are: the database used to estimate the costs of living per municipality, which mixed data from IBGE's PNAD (National Sampling Survey on Domiciles) and the 2000 demographic Census, as we further explain subsequently; the IFDM[42] (Firjian Index of Municipal Development); and the number of academics per municipality in 2005. The IFDM is a yearly index of municipal development which is calculated by a methodology similar to the one used for the HDI (Human Development Index) calculation. This index is used as one of the explanatory variables in our regressions, and serves as a measure of development and equality for each municipality. The reason why we have chosen this index instead of the decennial HDI-M (Human Development Index –

[40] The reason for that is because we are trying to focus on how the Brazilian Amazon region exchanges population with the rest of Brazil. Intra-regional flows, in this sense, would only make it difficult to isolate the differences in responses of the three kinds of migrants to each push or pull factor. This is so because, for example, one individual who moves from an Amazonian city towards another Amazonian municipality is at the same time an Amazon immigrant and an Amazon emigrant. Thus, this individual may pursue, at the same time, characteristics from both of these types of migrants. And being our goal to compare between these profiles, including this type of individual might only obfuscate the analysis.

[41] *Instituto Brasileiro de Geografia e Estatística.*

[42] *Índice FIRJAN de Desenvolvimento Municipal.*

Municipal) calculated by PNUD[43] (United Nations' Program of Development) using census data, is because the IFDM was available for the year of 2005, which matches exactly the first year of the migration flows considered in this study, while the HDI-M was only available for the years of 2000 and 2010. Another important aspect of the IFDM index is that, once again as the HDI, it can be divided into three categories: Education, Health and Income. Throughout estimations, we use these categories separately many times. Finally, for the number of academics per municipality, data was obtained from the Brazilian Ministry of Education and Culture (MEC) Educational Census, also for the year of 2005.

Another data limitation

Besides all the restrictions already discussed in the methodology section, there is another limitation, still unmentioned, which is imposed by the database which we chose to use in this study: in our estimation procedures, we are not including any punctual exogenous shocks that may drive migration flows towards some specific destinations. One example of this kind of shock would be the construction of a hydroelectric plant in one of the Brazilian Amazon municipalities. It is common knowledge that a project such as this one would directly drive migration flows towards this municipality, and probably the exogenous variables included in our model would not capture the whole set of incentives which this shock causes. However, even though these may be specific drivers of migration, we remind that the aim of this study is to comprehend long run patterns of migration in Brazilian Amazon, and punctual exogenous shock might not be classified as such, exactly due to their exogenous and punctual nature. Still, we are aware that our coefficients may probably be affected by the absence of this kind of shocks[44] in our regressions.

Estimating Costs of Living

As yearly data regarding costs of living per municipality were not available in any brazilian database, and this is an extremely important explanatory variable in the analysis of migration flows (by being one essential part of the real income differential between municipalities), we chose to try to estimate it, to then include it in our econometric approaches. We are aware that this option might not be ideal, since estimated variables might not capture the whole effects

[43] *Programa das nações unidas para o desenvolvimento.*
[44] We remind that one possible way to try to capture these shocks would be to include dummy variables representing the exact year which the individual has migrated. However, we are not able to do this, since we do not pursue such information, due to the fact that the migration database which we use here regards a "fixed date" migration database, as previously discussed. This means that the only information available in our data refers to where the individual used to live in 2005, and where he was living in 2010.

which the actual variable would; and also because including more steps of estimation in econometrical methods may lead to inefficiency associated the multicolinearity that this estimated variable will share with the other exogenous variables included in the regressions (since it is possible that some of the explanatory variables used to estimate the costs of living per municipality are also explanatory variables in the migration flows' regressions).

However, as the majority of the theoretical literature reviewed in this study consider costs of living as one of the most important variables to be included when studying the determinants of migration, by being part of the real income differentials, then omitting it would most likely be biasing in our empirical results, which would compromise our results even more than a multicolinearity issue. Results concerning such estimation, as well as a description of the costs of living predictions are shown in Appendix A.3 of this thesis.

4.5. Results and discussion: the determinants of Brazilian Amazon migration flows

In this section we discuss the main results of our econometric estimations. One important remark concerning these results is the fact that, as in most econometric approaches, we have included a large group of explanatory variables in the estimations, in order to avoid a possible bias caused by the omission of relevant variables. Each of these variables' inclusion was made following the theoretical framework previously revised. However, in the next subsections, we focus the discussion on the main results and the main interest variables, thus considering many of these explanatory variables as "controls". Nevertheless, these controls' coefficients and their respective significance tests are exhibited in the tables which present the results.

4.5.1. Results from the Hierarchical Heckit First Stage

Given the average individual characteristics of the migrants which compose our sample, we now turn to the specific results from our econometric estimations. In this subsection we present the results of the Hierarchical Heckit first stage, which provide the first evidence about migration determinants in Brazil. Table 4.6 shows the results of the first stage probit regression, where we estimate the probability of one Brazilian individual to migrate, given his individual characteristics and the characteristics of his origin municipality[45].

---

[45] Two important reminders: first, the sample used to estimate this regression refers to all Brazilian individuals, without separating the ones whose origin or destinations are within or outside Brazilian Amazon. Second, this regression stands as the first stage of all Hierarchical Heckit procedures implemented in this study in order to correct sample selection bias. Therefore, it is important to keep in mind that the main purpose of this regression

Table 4. 6: "Hierarchical Heckit" First Stage: individual probability of migration in Brazil, explained by push factors[46] and individual characteristics

Explanatory Variable	*Probit Coefficient*	*t-stat*	Variable Group
Age in 2005	-0.044***	(-184.055)	Individual Characteristics
Squared Age in 2005	0.001***	(101.234)	
Gender (1 if man, 0 if woman)	0.069***	(44.050)	
Etnic Group - Black (omitted: Caucasian)	-0.053***	(-19.429)	
Etnic Group - Oriental (omitted: Caucasian)	-0.044***	(-6.606)	
Etnic Group - Mullatto (omitted: Caucasian)	-0.061***	(-38.683)	
Etnic Group - Indians (omitted: Caucasian)	-0.053***	(-4.413)	
Previous (1991 - 2000) Labor Income Growth (Origin)	-0.020***	(-5.812)	Income & Employment
Average Labor Income (Origin) (log)	-0.022***	(-4.372)	
Avg. Cost of Living (Origin) (log)	0.070***	(13.840)	
Cost of Living (Origin) (log) - Standard Deviation	-0.001***	(-8.253)	
EAP / WAP ratio (Origin)	-0.018	(-1.069)	
% of "Formal" Working force	-0.003	(-0.281)	
Gini Index (Origin)	1.150***	(59.046)	Equality & Health
IFDM Health Index (2005)	-0.053***	(-4.374)	
% of Pop. w/ Access to Sanitary Treatment (Origin)	-0.133***	(-32.915)	
Population in the Origin City (2000)	0.001***	(-45.669)	Demographic & Urbanization
% of Urban Population (origin)	0.099***	(12.201)	
Pop. Density (Origin)	0.001***	(7.235)	
Squared Pop. Density (Origin)	0.001***	(5.611)	
% of Labor Force in Services and Commerce (Origin)	-0.490***	(-49.807)	
Previous (1991 - 2000) Migration Inflow (Origin)	-1.667***	(-48.068)	
Squared Previous (1991 - 2000) Migration Inflow (Origin)	2.653***	(37.094)	
% of Iliteracy (Origin)	-0.238***	(-13.223)	Education
Number of Academics (2005) (Origin)	0.001***	(44.618)	
% of Pop. w/ College Degree (Origin)	0.932***	(25.386)	
IFDM Education Index (2005)	-0.302***	(-31.027)	
Constant term	-0.427*	(-1.803)	
Number of obs = 12,049,139 Number of Strata = 10,184 Population size = 113,222,203			

Note 1: *** p<0.01, ** p<0.05, * p<0.1

Note 2: All variables refer to the year of 2000, except when described in the variable's name.

Note 3: Besides the coefficients shown in this table, the probit model also included dummies for the economic sector in which the individual was working in 2010, as control variables, in order to capture possible relevant fixed effects. These coefficients are ommited in this table in order to save space and due to the fact that those are not our main interest variables. The complete table including these coefficients is attached in the Apendix.

SOURCE: Own Elaboration

is to serve as the basis for the estimation of the inverse Mills ratios used in the econometric strategy of estimation of the second stage regressions, as we have previously discussed.

[46] Besides the explanatory variables shown in table 4.8, we also have included other 27 intersection dummy variables representing each Brazilian State, with these variables assuming the value 1 for the State in which the individual has lived in 2005, and zero otherwise. These were included in order to capture regional fixed effects which may influence the individual whether to migrate or not. For example, it is possible that one peculiar state have a policy which hinders that individuals leave their origin cities. In this case, as this policy cannot be included as an explanatory variable, then the State intersection dummies may capture this particular effect, along with other possible “fixed effects”. However, as these variables are not our main interest ones, and the main objective of this regression is simply to estimate the Inverse Mills Ratios to be used in the second stages, then we have omitted these coefficients from table 4.8 in order to save space. The regression with the full set of variables, including these dummies, is present on the Appendix A.4 of this work.

In order to ease the coefficients' interpretation from table 4.8, we have grouped variables into five categories, according to their characteristics. The first thing that calls attention in table 4.8 is that most of the variables are significant in our estimations. Given that our sample is composed by 12 millions observations whose weights represent the 113 million inhabitants (the size of brazilian population with over 18 years in 2005), one might argue that these statistic significances might be influenced by this huge sample size of the survey, as the coefficients' variances tend to get smaller the higher the number of observations in the sample. However, this is not the case in our estimations, since we have corrected all coefficients' variances in this stage by making use of the variance linearization[47] process suggested by Korn (1990). Thus, the significance found in our estimations of table 4.8 may be considered statistically reliable and robust.

Among the results regarding the individual characteristics' explanatory variables, we can notice that our estimations predict that Caucasian individuals are the ones who are most likely to migrate, as the dummy variable representing this ethnic group is the one omitted[48] in the regression, and all ethnic groups' coefficients are significantly negative. Moreover, younger people[49] are more likely to migrate, as is the Brazilian masculine population (when compared to the feminine). Fortunately, all these results are in accordance to table 4.7, where individual characteristics of the migrants are summarized: there, caucasians represent the biggest share of migrants for Brazil.

By their turn, income and employment variables at the origins corroborate the predictions of the theoretical framework revised in this chapter: individuals are less likely to migrate from origins which exhibit higher average labor nominal income and lower costs of living. This evidence is also found in our second stage regressions, but relevant differences appear regarding this variable when we consider specifically the migration flows of Brazilian Amazon.

---

47 This procedure is implemented by the "svy" commands in Stata. To more details, see the Stata survey manual (2009).

48 In order to avoid the classic econometric issue called the "dummy trap".

49 It is important to mention at this point that, as in the first stage we have include the age of the individual in our estimations, then the shares of different age groups in the municipality's population were excluded from the first stage regressions. The reason for that is because as in the first stage we are concerned in explaining the individual choice regarding the migration decision, it seems more likely that each individual's own age is what counts at the time that the decision is taken. As in the second stage regressions our sample defined at the municipality level, it is necessary to include the shares of each age groups of the population per municipality among the regressors, otherwise the influence of population's aging would be omitted, possibly causing inconsistency in our results.

Moreover, the previous income growth in the origin municipality coefficient presents a significant and negative sign, as well as the coefficient of the previous immigration flows (from 1991 to 2000) variable. This suggests that cities which have presented a large number of emigrants in the recent past now tend to relatively retain population. Moreover, equality explanatory variables also show the expected signs and significance in the first stage Heckit regression. Migrants are less likely to move away from cities with higher initial development level (measured by the IFDM), with higher access to sanitary treatment or lower levels of income inequality (measured by the Gini Index). Once again, this is in accordance with the theoretical framework revised here.

Furthermore, regarding the urbanization variables, a dubious result is found at this first stage: on the one hand, the share of urban population and population density present positive and significant signs, whereas the share of employment in the commerce and services sector coefficient is negative, which hinders us to draw any conclusion in this regard. Possibly, this may be happening due to multicolinearity among these variables, as they all capture a similar economic aspect. In this case, it would be recommended to withdraw at least one of these variables from the analysis, if the main goal of this analysis were to analyze the urbanization coefficients' sign. However, we chose to keep all the three of them among our regressors due to two reasons: although the results of these variables coefficients are dubious at this stage, identifying the sign of these is not our main goal at this point. And more importantly, these three measures capture different aspects of urbanization (as discussed in the previous chapter of this thesis). Therefore, excluding one of them would be an arbitrary decision, which could even lead to the omission of a relevant variable bias, if the omitted characteristic were significant. Moreover, when we include the urbanization level of the destination municipalities on the second stages of our Heckit procedures, this duality in the results disappear, and as the first stage estimations serve mostly to predict the inverse Mills ratios, we have chosen to maintain all of these three variables in our first stage regression[50].

A similar dual picture occurs regarding the educational variables' coefficients in this first stage regression: on the one hand, the IFDM education index coefficient indicates that higher educational levels of one origin municipality tend to encourage migrants to remain in this

[50] In the case of Brazilian data, it is possible that the share of employment in the tertiary sector might not be considered as a good representative of urbanization, due to the fact that in many small cities, this sector is mostly composed by public services and/or public administration. Nevertheless, as it is just one among many other explanatory variables, we have chosen to keep such variable as one possible representative of urbanization, also in order not to contradict empirical literature on urbanization, as reviewd previously in this chapter.

origin. On the other hand, the origins' illiteracy rates, the share of population with college degree and the number of academics in the origin municipality coefficients all point to the opposite direction. Once again, there is a high colinearity between these variables which may be driving such results. However, as each of these variables capture different aspects of education, and as the specific analysis of these coefficients is not our main objective at this point (at the first stage regression), we once again chose to keep all three of them in our probit regression. Furthermore, as in the case of the urbanization variables, this duality concerning educational results disappear when we include the educational levels of the destination cities (measured by these same variables) at the Heckit second stages' regressions.

Furthermore, as the main goal of this first stage estimation is to produce reliable predictions of the average individual probability of migration (and afterwards, per municipality), we have produced and tested such individual predictions from our models. Table 4.9 brings the average individual probability of migration, predicted by this probit regression, for the sample divided in two different population groups: the group containing only migrants, and the group containing only non-migrants.

As we can observe, our first stage estimations may also be considered successful in this sense, because our predictions tend to estimate a three times higher probability of migration for the individuals who have actually migrated in our sample, when compared to the predicted probability of migration for the non-migrants, also produced by our first stage regression.

Table 4. 7: Hierarchical Heckit First Stage Fit: comparing estimated probabilities of migration between Migrants and non-migrants

Avg. Estimated Probability of Migration for "only migrants" sample group		***Avg. Estimated Probability of Migration for "only non-migrants" sample group***	
0.16	Obs: 911,518	0.06	Obs: 11,139,501

SOURCE: Own Elaboration

4.5.2. Results from the Hierarchical Heckit Second Stage Regressions

Given the average inverse Mills ratios and the results from the Hierarchical Heckit first stage regression, we now analyze the main results of this chapter, which refer to the results of the Hierarchical Heckit second stages regressions.

In order to ease the understanding of these results, they were split in two following subsections: In the first subsection, we show and discuss the second stage results of the regressions regarding the Brazilian Amazon immigration flows; and in the second subsection, we show and discuss the second stage results regarding the Brazilian Amazon emigration flows. Throughout these subsections, we compare the differences and similarities found for each of these kinds of flows, also comparing them to the results regarding the general Brazilian migrations flows.

Furthermore, in order to ease the construction of the arguments brought by the evidence found in these tables, we have grouped the variables according to the categories described in the "Variable Group" column of table 4.8. Among these groups, there is one called "Other Control Variables", which contains variables that were mostly included to avoid the bias of omitting relevant variables, but whose coefficients interpretation we have not included in our discussion. The reasons for this choice are three: 1) some of these variables' coefficients, differently from the main variables of the other groups (real income variable, for example), do not pursue a previously expected sign due to their complex nature (such as the Gini Index, previous migration flows and real income growth, and the EAP/WAP ratio[51]) which creates different possible interpretations to these coefficients, thus might lead to incorrect arguments; 2) some of these variables' coefficients interpretation have not contributed to the evidence built over the main explanatory variables' coefficients, by not exhibiting statistical significance or only corroborating the other main results; 3) some of these variables do not specifically contribute to bring any new evidence on migration determinants, serving mostly as controls to the regressions. "Population size" is an example of it, since more populated cities both receive and send greater migration flows.

Hierarchical Heckit Second Stage Results: Brazilian Amazon immigration determinants

Table 4.8 brings the results of the second stage of the hierarchical Heckit model which focus on Amazon immigration flows[52]. As further explained previously, results from this table bring evidence on the answer to the following question: what variables in the origin and destination municipalities make migrants to choose to move towards Brazilian Amazon (coming from a different part of Brazil), instead of moving to other country regions?

---

[51] Economic Active Population divided by the Working Age Population. This ratio is also referred as the rate of participation of one municipality, and measures the share of the population of one city which is able to work that is is actually working by the date which the ratio is measured.

[52] As a reminder, this means that it brings the results from the second stage Heckit regressions which includes all migration flows in Brazil as observations, but which also includes interaction and intersection dummies of the Amazon immigration flows for all explanatory variables, as explained in the methodology section.

Table 4. 8: "Hierarchical Heckit" Second Stage - Migration in Brazil focusing on the Brazilian Amazon Immigration Flows

Explanatory Variable	Brazil		AMZ Imigration Interaction Dummies		(Ho: β + α = 0) Joint Test, p-value	Variable Group
	Coefficient β	*t-stat*	*Coefficient α*	*t-stat*		
Avg. Labor Income (Destination) (log)	3.798*	1.8	13.179***	4.6	0.0000***	**Income and Employment**
Avg. Labor Income (Origin) (log)	-1.640	-1.0	3.317	1.2	0.4502	
Avg. Cost of Living (Destination) (log)	2.931**	2.0	-16.059***	-5.4	0.0000***	
Avg. Cost of Living (Origin) (log)	-2.960**	-2.2	3.769	1.6	0.6906	
Unemployment rate (Destination)	18.690***	2.8	-48.308***	-5.1	0.0000***	
Unemployment rate (Origin)	-23.275***	-4.6	42.799***	3.7	0.0557*	
% of Pop. w/ Access to Sanitary Treatment (Origin)	7.364***	6.6	4.650**	2.3	0.0000***	**Health**
% of Pop. w/ Access to Sanitary Treatment (Destination)	-1.425	-0.8	21.393***	6.5	0.0000***	
% of Labor Force in Services and Commerce (Origin)	70.386***	22.9	-64.804***	-13.9	0.1075	**Urbanization & Age**
% of Labor Force in Services and Commerce (Destination)	42.756***	8.6	-21.032***	-3.4	0.0000***	
Pop. Density (Origin)	0.002***	2.9	-0.002**	-2.4	0.6595	
Pop. Density (Destination)	0.002**	2.5	0.006	0.7	0.3053	
% of Pop. aged 22 - 29 (Origin)	233.119***	7.0	19.919	0.3	0.0000***	
% of Pop. aged 55 - 60 - "about to retire" (Origin)	-633.771***	-8.9	65.334	0.6	0.0000***	
% of Pop. aged more than 60 - "retired" (Origin)	-453.917***	-13.2	603.860***	10.6	0.0005***	
% of Pop. aged 22 - 29 (Destination)	145.578***	4.5	-249.532***	-4.8	0.0110**	
% of Pop. aged 55 - 60 - "about to retire" (Destination)	-169.428**	-2.1	-589.475***	-4.6	0.0000***	
% of Pop. aged more than 60 - "retired" (Destination)	-474.509***	-13.7	223.189***	4.1	0.0000***	
% of Iliteracy (Origin)	51.404***	9.7	-26.519***	-3.0	0.0006***	**Education**
% of Iliteracy (Destination)	4.196	0.6	-21.037**	-2.5	0.0015***	
% of Pop. w/ College Degree (Origin)	243.568***	9.6	-145.002***	-4.7	0.0000***	
% of Pop. w/ College Degree (Destination)	325.642***	13.1	-265.300***	-3.8	0.3526	
Average Mills Ratio - "propensity to migrate" (Origin)	-131.103***	-23.5	85.485***	10.1	0.0000***	**Sample Selection**
Average Mills Ratio - "propensity to migrate" (Destination)	-23.504**	-2.2	-7.968	-0.7	0.0000***	
Centroid Distance (Origin - Destination) (log)	-25.084***	-27.8	2.039***	11.1	0.0000***	**Spatial**
Constant term	346.642***	15.0	-41.686	-1.2	0.0000***	
per capita GDP (Destination) (log)	-3.113***	-2.9	1.648	1.2	0.0744*	**Other "control" variables**
per capita GDP (Origin) (log)	-1.306**	-2.0	2.606**	2.2	0.1742	
EAP / WAP ratio (Destination)	32.477***	8.0	-20.330***	-3.6	0.0028***	
EAP / WAP ratio (Origin)	16.300***	4.3	-4.838	-0.7	0.0740*	
Previous (1991 - 2000) Labor Income Growth (Destination)	-5.030***	-4.6	2.813**	2.2	0.0009***	
Previous (1991 - 2000) Labor Income Growth (Origin)	-1.467**	-2.2	13.165***	9.7	0.0000***	
Gini Index (Origin)	-52.645***	-6.1	-22.712*	-1.7	0.0000***	
Gini Index (Destination)	13.947	1.0	-69.192***	-4.2	0.0000***	
Previous (1991 - 2000) Migration Inflow (Destination)	15.660	1.3	-40.467***	-3.0	0.0000***	
Previous (1991 - 2000) Migration Inflow (Origin)	-8.541	-1.9	5.011	0.5	0.6844	
% of Pop. Born in the City (Destination)	10.120	0.9	12.254	1.0	0.0000***	
% of Pop. Born in the City (Origin)	120.409***	21.6	-81.661***	-9.7	0.0000***	
Pop. (Origin)	0.001***	9.0	-0.0001***	-4.7	0.0010***	
Pop. (Destination)	0.001***	8.2	-0.00001	-1.2	0.0222**	
Number of Academics (2005) (Origin)	-0.002***	-7.1	0.001***	3.7	0.0033***	
Number of Academics (2005) (Destination)	-0.002***	-7.2	0.002**	2.5	0.9260	

Note 1: *** p<0.01, ** p<0.05, * p<0.1

Note 2: All variables refer to the year of 2000, except when described in the variable's name.

SOURCE: Own Elaboration

Income and Employment: The first results observed in both tables 4.8 bring evidence that the immigration flows towards Brazilian Amazon are positively influenced by the migrant's search for higher real income (both in terms of higher nominal income and lower cost of living), which goes along with most of the theoretical models of migration, as the neoclassic growth models and the ones from New Economic Geography and Urban Economics. This evidence is found significantly in our regressions, even though the average real income in Amazon is lower than the average real income in the rest of Brazil, when these averages are weighted by the overall population from each region, as Table 4.9 shows.

More specifically, results from table 4.8 show that the higher nominal labor income and the lower the costs of living in the municipalities of destination located within Brazilian Amazon, the greater are the flow of immigrants towards these cities. In fact, the coefficients' interpretation suggest that higher levels of real labor income in the destination cities work as a force of attraction to all brazilian migration flow, but this attraction force is even greater when the destination municipality is located within Brazilian Amazon, as the coefficient of the interaction dummy of Brazilian Amazon with the nominal income at destination coefficient is positive and significant.

Furthermore, regarding the average costs of living at the destination municipality, the sum of the coefficient for brazilian migration flows in general with the coefficient of the interaction dummy of Brazilian Amazon with this variable ($\beta + \alpha$) show that higher costs of living in the destination exert negative influence on migration flows if the destination is located within Brazilian Amazon, even though this influence is positive if the destination city is located within the rest of Brazil (as the β coefficient is positive). This further corroborates the evidence on the attractive force which real income within the destination municipalities exert over immigration flows towards Amazon.

In turn, nominal income and costs of living within the migrants' municipalities of origin do not exert any effect on pushing out Amazon immigrants. On the other hand, for those who migrated to other locations of Brazil, higher average costs of living in the city of origin reduces the size of these migration flows, which is unexpected, since higher costs of living in the city of origin is mostly considered as a push factor. This opposite result for non-Amazon immigrants may be resulting from the fact that higher costs of living can be capturing other non-observed effects, such as living conditions in these origin cities. In this case, this variable may not only be considered as costs to be discounted from the individual's nominal earnings, that is, it may not be representing a only a part of the average real income in these municipalities. In this sense, this result can be considered as another evidence that immigration flows towards Brazilian Amazon seem to be more heavily influenced by the search for higher real income in the short run, whereas migration flows towards other regions of Brazil, on average, are probably less relatively influenced by such direct search.

Therefore, summing up the results concerning real income, results suggest that Brazilian Amazon immigrants, on average, seem to choose to migrate influenced by the search of a positive real income differential, as theory would predict. Such result may be considered counterintuitive at first, if one considers that average real income in the Rest of Brazil is higher than the average real income within Amazon, when these averages are calculated using all brazilian cities (according to our calculations, average real income in the Rest of Brazil is R$ 454.28 per month, against an average real income of R$338.73 per month in Brazilian Amazon, weighting these averages by the populations of these municipalities). However, further descriptive analysis of real labor income in cities specifically involved with the migration flows, and also considering the migration flows as the weights of these averages, shows that this econometric results may be considered reasonable.

Table 4. 9: : Average Labor Income and Cost of Living, weighted by the municipalities populations

In AMZ Destination cities of Migrants coming from the RoB Municipalities			**In RoB Origin cities of Migrants migrating to the AMZ Municipalities**		
Average Labor Income (R$ / Month)	*Cost of Living (R$/month)*	*Real Income (R$ / Month)*	*Average Labor Income (R$ / Month)*	*Cost of Living (R$/month)*	*Real Income (R$ / Month)*
506.31 (187.96)	167.58 (71.52)	338.73	706.34 (292.29)	252.10 (114.91)	454.24

Notes: Standard Deviation in Brackets; AMZ - "Amazon region"; RoB - "Rest of Brazil"

Source: Own Elaboration, IBGE Census and PNAD data.

In order to build such analysis, firstly we justify that to what concerns our goals, it does not seem correct to simply compare the average real income of all municipalities from Brazilian Amazon with the average real income of all municipalities in the rest of Brazil, weighing up such averages by the populations of these, when it comes to analyze migration flows. Two reasons justify this affirmation: 1) many of these municipalities are not origins or destinations of any migrants who compose the flows; 2) the most correct weighting for this calculation should be the intensity of migration flows between each origin and destination, since we want to compare the levels of real income at the migrants' origins against destinations, but always giving more importance to the incomes of municipalities that receive or send the larger shares of migrants.

Therefore, we have calculated the average real income of the municipalities that are destination of migrants within Amazon, pondering this average by the flow of migrants which arrived in each destination, and compared this average with the average income in the municipalities of origin (located in the Rest of Brazil) of immigrants who moved towards Amazon, weighted by the migration flow of exit in each origin city. These calculations resulted in a different perspective: with migration flows as weights, the average real income of the destination municipalities within Amazon becomes greater than the average real income of the origin cities within the rest of Brazil (of the Amazon immigrants), as tables 4.10 and 4.11 show.

Table 4. 10: Average Labor Income and Cost of Living in AMZ destinations, weighted by the destination inflow of migration

Average Labor Income (R$ / Month)	*Cost of Living (R$/month)*	*Real Income (R$ / Month)*
411.98 (187.97)	135.17 (54.26)	276.82

Notes: Standard Deviation in Brackets; AMZ - "Amazon region"; RoB - "Rest of Brazil"
SOURCE: Own Elaboration, IBGE Census and PNAD Data

Table 4. 11: : Average Labor Income and Cost of Living in RoB Origins, weighted by the origin' outflow of migration

Average Labor Income (R$ / Month)	*Cost of Living (R$/month)*	*Real Income (R$ / Month)*
430.49 (202.67)	163.73 (85.39)	266.76

Notes: Standard Deviation in Brackets; AMZ - "Amazon region"; RoB - "Rest of Brazil"
SOURCE: Own Elaboration, IBGE Census and PNAD Data

At this point, we highlight that the previous analysis on the regional characteristics of migration flows corroborates such evidence. As previously shown, the largest share of Amazon immigrants stems from the Northeast region of Brazil. By calculating the average real income of labor within this region and comparing it to the real income in Amazon destination cities, we find that, in general, real income in Northeast municipalities of origin are lower (R$ 300.57 per month in Northeast against R$ 338.73 per month in Amazon).

Similar results were observed with respect to unemployment on table 4.8: individuals who migrate to the Amazon region tend to leave origins with higher unemployment rates and migrate to cities with lower rates. This evidence is corroborated by comparing the unemployment rates between the origin cities of migrants coming from the rest of Brazil with those of the destination cities of migrants moving towards Amazon[53], as in Table 4.12, which shows that unemployment tend to be lower in the Amazon destination cities, thus justifying

[53] Again, considering the flows of migration as the criterion of weighting, due to the reasons discussed before.

the unemployment push force in the origin, and the unemployment attraction force in destinations.

On the other hand, this pattern is unexpectedly reversed when dealing with migrants who have migrated to regions of Brazil other than Amazon: results from table 4.8 suggest that migration flows are larger the higher the unemployment at the destination municipality, and the smaller the unemployment rate in the origin. As this is an unexpected result, as it is hard to argue that in fact individuals migrate towards cities with higher unemployment, we interpret such evidence in a more conservative way: it seems more reasonable that, in fact, individuals who migrate to (and between) areas of Brazil outside Amazon are in fact searching for cities with characteristics other than low unemployment, and such characteristics happen to be located at cities that also present high unemployment rates. In this sense, migration would be flowing to cities in the rest of Brazil with high unemployment rates, despite such unemployment size.

Nevertheless, an even more conservative interpretation allows us only to consider this result regarding unemployment as a confirmation of a distinctive pattern between migrants who intend to Brazilian Amazon, in comparison to migrants who recently moved to other regions of Brazil: whereas our evidence point that Amazon immigrants tend to move to cities with higher real income level and lower unemployment rates, these same variables do not seem to be as relevant to the migration decision of brazilian migrants heading to other places of Brazil.

Table 4. 12: Average Unemployment Rate

In AMZ Destination of migrants coming from RoB, weighted by the destination's inflow of migration	In RoB Origin of migrants going to AMZ, weighted by the origin's outflow of migration	Difference
11.36% (5.88%)	12.40% (5.92%)	-1.04%

Notes: Standard Deviation in Brackets; AMZ - "Amazon region"; RoB - "Rest of Brazil"

SOURCE: IBGE Census data. Own Elaboration

Health: Results concerning access to sanitary treatment are somewhat controversial as push and pull factors. On the one hand, as expected, immigrants tend to move to the Amazon region as a destination seeking better sanitary conditions of living, measured by the share of population with access to sanitary treatment in target municipalities. On the other hand, however, these same flows of immigration towards Amazon tend to be larger the higher the access to sanitary treatment in the cities of origin of these migrants. This controversial result (specially for the for the origin cities) is even more unexpected regarding migrants whose destination and origins are located within the rest of Brazil: in this case, migrants tend to leave cities with better sanitary conditions, and access to sanitary treatment is no longer significant in their destinations. Therefore, evidence regarding a search for better health conditions in the migration decision is inconclusive from our results, and we conservatively assume that if health plays a role in determining migration flows, this is not being properly captured by our estimations[54].

Urbanization and Age: With respect to the variables which represent urbanization, results from table 4.8 once again indicate that immigrants who move to the Brazilian Amazon also present a distinct profile of those who decide to migrate to other places in Brazil. Specifically, brazilian migrants who neither originate nor are intended to Amazon tend to migrate from highly urbanized centers towards to cities which are also highly urbanized, according to the coefficients of the share of employment in the tertiary sector of the economy and population density (which for both origins and to destinations in Brazil as a whole, excluding Amazon, are positive and significant).

More specifically, immigrants who go to Amazon also tend to choose destination cities with the highest rates of urbanization measured to the share of jobs in the tertiary sector[55]. On the other hand, this same variable at the origin do not act as a force of expulsion for these same immigrants, since the sum of the coefficients of the interaction Amazon dummy with the coefficients of the share of employment in the tertiary sector for Brazil is not statistically different from zero. Likewise, urbanization measured by population density in the origin or destination do not push migrants towards Amazon, according to the Wald test on the sums of coefficients related to this variable, unlike the evidence found for the rest of Brazil, as described above.

Therefore, we find evidence that in Brazil as a whole, which has already reached a more consolidated urbanization level when compared to the Amazon region, people tend to migrate

---

[54] It is possible that such inconclusive and controversial result is given by the fact that the true health conditions of the origin and destination municipalities are poorly represented in our models, due to lack of precise data which we could use in our regressions. Nevertheless, according to the literature which supports our estimations, health was never specifically pointed out to be one of the most important determinants of migration : most of the times it is only one among several representatives of quality of life. Therefore, we assume that our results are not enbiased by such lack of health conditions representation in our modelling.

[55] Since the sum of the coefficient for Brazil with the interaction between the Amazon dummy and the terciary sector percentage variable is positive and significant, even though the coefficient of the interaction alone is negative, indicating that the force of attraction of greater urbanization of fate is weaker for those intended to Amazon, but it is still positive

without abandoning the urban centers, that is, to move from densely urbanized areas towards other highly urbanized locations as well. On the other hand, for the Brazilian Amazon (which is still going through a process of urbanization), evidence found points to an immigration process that supports urbanization as a process in evolution: individuals who immigrate to Amazon tend to seek the more highly urbanized centers, but they do not come necessarily from places as urbanized as their destinations. In other words, they tend to come from both more or less urbanized cities.

Descriptive analysis confirms this result (see table 4.13). The average percentage of jobs in the tertiary sector in Brazil was around 81.60% in 2000. In turn, this average was 58.99% for the destination cities of the immigrants of Amazon in this same year, while this average percentage in the origin municipalities of these immigrants (located in Rest of Brazil) was 65.62%[56]. Thus, firstly, a comparison between these averages shows how much the process of urbanization is more consolidated in Brazil as a whole. Secondly, it evidences that the origins of immigrants intended to Amazon are much less urbanized than Brazil as a whole, even though these cities are also located in the Rest of Brazil (therefore, indicating that immigrants of Amazon come from both urbanized and rural areas[57]).

Table 4. 13: % of Employment in the Tertiary sector

In Brazil, weighted by municipalities' populations	In AMZ Destination of migrants coming from RoB, weighted by the destination's inflow of migration (A)	In RoB Origin of migrants going to AMZ, weighted by the origin's outflow of migration (B)	Difference (A - B)
81.60% (21.98%)	58.99% (20.83%)	65.62% (21.80%)	-6.63%

Notes: Standard Deviation in Brackets; AMZ - "Amazon region"; RoB - "Rest of Brazil"

SOURCE: IBGE Census data. Own Elaboration

With respect to the demographic variables, results from table 4.8 present some clear patterns in migration decisions. First, we see that cities with a higher percentage of younger population (aged between 22 and 29) are the most dynamic in terms of mobility, as migration flows tend to be higher in both origins and destinations whose share of this kind of population is higher. Similarly, the higher the percentage of the elderly population in the municipalities of

[56] Due to the same reasons as described in the results concerning income and cost of living, we used the flows of migrants as weights in these average calculations.

[57] Ideally, in order to confirm this evidence of rural exodus, it would be necessary to make a specific analysis of migration which includes the migrant within the same county or state who migrated from rural areas to urban areas as the object of study. Unfortunately, the database of the 2010 Brazilian Census does not include the information regarding if the migrant used to live in a rural or in an urbanized location in his origin municipality, and due to this reason, we were not able to implement such experiment in this work.

destination or origin in Brazil as a whole, the lower the flow of migration related to those cities. This indicates that the aging of the population makes a city to become less dynamic in terms of mobility. This result is in agreement with LeSage (2005) who argues that these variables are to be included as explanatory in migration analysis precisely in order to capture this kind of dynamic. According to the author, it is expected that younger population should be naturally more willing to move in search of new jobs, higher income or other perspectives of living, due to their longer life expectancy.

Education: Results concerning education variables one again suggest important differences between immigrants moving towards the Brazilian Amazon and the other general brazilian migrants. First, we notice that for those who migrate to any region of Brazil, higher levels of illiteracy in the city of origin tend to act as a push factor, which points out to a first evidence that the search for higher levels of education might be considered a relevant motivation in the migration decision for Brazilians in general, as the Human capital theories of migration would predict (see Becker, 1993; Borjas, 1989). For individuals who immigrate towards Amazon, this result goes in the same direction in terms of sign but with less intensity[58]. On the other hand, low percentages of illiteracy in the destination municipality does not act as a factor of attraction for Brazilian migrants in general, whereas for the Amazon immigrant, this variable also acts as a pull factor[59].

A similar result occurs with respect to the percentage of population with college degree. On the one hand, results suggest that this kind of population is naturally more dynamic in terms of mobility. Higher percentages of this kind of population in both origins and the destinations encourage greater flows of migration for all cities in Brazil, on average. This result is expected, since more educated individuals tend easily place themselves in the labor market, which in turn provides them greater possibility of quicker adaptation to new locations. However, these variables demonstrate a slightly different performance concerning the individuals who decide to immigrate to Amazon. In this case, only the percentage of individuals with college degree at the origin remains significant and positive, and also less

[58] As the coefficient of the interaction Amazon dummy for this variable is negative, but the sum of this dummy coefficient with the one for the same variable for brazilian migrants in general is still positive and significant, according to the Wald test.
[59] Although this evidence might be considered weak, since this coefficient proves significant in only one of the models with only immigrants of Amazon in the sample.

intensively than for migrants from Brazil in general[60]. Thus, higher levels of population with College degree in Amazon destination cities do not exert attraction to migrants in general, according to our results.

One important remark, which has already been discussed in the literature review section of this chapter, must be made at this point. Such point is: migration flows towards cities with higher education levels may also be interpreted as a real income search pull factor of migration, with the difference that migrants seeking education are more concerned to earn higher levels of real income in the long run, differently from migrants who move seeking immediate higher real income levels at new jobs in new cities. This is exactly the main argument presented by Human Capital theories (see Becker, 1993; Borjas, 1989): individuals seek to increase their education levels in order to increase their future (long run) real income. From this perspective, positive and significant coefficients of education variables in our models might also be interpreted as a search for higher real income levels by the migrants, but in an inter-temporal context. Still, to what concerns our goals, the objectives of migrants who seek higher levels of immediate real income may be considered distinct from the goals of the individuals who seek long term real income increases, since one municipality education level and its job market are different drivers, and should be considered separately as drivers of migration[61].

Therefore, interpreting this evidence regarding education along with the previous results from the other explanatory variables, we find that the Amazon immigrants seem to present a pattern of migration which is more oriented towards the search of better immediate (or in the short run) socio-economic conditions, measured by real income, unemployment and urbanization, with education (or the search for higher real income in the long run) appearing as a secondary goal to these individuals. Such pattern is in opposition to the one presented by the average Brazilian migrants according to our results, since this latter pattern seem to be more oriented towards cities with higher levels of education, and less oriented towards cities with higher immediate real income or unemployment.

---

[60] As the coefficient for the Amazon dummy interaction (α) is negative, while the sum of this dummy with the percentage of population with college degree in the origin for all brazilian migrants ($\alpha + \beta$) remains positive and significant according to the Wald test applied.

[61] In policy implication terms, for example, if government wants to reduce migration flows towards a city which receives many individuals seeking higher immediate levels of real income, a policy by which the city's educational level will be changed will probably be relatively inneffective.

Corroborates this evidence the results from table 4.14, which show that the share of population with college degree in the Amazonian cities of destination is about half the share of this same kind of population in the origins (located at the rest of Brazil) of these immigrants. Furthermore, table 4.15, which refers to illiteracy in different locations in Brazil, also corroborates it: although levels of illiteracy in Brazil as a whole tend to be lower than in the destination municipalities within Amazon, if we compare the latter with the levels of illiteracy in the origins of Amazonian immigrants, we find that the percentages of illiterates in these origins and destinations are very close, which by its turn indicates that among the various origins Brazil, Amazon immigrants come from places with higher illiteracy on average. Contributes to this evidence the fact that most immigrants come from the Northeast of Amazon, which also exhibits the lowest levels of education in Brazil.

Table 4. 14: Share of Population with College Degree

In Brazil, weighted by municipalities' populations	In AMZ Destination of migrants coming from RoB, weighted by the destination's inflow of migration	In RoB Origin of migrants going to AMZ, weighted by the origin's outflow of migration	Difference
6.36%	1.65%	3.33%	-1.68%
(5.30%)	(2.07%)	(3.21%)	

Notes: Standard Deviation in Brackets; AMZ - "Amazon region"; RoB - "Rest of Brazil"
SOURCE: IBGE Census data. Own Elaboration

Table 4. 15: Share of Population Illiteracy

In Brazil, weighted by municipalities' populations	In AMZ Destination of migrants coming from RoB, weighted by the destination's inflow of migration	In RoB Origin of migrants going to AMZ, weighted by the origin's outflow of migration	Difference
14.25%	23.52%	20.52%	3.01%
(11.68%)	(11.04%)	(13.06%)	

Notes: Standard Deviation in Brackets; AMZ - "Amazon region"; RoB - "Rest of Brazil"
SOURCE: IBGE Census data. Own Elaboration

This evidence regarding the educational level of municipalities as push-pull factors in the Amazon immigration decision bring an interesting message, if we try to interpret it focusing on the labor market aspect. Apparently, the Amazon immigrants group may be composed by a workforce with intermediary skills in educational terms: on the one hand, workers moving towards Amazon are not the most educated individuals, as college degree results suggest. On the other hand, they neither are the less qualified workers available in Brazil, as they seem to be trying to move away from labor markets with lower qualification. This is an important result, and again evidences an immigration pattern typical of developing destination cities going through a transitional process of urbanization, as it is expected that more economically mature cities tend to exhibit higher specialization of its workforce, or putting into Jacobs (1970) terms, with a higher share of "new dynamic work" already established, which is not the case of Brazilian Amazon yet.

Selection and Spatial Variables: In table 4.8, the distance between origins and destinations variable coefficients exhibit a negative and significant sign, as expected: migration flows tend to occur between nearest cities due to lower transportation costs and higher ease of adaptation.

By their turn, the Inverse Mills ratios representing the average probability of individuals to migrate given their individual characteristics in the origin and in their destinations are also significant, which according to Heckman (1979) and Wooldridge (2002) suggest that sample selection is actually present in the estimations, which by its turn means that the estimation strategy through the use of the hierarchical Heckit models seems to be the correct choice in order to avoid selection bias.

Hierarchical Heckit Second Stage Results: Brazilian Amazon emigration determinants

Tables 4.16 brings the results of the hierarchical Heckit model which focus on Amazon immigration flows[62]. Their interpretation should be made similarly to the ones given to table 4.8 (Amazon immigration regression), and likewise, we discuss the results from each of the two tables altogether, once again grouping these according to the explanatory variables' categories described in the "Variable Group" column.

Income end Employment: results regarding the influence of real income on table 4.16 indicate a first evidence of a distinct average behavior between individuals who immigrate to the Amazon, analyzed in the previous section, and those who emigrate from this region. First, however, we note that with respect to the general Brazilian migrant whose origin is not located in Amazon, results are very similar to those observed for this same brazilian average migrant in table 4.8: once again, migration flows in Brazil as a whole (excluding the one originated within Amazon borders) are greater the higher is the nominal income in these migrants destination cities, whereas coefficients of nominal income and living costs in their origins are not significant at all (and neither are the coefficients of living costs at the destination). However, it is noteworthy that the signs of the variables cost of living are identical to those observed in the immigration decision regression. Thus, from both

[62] As a reminder, this means that it brings the results from the second stage Heckit regressions which includes all migration flows in Brazil as observations, but which also includes interaction and intersection dummies of the Amazon immigration flows for all explanatory variables, as explained in the methodology section.

emigration and immigration results, we have obtained the same evidence regarding the average Brazilian migration flows (excluding those which have Amazon as their origins or destination): migration flows are higher towards destinations with higher real income[63], even tough real income seems to be a less important driver of migration for the average brazilian migrants, than it is for immigrants heading towards Amazon, who tend to go for cities with higher levels of real income both in terms of higher nominal income and lower costs of living.

On the other hand, for the Brazilian Amazon emigration flows, an opposite evidence is observed regarding real income: coefficients regarding income and costs of living variables in table 4.16 suggest that migration flows of individuals who leave the Amazon region are greater the higher the costs of living in their destinations, and the smaller the nominal income in these same destinations. These is an opposite and unexpected result in theoretical terms.

Such unexpected result is hard to justify, since there is no theoretical background which would support this result. Therefore, once again we conservatively interpret this result as an econometric possibility by which nominal income and costs of living might be capturing the effects of other unobserved (in our model) explanatory variables. It is possible, for example, that costs of living may be reflecting quality of life indicators[64], since it seems reasonable to believe that there is a price to paid in exchange for better quality of life conditions. In this sense, it is possible to interpret this result as if Brazilian Amazon emigration flows occur due to other unobserved variables positively correlated to real income, and real income is simply reflecting the effects of these unobserved variables. In this sense, we may argue that Amazon emigration flows are occurring despite the real income differential, that is, migrants who leave Amazon tend to do so due to reasons other than the search of real income, and in fact, they do so even if real income at their destinations is small, in the sense that the other reasons which drive his decision more than compensate to them.

Or, put in a more conservative way, we may argue that such unexpected result indicates that in fact, immediate (or in the short run) real income differentials is probably not among the most important variables when it comes to explaining Amazon emigration flows. In fact, as

[63] As expected by neoclassic, NEG and other theories revised here

[64] Since we are aware that our explanatory variables do not include all quality of life indicators on each municipality, since many of those are not available for every brazilian municipality (such as air pollution levels or average distance to work). In an attempt to include one measure of life quality, we have tried to include the HDI-M index in the year 2000 as a explanatory variable. However, as such index is composed by income, health and education, it proved to be highly correlated to the other explanatory variables included in our estimations, and interpretation of the coefficients became dubious and hard due to that inclusion. Therefore, our estimations do not include such index.

we will see in the next paragraphs, results from the other explanatory variables suggest that the Amazon emigration flows in the period considered tend to respond more to education (or, as previously discussed, the search for higher real income in the long run), urbanization and health variables than they do to real income, thus corroborating such argument line.

Table 4. 16: "Hierarchical Heckit" Second Stage - Migration in Brazil focusing on the Amazon Emigration Flows

Explanatory Variable	Brazil		AMZ Emigration Interaction Dummies		(Ho: β + α = 0) Joint Test, p-value	Variable Group
	Coefficient β	*t-stat*	*Coefficient α*	*t-stat*		
Avg. Labor Income (Destination) (log)	5.202**	2.4	-19.124***	-5.9	0.0000***	**Income and Employment**
Avg. Labor Income (Origin) (log)	-2.465	-1.5	1.228	0.4	0.5904	
Avg. Cost of Living (Destination) (log)	2.199	1.5	9.580***	3.7	0.0000***	
Avg. Cost of Living (Origin) (log)	-1.908	-1.4	-0.452	-0.2	0.2657	
Unemployment rate (Destination)	15.540**	2.4	-55.393***	-5.0	0.0000***	
Unemployment rate (Origin)	-17.385***	-3.5	-1.597	-0.2	0.0188**	
% of Pop. w/ Access to Sanitary Treatment (Origin)	7.357***	6.7	3.982	1.0	0.0029**	**Health**
% of Pop. w/ Access to Sanitary Treatment (Destination)	-2.099	-1.2	5.380**	2.3	0.0344*	
% of Labor Force in Services and Commerce (Origin)	69.574***	22.8	-48.950***	-8.3	0.0000***	**Urbanization & Age**
% of Labor Force in Services and Commerce	42.932***	9.2	-25.684***	-4.2	0.0000***	
Pop. Density (Origin)	0.001**	2.3	0.009	1.4	0.0988*	
Pop. Density (Destination)	0.002**	2.5	-0.003***	-2.9	0.1519	
% of Pop. aged 22 - 29 (Origin)	281.720***	8.3	-268.206***	-5.2	0.7256	
% of Pop. aged 55 - 60 - "about to retire" (Origin)	-549.465***	-8.0	541.144***	3.4	0.9537	
% of Pop. aged more than 60 - "retired" (Origin)	-434.157***	-13.1	332.905***	4.7	0.1060	
% of Pop. aged 22 - 29 (Destination)	114.928***	3.8	296.577***	4.9	0.0000***	
% of Pop. aged 55 - 60 - "about to retire" (Destination)	-261.265***	-3.4	207.314*	1.9	0.4853	
% of Pop. aged more than 60 - "retired" (Destination)	-491.062***	-14.1	228.142***	3.9	0.0000***	
% of Iliteracy (Origin)	50.482***	9.6	-26.836***	-3.2	0.0003***	**Education**
% of Iliteracy (Destination)	4.008	0.6	-3.005	-0.3	0.8950	
% of Pop. w/ College Degree (Origin)	236.383***	9.7	-247.148***	-5.1	0.7998	
% of Pop. w/ College Degree (Destination)	329.208***	13.0	-112.151***	-2.7	0.0000***	
Average Mills Ratio - "propensity to migrate" (Origin)	-129.531***	-23.5	102.438***	12.6	0.0000***	**Sample Selection**
Average Mills Ratio - "propensity to migrate"	-24.691***	-2.6	-18.715*	-1.7	0.0000***	
Centroid Distance (Origin - Destination) (log)	-24.720***	-28.1	18.612***	16.3	0.0000***	**Spatial**
Constant term	341.410***	15.1	-195.599***	-5.5	0.0000***	
per capita GDP (Destination) (log)	-3.652***	-3.5	2.355*	1.8	0.1166	**Other "control" Variables**
per capita GDP (Origin) (log)	-0.884	-1.4	1.149	1.2	0.7098	
EAP / WAP ratio (Destination)	25.251***	6.4	-14.573**	-2.3	0.0254**	
EAP / WAP ratio (Origin)	22.557***	5.9	-14.650**	-2.1	0.1646	
Previous (1991 - 2000) Labor Income Growth	-5.230***	-5.0	6.718***	4.6	0.1352	
Previous (1991 - 2000) Labor Income Growth (Origin)	-1.016	-1.5	1.508	1.3	0.6016	
Gini Index (Origin)	-53.016***	-6.2	56.140***	4.2	0.7642	
Gini Index (Destination)	10.739	0.8	-5.896	-0.4	0.6084	
Previous (1991 - 2000) Migration Inflow (Destination)	10.987	1.1	-32.106***	-2.6	0.0041***	
Previous (1991 - 2000) Migration Inflow (Origin)	-6.666	-1.3	4.593	0.7	0.6369	
% of Pop. Born in the City (Destination)	7.899	0.8	25.903**	2.3	0.0000***	
% of Pop. Born in the City (Origin)	122.809***	21.3	-105.719***	-13.6	0.0010***	
Pop. (Origin)	0.000***	9.1	-0.000***	-3.5	0.3515	
Pop. (Destination)	0.000***	8.3	-0.000***	-4.0	0.1210	
Number of Academics (2005) (Origin)	-0.002***	-7.3	0.002***	3.5	0.6278	
Number of Academics (2005) (Destination)	-0.002***	-7.3	0.002***	3.8	0.3469	

Note 1: *** p<0.01, ** p<0.05, * p<0.1

Note 2: All variables refer to the year of 2000, except when described in the variable's name.

SOURCE: Own Elaboration

In order to try confirm if this interpretation may be correct, we firstly have made a comparison between the average[65] costs of living in the origins cities within Amazon with the average costs of living in destinations of Amazon emigrants located in the rest of Brazil (see the middle column in tables 4.17 and 4.18). This comparison shows one evidence in favor of this interpretation: average costs of living in destinations are indeed found to be larger[66]. However, we also compared the average real and nominal incomes from these same two types of origins and destinations (tables 4.17 and 4.18 once again), and found that the average nominal income in these destinations is also larger, and in fact larger enough to make average real income in these destinations also greater than in the origins. Thus, the results concerning real income remain controversial concerning Amazon emigrants, in terms of what it is to be expected by theoretically.

Table 4. 17: Average Labor Income and Cost of Living in RoB destinations, weighted by the destination inflow of migration

Average Labor Income (R$ / Month)	*Cost of Living (R$/month)*	*Real Income (R$ / Month)*
483.37 (216.40)	188.24 (95.08)	295.13

Notes: Standard Deviation in Brackets; AMZ - "Amazon region"; RoB - "Rest of Brazil"

SOURCE: Own Elaboration, IBGE Census and PNAD Data

Table 4. 18: Average Labor Income and Cost of Living in AMZ origins, weighted by the origin's outflow of migration

Average Labor Income (R$ / Month)	*Cost of Living (R$/month)*	*Real Income (R$ / Month)*
364.56 (170.62)	122.78 (46.84)	241.78

Notes: Standard Deviation in Brackets; AMZ - "Amazon region"; RoB - "Rest of Brazil"

SOURCE: Own Elaboration, IBGE Census and PNAD Data

However, in terms of our goals in this study, we can interpret this result as an evidence that whereas Amazon immigration flows are more clearly influenced by higher immediate real income differentials between their cities of origin and destination, as it is to be expected by

[65] Once again, we highlight that these averages were weighted by the migration flows instead of the populations of each municipality, for the reasons previously discussed.
[66] Once again, we calculate these averages using the flows as the weights, in order to give more importance to cities that provided or received more migrants on calculations, in accordance to the argument discussed previously.

literature (reviewed previously), Amazon emigration flows (as well as the rest of the brazilian migration flows), on the other hand, are much less influenced by this such income differential in the short run. In fact, as we will see in the evidence brought the other explanatory variables, instead of being determined by real income differentials in the short run, Amazon emigration flows seem to be more oriented towards cities with higher levels of education, which as we have previously pointed out, may be interpreted as a search for higher real income differentials in the long run, since according to Human Capital theories (Becker, 1993; Borjas, 1989), higher levels of education increase the individuals' income flows throughout the rest of his life, but reduces his income level during the period by which such education is being acquired.

Regarding the intensity of Amazon emigration flows in response to unemployment at the origins and destinations, the pattern observed for Brazilian migrants is once again the same as the one found in tables 4.8: surprisingly, the coefficients indicate that migration flows are stronger the smaller the unemployment level in origin and the lower the unemployment in the destination. As for the emigrant coming from Amazon, we find that lower unemployment in destinations act as a pull factor, although the inverted signal with respect to the origin remains similar to those from migrants in general, that is, negative. Comparing the average values (weighted by migration flows, in table 4.19) of unemployment on the origin and destination of emigrants from Amazon, we see that the level of unemployment is indeed lower in Amazon origins, and yet, emigration still occurs.

Thus, results of emigration regressions concerning unemployment, as the results concerning real income, are controversial and unexpected in theoretical terms. Nevertheless, in this case, one important data aspect might be held responsible: unemployment is a variable observed only for the year of 2000[67] , and migration flows considered between for the 2005-2010 period. As unemployment is a variable that can change rapidly in the short run, it is possible that changes in this variable levels on municipalities may have occurred between 2000 and 2005, possibly explaining the unexpected signs in regressions. Of course, this argument is also valid for the immigration regressions, where results concerning unemployment were more coherent with traditional economic theory.

[67] If we wish to include informal labor and also all municipalities in Brazil, because this kind of data is only available in brazilian Census data every 10 years.

That being put, once again, a more conservative interpretation leads us to infer that, in respect to unemployment, as in the real income variables case, Amazon immigration flows respond differently than do its emigration ones: once again, based on the fact that result are incoherent with economic theory, we conclude that Amazon emigration occurs in spite of unemployment, and in that sense, unemployment seems to be less important to explain Amazon emigration than it is for Amazon immigration (due to the same arguments previously discussed for the case of real income variables).

Table 4. 19: Average Unemployment Rate

In RoB Destination of migrants coming from AMZ, weighted by the destination's inflow of migration	In AMZ Origin of migrants going to the RoB, weighted by the origin's outflow of migration	Difference
13.11% (5.75%)	10.64% (6.18%)	2.47%

Notes: Standard Deviation in Brackets; AMZ - "Amazon region"

Source: IBGE Census data. Own Elaboration

Health: Regarding the access to sanitary treatment in the origins and destinations, coefficients with respect to brazilian migrants in general are very similar (and as controversial as) to those in the Amazon immigration analysis (table 4.8), showing that flows tend to be higher the higher the share of population with access to sanitary treatment at the origin, a result somewhat unexpected considering that, in general, individuals are not expected to leave municipalities with better sanitary conditions. Regarding the decision to emigrate from Amazon, results are less controversial, since on the one hand, higher shares of the population with access to sanitary treatment at the destination municipalities act as a pull factor, as expected, but on the other hand, as in the case of general brazilian migrants, higher shares of population with access to sanitary treatment at the origin also result in higher migration flows. Therefore, as in the case of Amazon immigration (table 4.8), we conclude that evidence regarding health conditions of the municipalities as determinants of migration is inconclusive from our estimations, which leads us to conservatively assume that if health plays a role in determining migration flows, this is not being properly captured by our estimations.

Urbanization e demographic variables: results regarding urbanization in origins and destinations in emigration regressions confirm the previous results for brazilian migrants in general observed in table 4.8: in Brazil, higher migration flows to occur between locations with high urbanization rates. Specifically concerning Amazon emigration, results from table 4.16 also show that the average emigrant coming from Amazon follows this trend as well, when we measure urbanization by the percentage of the population working in the tertiary sector.

More importantly than that, in terms of what concerns our analysis, is that this is a distinct pattern of behavior compared to the one observed for the Amazon immigrants. As a reminder, we have observed that Amazon immigrants also tend to move towards more urbanized centers, but coming from both more or less urbanized origins. As for Amazon emigration, the migrants' origins are typically more urbanized only. This result is somewhat expected, simply because urbanization is at different stages within Amazon and in the rest of the country, which means that migration flows towards Amazon most likely follow a growing urbanization pattern, whereas emigration simply occurs between two highly urbanized areas.

With regard to the influence of the population's age structure in migration flows, once more the standard for brazilian migrants in general found in the previous immigration regressions are corroborated, and also for the Amazon emigrants: flows tend to be more intense between origins and destinations that have a higher percentage of young people instead of elderly population, according to the coefficients analysis. It is important, at this point, to emphasize that this is also the result previously found for immigrants of Amazon, and thus, in this sense, there is a similarity in patterns of Amazon emigrants and immigrants. However, for emigrants from Amazon, this result might be considered slightly weaker in terms of robustness: Wald tests on the sums of interactions with "brazilian" coefficients" indicate that municipalities in Brazil with the highest share of young people tend to attract larger migratory flows, as well as municipalities with the highest percentage of people over 60 years tends to attract fewer migrants, on the one hand. On the other hand, these same variables are not significant in the origins, implying that the outflow does not necessarily occur in origins with more young people within the Amazon region, as they do in the rest of Brazil.

Education: along with the results found for real income and unemployment variables, results related to the educational level in the origins and destinations indicate the second main difference found in this study between the behavior patterns of the Amazon immigrant and the Amazon emigrant, concerning the importance of these variables as drivers of migration. First, however, we emphasize that for the brazilian migrants in general, once again the results from emigration regressions are very similar to those found in the immigration models (table 4.8), which gives robustness to such results: higher illiteracy rates at the origin act as a pull factor, as expected theoretically, and a larger share of the population with college degree in both origins and destinations both act as a push factor, indicating that this is a kind of population with greater mobility in Brazil as a whole, which might be justified according to the arguments already mentioned previously by LeSage (2005 and 2008).

In turn, the results for the Amazon emigrant concerning these variables are partially distinct from the ones found for the average brazilian migrant: on the one hand, illiteracy in the origins (within Amazon) also act as a push factor. However, only the share of population with college degree at the destination (in the Rest of Brazil) remains positive and significant for individuals leaving Amazon. This indicates that, unlike the case for the Amazon immigration, the Brazilian Amazon emigration flows are more clearly oriented towards places with higher levels of education in both basic (lowest illiteracy) and advanced (greater share of college degree) levels. Such evidence is confirmed by comparing the averages of illiteracy and the share of population with college degree in the origins and destinations of Amazon migrants (once again weighted by the migration flows), as in tables 4.20 and 4.21.

Table 4. 20: Share of Population with College Degree

In RoB Destination of migrants coming from AMZ, weighted by the destination's inflow of migration	In AMZ Origin of migrants going to the RoB weighted by the origin's outflow of migration	Difference
4.07% (3.65%)	1.18% (1.54%)	2.89%

Notes: Standard Deviation in Brackets; AMZ - "Amazon region"

SOURCE: IBGE Census data. Own Elaboration

Table 4. 21: Share of Population Illiteracy

In RoB Destination of migrants coming from AMZ, weighted by the destination's inflow of migration	In AMZ Origin of migrants going to the RoB weighted by the origin's outflow of migration	Difference
17.49% (12.08%)	25.86% (10.44%)	-8.36%

Notes: Standard Deviation in Brackets; AMZ - "Amazon region"; RoB - "Rest of Brazil"

SOURCE: IBGE Census data. Own Elaboration

As expected, the average share of the population with higher education level in destinations located in Rest of Brazil is about four times higher than the average education level within the origins in Amazon, which corroborates the evidence that individuals leave Amazon in search for higher educational levels. Moreover, comparing tables 4.14 and 4.15 with tables 4.20 and 4.21, we find that the difference in the share of population with College Degree between destinations (located in the rest of Brazil) and the origins (located in Amazonia) of Amazon emigrants is two times greater than the difference between the origins (in the rest of Brazil) and destinations (Amazon) of the this same variable of Amazon immigrants. This, once again, indicates that the search for municipalities with more advanced levels of education indeed seem to be more reasonable for those who leave the Amazon region, than to those who move into it. If we make the same comparison using illiteracy rates instead of share of population with college degree, this is once again corroborated: moving away from cities with lower levels of education makes more sense for those who are leaving the Amazon region than for those arriving at it.

Therefore, the most important evidence we have found regarding the educational level variables as migrations drivers, is that Amazon emigration flows are more clearly oriented towards the search for higher educational levels in their migration decision than are the Amazon immigrants. These differences between trends of immigration and emigration can be interpreted as an evidence that Amazon immigration in the period analyzed (2005-2010) seems to be more clearly oriented by immediate real income differentials (or in the short run), whereas Amazon emigration flows seem to respond more to possible long run real income differentials, following the Human capital theories argument (see Becker, 1993 and Borjas, 1989) by which migrants may move seeking to obtain higher levels of education, in order to earn higher levels of real income in the future (or in the long run).

Still, it important to remark that as the explanatory variable representing education refers to the city level, it is possible that the migrant might be not necessarily be seeking to increase his education level by migrating, but instead, it is possible that the education level of the municipality may be capturing the effects of other unobserved and positively correlated (to education) variables, especially if we take in consideration that many relevant variables cannot be observed or inserted in the regressions. Nevertheless, it is possible to affirm from our results that Amazon immigration is more clearly oriented towards immediate real income, whereas emigration seem to be more oriented to the municipalities education variables, which already represents a significant difference in these two kinds of flows. One less conservative interpretation of these result, which need to be confirmed by different approaches, would be that Amazon immigrants exhibit preferences which are distinct from the preferences of the Amazon emigrant, in the sense that the first is probably more concerned about his income flow in the short run, whereas the latter worries more about his lifecycle income. Still, such evidence needs to be addressed by a more specific approach regarding these variables, probably with econometric models entirely built on the individual level.

Spatial and Selection variables: Results observed for the spatial variable (distance) in table 4.16, and also for the variables included to correct the problem of selection bias, proved to be as expected in theory, and also very similar to the one observed in the immigration regression (tables 4.8).

Specifically, we found that Amazon emigration flows are greater the lower the distance between origin and destination, both for general brazilian migrants and the Amazon emigrant. As we have argued previously, this is an expected result because costs of migration and adaptation tend to be lower among closer distances. Besides, the Inverse Mills representing the average propensity towards migration in the origins and destinations of Amazon emigrants, and also for brazilian migrants in general, are all significant, which according to Heckman (1979) means that selection sample bias might be present if these were not included in the analysis, which justifies the methodology chosen in this work.

4.6. Conclusions and final remarks

In this chapter we have tried to find evidence on which are the main drivers of the migration flows between Brazilian Amazon and the rest of Brazil in the period 2005-2010, comparing the main differences and similarities between the determinants of immigration to and emigration from Brazilian Amazon. The main goal of such study is to contribute to the better understanding of the recent population growth and urbanization process which the region has been going through in the last few decades, since migration flows are historically closely correlated to population dynamics. In order to do so, we have implemented an Hierarchical Heckit econometric model, which includes explanatory variables such as real income differentials, educational levels and other origins' and destinations' characteristics. Our estimations tried to account for the sample selection problem regarding migration as the dependent variables, as pointed out by literature, which may happen if migrants and non-migrants exhibit differences in skills (see Chiswick, 1999 & 1978; Borjas, 1994; dos Santos, Menezes & Ferreira, 2005; Sjastaad, 1962).

In this sense, a first simple descriptive analysis of the database provided the first important evidence regarding migration in Brazilian Amazon. More specifically, we have calculated the fixed date migration flows between 2005 and 2010, and compared the size of them with the fixed date flows between 1995 and 2000, using Census data. Surprisingly, we found that the size of the migration flows involving the Amazon regions have slightly declined from the previous period to the more recent one. Moreover, Amazon emigration flows were also slightly bigger than the Amazon immigration flows from 2005 to 2010. Therefore, since both the region's population and urbanization rates continued to grow throughout the decade of 2000, this suggests that such population growth is currently more based on an internal vegetative growth dynamics, that is, on the reproduction of the population who already have been living in Brazilian Amazon. Such evidence may be considered important when it comes to the Amazon region, since the region had been historically occupied by big waves of immigration flows, especially until the 1990 decade. Nevertheless, current migration should still be regarded as an important element of such population growth and urbanization dynamics, since it may change the local population characteristics in the long term, which by its turn may affect the future vegetative growth trend. Therefore, we have proceeded with the analysis and comparison of the main drivers of Brazilian Amazon immigration and emigration flows.

In this sense, our econometric results point that Amazon immigration flows (from the rest of Brazil) drivers can be considered essentially different from what have been determining emigration from this region to the rest of Brazil. Specifically, they suggest that Amazon immigration flows are mostly driven by young population coming from both more and less urbanized cities in Brazil, in majority located at the poorest Northeast macro-region of the country, and moving towards the more urbanized centers of Amazon. According to our findings, short run real income differentials between destinations and origins act as one of the main drivers of such immigration, as most of the economic theories would predict. On the other hand, Amazon emigration flows are also mostly composed by young individuals moving to highly urbanized municipalities of Brazil. However, they also tend to leave more highly urbanized cities of Brazilian Amazon. Furthermore, in opposition to Amazon immigration flows results, Amazon emigration does not seem to be driven by short run real income differentials between origins and destinations, but instead, our results point that such emigration flows are mostly driven by the differentials in basic and superior education levels of the migrants' origins and destinations. Such result does not necessarily mean that these migrants move seeking to increase their own educational levels. In fact, it is possible to interpret this evidence in two different ways: a) it is possible that these migrants are in fact searching to increase their education level, as Human Capital theories would predict (see Becker, 1993 and Borjas, 1989), in order to elevate their future (or long run) real income earnings; b) it is possible that the municipalities' education level might be capturing the effects of other unobserved (and positively correlated to education) explanatory variables which were not included[68] in our estimations, such as culture or infra-structure.

Nevertheless, in a conservative interpretation, our results suggest that Brazilian Amazon immigration and emigration can be considered essentially different in terms of their main determinants. And such differences are to be taken in consideration when it comes to analyzing future population and urbanization trends of the region, especially considering that the urbanization process going on in Brazilian Amazon will probably keep its pace throughout next decades, since the region’s urbanization rates still have not reached the same levels of the average brazilian ones. And as discussed in previous chapters, such urbanization tend to be accompanied by economic growth and development, which by its turn tend to increase the region’s relative real income level. This, by its turn, may be interpreted as a greater incentive to increase immigration towards the region, as long as the Amazon immigration flows response to short run real income differentials remains as observed in our results for 2005-2010.

[68] Due to data unavailability at the municipality level.

5. CONCLUSION AND POLICY IMPLICATIONS

Brazilian Amazon presents itself as a complex economic and environmental maze, due to its immense geographical size (over 5 million square kilometers) and its large socio-economic and environmental diversity. As Andersen et al (2002, pp. 200) point out, it seems unlikely that economic research by itself will be able to entangle all causes and consequences of its deforestation process. Nevertheless, this dissertation brings new perspectives and empirical evidence to some relevant aspects of such complexity.

More specifically, the findings of this thesis' two first essays characterize an ongoing tradeoff between economic development and deforestation, caused by urbanization and local population growth within the Brazilian Amazon during the previous two decades. It is true that evidence regarding such tradeoff had already been uncovered by previous literature, as in Igliori (2006), in which econometric estimations point that Amazon local economic agglomerations (which are closely related to urbanization) are non-linearly correlated with both economic growth and deforestation, affecting both variables likewise (in terms of signal) throughout the 1990 decade. Andersen et al. 2002 also argue (and find evidence) that from 1970 to 1995, the sprawl of local urban markets and the increase in the extension of local paved roads have played a significant role in terms of increasing the region's GDP, alongside with rainforest deforestation.

Nevertheless, this thesis contributes in three relevant ways to help disentangle and uncover relevant characteristics of this tradeoff. Firstly, it updates empirical evidence regarding it, until the year of 2010, showing not only that the tradeoff still persists, but in fact, it has increased in the last decade. Secondly (and mainly), it proposes a different approach to the analysis of local deforestation causes, focusing on the role of local population and urbanization as deforestation drivers, motivated by the fact that these have grown expressively within the region over the last 40 years. Thirdly, the methods applied address deforestation drivers from a "demand-side of the economy" viewpoint, that is, by measuring how much deforestation can be consider due to the local population consumption of goods and services, and how much local urbanization currently contributes to expand such consumption and (i.e.) deforestation. Moreover, our method also measures the size of demand-side deforestation drivers belonging to external markets, specifically from Brazilian regions outside the Amazon area, and also the ones belonging to abroad countries. Therefore, we are able to compare the sizes of local and abroad demand-side drivers of Amazon deforestation. Put in simpler terms, our analysis addresses deforestation as driven by the agents who consume the output produced in (previously) forest covered areas, and assign shares of responsibility for deforestation geographically, according to how much each region "consumes" the inputs and outputs produced in those areas. Such perspective is somewhat different and complementary to the mainstream of Amazon economic literature (so far), in which deforestation drivers are mostly analyzed from a "supply-side of the economy" perspective, and cattle raise and soybean production are pointed as the main drivers of deforestation, but the geographic destinations of such outputs (or who consume them) are generally overlooked.

In that sense, evidence from the spatial econometric approach applied in the first essay point out that the undergoing process of urbanization of Brazilian Amazon can be held partially responsible, among other factors, for the region's recent economic growth and development (represented by employment, per capita GDP and the HDI index). And even though results also point that such growth is still partially dependent of external markets, this result regarding urbanization and development suggests that the region's local economy seem to be moving towards a relatively endogenous growth path, thus probably reducing such external dependency in years to come, and increasingly relying its economic development on the dynamics of its own internal markets.

On the other hand, such urbanization and local population growth have been also contributing to increase local deforestation, according to the results achieved in the second essay. More specifically, evidence found show that from a "demand-side of the economy" perspective, even though Brazilian Amazon local population accounts for only 13% of total Brazilian population, it drives around 27% of the overall Amazon yearly deforestation. Also, families who live within the region's Metropolitan areas are responsible for more than a half of such deforestation, even though only 25% of the Brazilian Amazon population actually live within these areas. In per capita terms, our calculations point that the average demand of goods and services from one individual living within the Amazon region, but outside its Metropolitan areas, generates 2.2 more deforestation than one individual living outside Brazilian Amazon. Moreover, the consumption vector of one individual living within the Amazonian Metropolitan Regions causes a deforestation impact 7.7 times higher than the impact of one individual living outside the Brazilian Amazon.

Historically contextualizing, this trade-off may be interpreted as a primary result of intense government occupation policies between 1964 and 1990 (see Becker, 2013). In this period, Brazilian federal government promoted incentives to occupy and develop the region, mainly in order to protect the country borders. Infrastructure investments and paved roads construction increased drastically within the area, and resulted in a rapid subsequent population expansion (and Andersen, 2002). These policies have practically ceased in the early nineties, however, population growth and urbanization intensification persist until the present date. This suggests that current Brazilian Amazon occupation became relatively more endogenous in the last 20 years, that is, it became more dependent on the internal dynamics of local population reproduction (vegetative growth), as well as on the orientation of the region's migration flows towards its urban centers.

This suggestion is confirmed by results found in the third essay of this dissertation, in which a simple description of the 2010 Brazilian demographic census data shows that immigration flows intended to Brazilian Amazon are currently in balance with (or, in fact, slightly smaller than) the number of emigrants leaving the region[69] towards other Brazilian centers. This, by its turn, indicates that local vegetative growth was the main reason for the large increase in population experienced by Brazilian Amazon throughout the last decade, evidencing that the region has already reached the critic point in which a high rate of local population increase depends more on the current size of the population itself, than on flows of immigration towards the region.

Moreover, the econometric analysis regarding the determinants of Brazilian Amazon migration flows implemented in this same essay also show that the region's urbanization process plays an important role in terms of explaining the increasing endogeneization process of the region's economic growth: both kinds of migrants analyzed (Amazon immigrants and Amazon emigrants), that is, both groups of individuals who leave and who arrive at the Amazon region, tend, on average, to migrate towards more densely urbanized centers, indicating the average set of preferences of individuals who migrate in Brazil are in favor of increasing the urbanization rates all over the country, including Brazilian Amazon. This, as evidenced in the other two essays, tends to boost both deforestation rates and the endogenous economic development of the region. Complementarily, our results from the third essay also suggest that Amazon immigrants' motivation to move is more clearly leaned towards gains of real income in the short run, whereas its emigrants seem to be more clearly motivated by

[69] As described in chapter 4, these calculations consider only "fixed-period" internal (to the country) migration flows of individuals over 18 years old from 2000 to 2005.

opportunities to increase their real income in the long run, by moving into cities which present higher current educational levels.

Given such endogenous nature of the tradeoff between economic growth and deforestation presented, future perspectives for Brazilian Amazon regarding environmentally sustainable growth tend to become an increasingly complex matter. On the one hand, locally, despite all recent economic growth described, the region still presents itself as the current second poorest macro region of Brazil, with 17 million (out of 34 million) individuals considered to be living below the poverty line. On the other hand, globally, the region's increasingly endogenous process of economic growth is being accompanied by a local population rise, alongside with intense urbanization, which altogether tend to increase the Amazon overall deforestation in years to come. Therefore, this represents a negative externality in global scale. In terms of future implications, such scenario may present itself as a choice-based situation for policy makers, in the extreme case where no sustainable solution may be achieved. In such a case, this tradeoff becomes considered as inevitable, policy makers will be forced to choose between trying to avoid the global issue of deforestation, or trying to promote local economic growth. In this situation, it seems reasonable to argue that both history and economic theory would most likely predict that the choice would be made in favor of economic growth, thus, in detriment of avoiding deforestation.

Historically, as previously cited, policies were mostly oriented towards creating incentives for growth and occupation, especially from 1964 until 1990. But even after the rising of the environmental concern among Brazilian citizens and policy makers (in the early nineties), which have caused such explicit occupation policies to cease, the majority of policies implemented thereafter have focused on trying to stimulate both economic growth and environmental conservation at the same time, that is, policies adopted were not oriented towards preserving the forest in detriment of local economic growth.

A recent debate regarding the reform of the Brazilian Forest Code, created originally in 1965, helps to illustrate this point: for over the last 13 years (therefore, after the environmental concern upraise regarding the conservation of the Amazon forest), there has been a struggle in Brazilian government between "environmentalists" (policy makers in favor of forest conservation or at least trying to promote sustainable practices) against "farmers" (policy makers who are more clearly in favor of boosting agriculture and other economic activities, even if it prejudices forest conservation). To which regards Brazilian Amazon, the most controversial topic of discussion was around the maintenance of the legal requirement for landowners to preserve 50% of their properties' original forest covered areas as legal reserves (as approved in the original Code, in 1965), versus the possibility of increasing this percentage up to 80%. Such point of the original Forest Code has never been effectively applied, and most Amazon farmers, in fact, previously had cleared over 50% of their forest covered areas. Nevertheless, penalties regarding this infraction have poorly been applied so far. Even though an official law project has already been created in order to implement the reform, and its content has already been discussed and modified several times by each side, the debate persists until nowadays, and no changes have been properly implemented yet. Either way, the last version of the law project content is considered, from the "environmentalists" perspective, as a victory won by the "farmers'" side, as argued in Morello (2012), ESALQ (2012), IPEA (2011) and SBPC (2012).

Economically, to attribute higher probability of policy makers to choose economic growth over forest conservancy is also perfectly justifiable. Such a choice is a matter of collective bargaining in which environmental costs of deforestation are divided globally (that is, they will be borne by the entire world's population) on the one hand, and the socio-economic opportunity costs of forest conservation are borne mostly by the local Amazonian families. As the brazilian government policy makers are elected locally, political economy models based on the classic "median voter theorem" (see Black (1948)) predict that if such a choice had to be made, it is more likely that policy makers would try to attend to the preferences of the Amazonian median voter, whose share of socio-economic costs of forest conservation is probably greater than his share of environmental costs of deforestation, since the latter is divided by a larger number of individuals. A similar argument is presented by Andersen et al (2002, pp. 203-204), in which authors argue that the main reason by which Brazil have ignored the negative externalities associated to deforestation until recently is because such negative externalities are divided by all world citizens, whereas the positive externalities of forest clearing mostly benefit only the local population.

Fortunately, a future scenario in which a choice between local economic growth and environmental conservation has to be made by policy makers may still be considered as unlikely to happen in the next few years. Or, put in other terms, it seems reasonable to believe that there still may be room for implementing sustainable policies which permit both economic growth and conservation in Brazilian Amazon. And the reason to conclude so is based on a descriptive analysis of databases from PRODES (INPE[70]) regarding deforestation of Brazilian Amazon, along with recent literature review regarding environmental sustainable growth policies which have been implemented throughout the last decade within the region. More specifically, according to PRODES database, 78% of the original Amazon forest covered area still remained preserved by the end of 2013, in Brazil. Furthermore, deforestation rates have been decreasing significantly since the 2004[71]. According to Assunção, Gandour & Rocha (2012), who have applied econometric methods to infer the main causes of such decay, around 50% of this downfall may be attributed to price fluctuations in crops and cattle markets, whereas the other 50% are due to policies of conservation and sustainable growth introduced in 2004 and 2008 by the Brazilian Government.

Even though such decrease in deforestation rates may be interpreted as an improvement of perspectives in terms of future forest conservation, there certainly is still much to be done in terms of promoting sustainable growth in Brazilian Amazon, in its strict sense. A closer look in the set of policies implemented in 2003, 2004 and 2005 (to which, in general, a great share of the decrease in deforestation rates is attributed to) indicates that such policies were mostly based on command and control efforts: since 2004, government has created over 180,000 km^2 of Conservation Units[72] (CU) within the Amazon forest, divided in Indigenous Lands and National Parks. Moreover, it has also improved significantly its capacity of measuring and monitoring deforestation, mainly through the implementation of the Action Plan for the Prevention and Control of Deforestation in the Legal Amazon (*Plano de Ação para a Prevenção e o Controle do Desmatamento na Amazônia Legal* - PPCDAm) (see Assunção, Gandour & Rocha (2012)). And despite the fact that these policies were effective in terms of reducing deforestation, they still cannot be considered ideal in terms of reducing the tradeoff discussed so far in this dissertation. And this is so because the creation of CUs and improved monitoring indeed tend to reduce deforestation on the one hand, but also prejudice economic growth, on the other, since they simply reduce land availability for agriculture and other economic activities within the region, but without encouraging the substitution of land as a production factor by other factors (such as capital and labor), and also without increasing

[70] Instituto Nacional de Pesquisas Espaciais, from Brazilian Ministry of Sciences, Technology and Innovation.
[71] In 2004, Brazilian Amazon deforestation reached 27.772 km^2, whereas in 2013, this rate was 5,843 km^2. Moreover, in this period, the average annual decay of the deforestation rate has been around 14% per year.
[72] Forest areas in which no human intervention of any nature are allowed.

overall factors productivity in order to compensate for the imposed reduction of land availability.

In order to promote sustainable growth, and therefore reduce the tradeoff in the long term, incentive-based policies which conciliate economic activities and conservation are made necessary. Fortunately, in this sense, the set of policies adopted in 2008 by the Federal Government apparently have provided the first steps towards this direction, even though there is still a long way to be covered. Specifically, in this year, a policy which lately became known as the “Green Municipalities” (*Municípios Verdes*) has been implemented, in which Brazilian federal government has imposed credit restrictions to rural establishments located within municipalities which presented the highest deforestation rates in the last three years, or in municipalities with the largest rates of accumulated deforestation (relative to their original forest covered areas). This policy provided incentives for local land owners to make efforts to try to reduce deforestation, but without forcing them to abandon their previously chosen economic activity or their lands. Putting in economic terms, land owners were encouraged to modify their production function, technology and input combinations of land, labor and capital, in order to substitute the use of forest covered areas by newer combinations of the other inputs. Lately, the results of this policy were considered quite satisfactory. Deforestation rates on these previously enlisted municipalities have been monitored in the subsequent periods which followed the credit restrictions imposition, and many municipalities were then excluded from the government “black list” selection, due to significant decays in their deforestation rates. Paragominas, which is one of the most famous examples of cities which have been affected by this policy, has reduced significantly its deforestation rates, due to the combined effort of local government, entrepreneurs and local land owners. The city became considered one of the first cities in Amazon to be known as a “Green Municipality”. Moreover, Assunção, Gandour & Rocha (2012) also find evidence that a great share of the deforestation rates decay observed in subsequent periods may be attributed to this policy.

Unfortunately, incentive-based solutions as the one described above are still very incipient in terms of the Brazilian Amazon. As Andersen et al (2002), Becker (2013) and Igliori (2006) point out, even though other policies such as ecotourism and payment for environmental services (which can serve as possibilities to replace activities which cause higher deforestation) may be feasible in the region, they were practically nonexistent within the region until a few decades ago, and as Andrade Filho (2008) point out, only recently they have gained (little) strength. Rapidly implemented and effective policies which may subsidize these kind of activities, as well as incentives to the development of more environmental sustainable technologies and innovations are goals to be pursued by the Brazilian government and also the world's population, if one wishes to reduce deforestation with the least comprising of local economic activity.

Finally, considering the main results of this thesis, we conclude by arguing that, unfortunately, the set of existing policies in Brazilian Amazon are still far from reaching an environmentally sustainable pattern of economic development. The continuous upraise of local markets caused by the recent (and rapid) increase of the region's population and urbanization rates poses as a threat to the forest in the long run, as the output produced in forest covered areas tend to become increasingly largely consumed by local families, due to lower transportation costs and positive agglomeration externalities. And such threat is even aggravated considering that urbanization affects positively the region's economic growth and development.

Moreover, as this population grows and local economy becomes more endogenous and self-sustainable, the set of policies which may reduce deforestation in detriment of economic growth tend to become less feasible, politically, since the socio-economic costs associated to low economic growth are borne only by local population (voters), whereas the negative externality of forest clearing is divided globally (by voters and non-voters). And this situation is even worsened taking in consideration that local population is still among the poorest of the country. Also, as local markets for cattle raising and other agriculture activities increase, such activities tend to become more profitable locally, and incentive-based policies which try to promote the replacement of such activities by others considered less environmentally hazardous (such as ecotourism or payment for environmental services) also tend to become increasingly less feasible.

Therefore, in order to properly promote the desired sustainable economic development which Brazilian Amazon needs, without causing major irreversible losses to the world's largest forest, it is necessary that more incentive-based policies of sustainable practices are implemented as soon as possible, also combining them with some command and control practices. Furthermore, considering the results brought by this thesis, it is crucial that these future policies take into account the recent and expressive growth of the Brazilian Amazon population, its undergoing and rapid urbanization process, and the relative endogeneization process of the region's economy. Such considerations have to be taken both in terms of elaboration and enforcement of these policies.

In order to contribute to the achievement of such goal, we propose five major guidelines. In fact, the first four have been previously suggested by literature regarding Brazilian Amazon forest conservation (see, for example, Igliori (2002), Andersen et al (2002), Assunção, Gandour & Rocha (2012), Becker (2013) and Andrade Filho (2008)). Thus, in this sense, the fifth guideline is the only one that may be considered as a specific suggestion which results from the contributions of this study, that is, it is the guideline hereby presented that is specifically related to the empirical results found in this thesis. Nevertheless, we consider that the results of such guidelines will bring better results, in terms of achieving sustainable growth, if implemented altogether, and therefore, we have chosen to describe them all in the next paragraphs.

i) Policies should promote the increase (and future maintenance) of forest areas belonging to conservation units, national parks and (indigenous and other kind) natural reserves, alongside with the further institutional strengthening of their borders, so as to limit the long-term expansion of land use by agriculture and cattle raising pastures, which may compete with currently forest covered areas in the years to come.

ii) In the long run, it is important that policies promote incentives to increase land productivity, along with the development of technologies and innovations which may lead to the relative substitution of land use by labor and capital, so that the growing local demand for goods and services may be properly attended by local producers through the intensification of non-land inputs use.

iii) Ideas such as the "National parks agreement", as suggested by Andersen et al (2002, pp. 204-208), should be implemented. As described by the authors, such "agreement" consists in elaborating international payment schemes in order to avoid clearing of the Amazon forest. Specifically, they suggest that, as forest clearing generally results in a negative externality borne by the entire world population (mainly due to biodiversity loss and increases in greenhouse gases emissions), and the opportunity costs of keeping forest covered areas are paid by local producers, then international community should pay local producers in order to maintain the forest, so that these opportunity costs are properly covered, and less incentives for forest clearing are created.

iv) Future policies should encourage sustainable practices in forest covered areas, instead of economic activities of higher environmental impacts (such as cattle raising). Incentives to ecotourism, payments for environmental services and the “green municipalities” policies implemented in 2008 by Brazilian federal government are good examples to be followed.

v) As the innovative general guideline, we suggest that future policies must take into account the expressive growth of Brazilian Amazon local population and the urbanization process undergoing within the region recently. More specifically, since evidence found in this thesis point out that local population growth has been increasingly implying significant deforestation, and such growth is majorly being caused by vegetative growth (as immigration and emigration flows are currently in balance within the region), then it seems crucial that such local population excessive growth must be discouraged in the first place. Furthermore, as shown in Chapter 3, the demand for goods and services from families living in large metropolitan areas of Brazilian Amazon tend, on average, to cause relatively higher deforestation impacts, when compared to the impacts caused by the consumption vectors from families living in small and medium-sized municipalities. On the other hand, metropolitan regions are heavily urbanized areas, and as shown in Chapter 2, higher local urbanization also leads to higher local economic growth. Therefore, one apparently reasonable solution would be to elaborate and implement plans of urbanization for the Amazon region, in which patterns of urbanization leading to large metropolitan areas are to be avoided, and the urbanization of medium-sized and small cities would be relatively encouraged, so that economic development could be pursued more neatly and evenly (geographically), and in an less environmentally impactful way.

6. BIBLIOGRAPHY

ABREU, Maria; DE GROOT, Henri; FLORAX, Raymond. **Space and growth: a survey of empirical evidence and methods**. 2004.

ACEMOGLU, Daron; ANGRIST, Joshua. **How large are the social returns to education? Evidence from compulsory schooling laws**. National Bureau of Economic Research, 1999.

ANDERSEN, Lykke E. (Ed.). **The dynamics of deforestation and economic growth in the Brazilian Amazon**. Cambridge University Press, 2002.

ANSELIN, L. (1988). **Spatial Econometrics: Methods and Models**. Boston: Kluwer Academic Publishers.

ANSELIN, Luc. Spatial externalities, spatial multipliers, and spatial econometrics. **International Regional Science Review**, v. 26, n. 2, p. 153-166, 2003.

ARMSTRONG, Harvey W. Convergence among regions of the European Union, 1950–1990. **Papers in Regional Science**, v. 74, n. 2, p. 143-152, 1995.

ARRAIZ, Irani et al. A spatial cliff-ord-type model with heteroskedastic innovations: small and large sample results. **Journal of Regional Science**, v. 50, n. 2, p. 592-614, 2010.

ASSUNÇÃO, Juliano; GANDOUR, C. C.; ROCHA, Rudi. Deforestation slowdown in the Legal Amazon: prices or policies. **Climate Policy Initiative (CPI) Working Paper**, Pontífica Universidade Católica (PUC), Rio de Janeiro, RJ, Brazil. p, v. 3, 2012.

BARRO, Robert T. et al. Regional growth and migration: A Japan-United States comparison. **Journal of the Japanese and International Economies**, v. 6, n. 4, p. 312-346, 1992.

AZZONI, Carlos R.; CARMO, H.; MENEZES, Tatiane. Índice de custo de vida comparativo para as principais regiões metropolitanas brasileiras: 1981-1999.**Estudos Econômicos**, v. 30, n. 1, p. 165-186, 2000.

BARRO, Robert; MARTIN, Sala-i. X., 1995. **Economic growth**. Boston, MA, 1995.

BECKER, Bertha. **A Urbe Amazonida**, Garamond, 2013.

BECKER, Gary S. **Human capital: A theoretical and empirical analysis, with special reference to education**. University of Chicago Press, 2009.

BLACK, D. (1948). On the rationale of group decision-making. **The Journal of Political Economy**, 56(1), 23-34.

BONJOUR, Dorothe et al. Returns to education: Evidence from UK twins. **The American Economic Review**, v. 93, n. 5, p. 1799-1812, 2003.

BORJAS, George J. Economic theory and international migration. **International migration review**, p. 457-485, 1989.

BORJAS, George J. The economics of immigration. **Journal of economic literature**, v. 32, n. 4, p. 1667-1717, 1994.

CHAGAS, André Luís Squarize; TONETO, J. R. Fatores Determinantes do Crescimento Local–Evidências a partir de dados dos municípios brasileiros para o período 1980-1991. **Pesquisa e Planejamento Econômico**, v. 33, n. 2, p. 349-385, 2003.

CHISWICK, Barry R. Are immigrants favorably self-selected?. **The American Economic Review**, v. 89, n. 2, p. 181-185, 1999.

CHISWICK, Barry R. The effect of Americanization on the earnings of foreign-born men. **The Journal of Political Economy**, p. 897-921, 1978.

CHOMITZ, Kenneth M.; THOMAS, Timothy S. Determinants of land use in Amazonia: a fine-scale spatial analysis. **American Journal of Agricultural Economics**, v. 85, n. 4, p. 1016-1028, 2003.

CONLEY, Timothy G.; UDRY, Christopher R. Learning about a new technology: Pineapple in Ghana. **The American Economic Review**, p. 35-69, 2010.

CROZET, Matthieu. Do migrants follow market potentials? An estimation of a new economic geography model. **Journal of Economic Geography**, v. 4, n. 4, p. 439-458, 2004.

DA MATA, Daniel. **Quais características das cidades determinam a atração de migrantes qualificados?**. IPEA, Instituto de Pesquisa Econômica Aplicada, 2007.

DE CASTRO, Newton; CARRIS, Larry; RODRIGUES, Bruno. Custos de transporte ea estrutura espacial do comércio interestadual brasileiro.**Pesquisa e Planejamento Econômico**, v. 29, n. 3, 1999.

DEFRIES, Ruth S. et al. Deforestation driven by urban population growth and agricultural trade in the twenty-first century. **Nature Geoscience**, v. 3, n. 3, p. 178-181, 2010.

DOS SANTOS JÚNIOR, Enestor da Rosa; FERREIRA, Pedro Cavalcanti; MENEZES-FILHO, N. Migração, seleção e diferenças regionais de renda no Brasil. **Pesquisa e Planejamento Econômico**, v. 35, n. 3, 2005.

DRUKKER, David M.; EGGER, Peter; PRUCHA, Ingmar R. On two-step estimation of a spatial autoregressive model with autoregressive disturbances and endogenous regressors. **Econometric Reviews**, v. 32, n. 5-6, p. 686-733, 2013.

DRUKKER, David M. et al. Creating and managing spatial-weighting matrices with the spmat command. **Stata Journal**, v. 13, n. 2, p. 242-286, 2013.

DRUKKER, David M.; PRUCHA, Ingmar R.; RACIBORSKI, Rafal. A command for estimating spatial-autoregressive models with spatial-autoregressive disturbances and additional endogenous variables. **The Stata Journal**, v. 1, n. 1, p. 1-13, 2011.

DURANTON, Gilles; PUGA, Diego. Nursery cities: Urban diversity, process innovation, and the life cycle of products. **American Economic Review**, p. 1454-1477, 2001.

DURANTON, Gilles. Urban evolutions: The fast, the slow, and the still. **The American Economic Review**, p. 197-221, 2007.

ELHORST, J. Paul. Specification and estimation of spatial panel data models. **International regional science review**, v. 26, n. 3, p. 244-268, 2003.

ESALQ. Agricultura e sustentabilidade: legislação deve garantir agricultura produtiva e ambiente protegido. **Revista Visão Agrícola**, ano 7, v.10, jan.-abr. 2012. Disponível em: <http://www.esalq.usp.br/visaoagricola/>.

FAMINOW, Merle D. Spatial economics of local demand for cattle products in Amazon development. **Agriculture, ecosystems & environment**, v. 62, n. 1, p. 1-11, 1997.

FINDEISEN, Sebastian; SUEDEKUM, Jens. Industry churning and the evolution of cities: Evidence for Germany. **Journal of Urban Economics**, v. 64, n. 2, p. 326-339, 2008.

FINGLETON, Bernard. Equilibrium and economic growth: spatial econometric models and simulations. **Journal of Regional Science**, v. 41, n. 1, p. 117-147, 2001.

FINGLETON, Bernard (Ed.). **European regional growth**. Springer, 2003.

FINGLETON, Bernard; IGLIORI, Danilo; MOORE, Barry. Cluster Dynamics: New Evidence and Projections for Computing Services in Great Britain*.**Journal of Regional Science**, v. 45, n. 2, p. 283-311, 2005.

FREGUGLIA, R. S. **Efeitos da migração sobre os salários no Brasil. 2007**. 2007. Tese de Doutorado. Tese (Doutorado em Economia)–Universidade de São Paulo, São Paulo.

FU, Shihe; DONG, Xiaofang; CHAI, Guojun. Industry specialization, diversification, churning, and unemployment in Chinese cities. **China Economic Review**, v. 21, n. 4, p. 508-520, 2010.

FUJITA, Masahisa; KRUGMAN, Paul R.; VENABLES, Anthony J. **The spatial economy: cities, regions and international trade**. Cambridge, MA: MIT press, 1999.

FUJITA, Masahisa; THISSE, Jacques-François. Economics of agglomeration: Cities, industrial locations, and regional growth. **Cambridge: CambridgeUniversityPress**, 2002.

FUJITA, Masahisa; THISSE, Jacques-François. Does geographical agglomeration foster economic growth? And who gains and loses from it?.**Japanese Economic Review**, v. 54, n. 2, p. 121-145, 2003.

GEIST, Helmut J.; LAMBIN, Eric F. Proximate Causes and Underlying Driving Forces of Tropical Deforestation: Tropical forests are disappearing as the result of many pressures, both local and regional, acting in various combinations in different geographical locations. **BioScience**, v. 52, n. 2, p. 143-150, 2002.

GERMANI, G. (1974). **Sociologia da modernização: estudos teóricos, metodológicos e aplicados a América Latina**. São Paulo, Mestre Jou, 261 p.

GLAESER, Edward L.; SCHEINKMAN, JoséA; SHLEIFER, Andrei. Economic growth in a cross-section of cities. **Journal of Monetary Economics**, v. 36, n. 1, p. 117-143, 1995.

GLAESER, Edward L. Cities, agglomeration, and spatial equilibrium. **OUP Catalogue**, 2008.

GLAESER, Edward L.; RESSEGER, Matt; TOBIO, Kristina. INEQUALITY IN CITIES*. **Journal of Regional Science**, v. 49, n. 4, p. 617-646, 2009.

GLAESER, Edward L.; RESSEGER, Matthew G. The complementarity between cities and skills*. **Journal of Regional Science**, v. 50, n. 1, p. 221-244, 2010.

GUILHOTO, J.J.M., U.A. SESSO FILHO. Estimação da Matriz Insumo-Produto a Partir de Dados Preliminares das Contas Nacionais. **Economia Aplicada**. Vol. 9. N. 2. pp. 277-299. Abril-Junho. 2005.

GUILHOTO, J.J.M.; SONIS, M.; HEWINGS, G.J.D. Linkages and Multipliers in a Multiregional Framework: Integration of Alternative Approaches. **Australasian Journal of Regional Studies** v.11, n.1, pp. 75-89. 2005.

GUILHOTO, J.J.M. **Análise de Insumo-Produto: teoria e fundamentos**. Unpublished Work in Progress. 2009.

GUILHOTO, J.J.M., C.R. AZZONI, S.M. ICHIHARA, D.K. KADOTA, E.A. HADDAD. **Matriz de Insumo-Produto do Nordeste e Estados: Metodologia e Resultados**. Fortaleza: Banco do Nordeste do Brasil. ISBN: 978.85.7791.110.3. 289 p. 2010.

GOLGHER, André Braz; ROSA, Carlos Henrique; ARAUJO, A. F. The determinants of migration in Brazil. **Belo Horizonte: UFMG/Cedeplar**, 2005.

HARBISON, Sarah F. Family structure and family strategy in migration decision making. **Migration decision making**, p. 225-251, 1981.

HARRIS, John R.; TODARO, Michael P. Migration, unemployment and development: a two-sector analysis. **The American Economic Review**, v. 60, n. 1, p. 126-142, 1970.

HECKMAN, James J. Sample selection bias as a specification error.**Econometrica: Journal of the econometric society**, p. 153-161, 1979.

HENDERSON, Vern; THISSE, Jacques-François (Ed.). **Handbook of regional and urban economics: cities and geography**. Elsevier, 2004.

HENDERSON, J. Vernon. Cities and development. **Journal of Regional Science**, v. 50, n. 1, p. 515-540, 2010.

HOTELLING, Harold. The economics of exhaustible resources. **The Journal of Political Economy**, v. 39, n. 2, p. 137-175, 1931.

IBGE – **Instituto Brasileiro de Geografia e Estatística**. Disponível em <http://www.ibge.gov.br>

IBGE **Censo Demográfico 2000**. Rio de Janeiro. 2001

IBGE **Pesquisa Nacional por Amostragem de Domicílios 2004**. Rio de Janeiro. 2005

IBGE **Censo Demográfico 2010**. Rio de Janeiro. 2011

IGLIORI, D.C., REIS, E., WEINHOLD, D. **Causal Forces of Deforestation in the Brazilian Amazon: Does Size Matter?.** Working Paper. 2009. Avaiable at SSRN: http://www.nemesis.org.br/sec-din5.php?id=0000000164&i=pt.

IGLIORI, Danilo Camargo. Deforestation, Growth and Agglomeration Effects: Evidence from Agriculture in the Brazilian Amazon. **Department of Land Economy, University of Cambridge**, 2006.

IGLIORI, D. C. (2009). **Spatial Economics of Conservation and Development: Topics on Land Use Change in the Brazilian Amazon**. 1. ed. Saarbrucken: VDM Verlag. v. 1. 160 p.

IGLIORI, Danilo; ABRAMOVAY, Ricardo; CASTELANI, Sergio. Urban evolution in Sao Paulo: employment growth and industrial location. **Regional Science Policy & Practice**, v. 4, n. 4, p. 447-477, 2012.

IMORI, Denise et al. The development of the brazilian amazon region and greenhouse gases emission: a dilemma to be faced!. **Available at SSRN 1851967**, 2011.

INPE. **Instituto Nacional de Pesquisas Espaciais**. MCTI. SSRN : < http://www.inpe.br/>

IOANNIDES, Yannis M.; ROSSI-HANSBERG, Esteban. Urban growth. **The new Palgrave dictionary of Economics, second edition, Palgrave Macmillan, forthcoming**, 2005.

IPEA. **Instituto de Pesquisa Econômica Aplicada**. SSRN : <http://www.ipeadata.gov.br>

IPEA. Código Florestal: implicações do PL 1876/99 nas áreas de reserva legal. **Comunicados do IPEA**, n° 96; 8/07/2011. Disponível em: <http://www.ipea.gov.br/portal/images/stories/PDFs/comunicado/110616_comunicadoipea96.pdf>.

JACOBS, Jane et al. The economy of cities. **The economy of cities.**, 1970.

KATZ, Eliakim; STARK, Oded. Labor migration and risk aversion in less developed countries. **Journal of labor Economics**, p. 134-149, 1986.

KELEJIAN, Harry H.; PRUCHA, Ingmar R. A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. **The Journal of Real Estate Finance and Economics**, v. 17, n. 1, p. 99-121, 1998.

KELEJIAN, Harry H.; PRUCHA, Ingmar R. A generalized moments estimator for the autoregressive parameter in a spatial model. **International economic review**, v. 40, n. 2, p. 509-533, 1999.

KELEJIAN, Harry H.; PRUCHA, Ingmar R. The relative efficiencies of various predictors in spatial econometric models containing spatial lags. **Regional Science and Urban Economics**, v. 37, n. 3, p. 363-374, 2007.

KELEJIAN, Harry H.; PRUCHA, Ingmar R. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. **Journal of Econometrics**, v. 157, n. 1, p. 53-67, 2010.

KORN, Edward L.; GRAUBARD, Barry I. Simultaneous testing of regression coefficients with complex survey data: Use of Bonferroni t statistics. **The American Statistician**, v. 44, n. 4, p. 270-276, 1990.

KRUEGER, Alan; ASHENFELTER, Orley. **Estimates of the economic return to schooling from a new sample of twins**. National Bureau of Economic Research, 1992.

KRUEGER, Alan B.; LINDAHL, Mikael. **Education for growth: why and for whom?**. National Bureau of Economic Research, 2000.

KRUGMAN, Paul. **Increasing returns and economic geography**. National Bureau of Economic Research, 1990.

KRUGMAN, Paul R. **Geography and trade**. MIT press, 1991.

KRUGMAN, Paul. Complex landscapes in economic geography. **The American Economic Review**, v. 84, n. 2, p. 412-416, 1994.

LEE, Lung-Fei. Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models. **Econometrica**, v. 72, n. 6, p. 1899-1925, 2004.

LEE, Lung-fei. Best Spatial Two-Stage Least Squares Estimators for a Spatial Autoregressive Model with Autoregressive Disturbances.**Econometric Reviews**, v. 22, n. 4, p. 307-335, 2003.

LEONTIEF, Wassily W. **The structure of American economy, 1919-1939: an empirical application of equilibrium analysis**. New York: Oxford University Press, 1951.

LESAGE, James P. An introduction to spatial econometrics. **Revue d'économie industrielle**, n. 3, p. 19-44, 2008.

LESAGE, James P.; PACE, R. Kelley. **Spatial Econometric Modeling of Origin-Destination flows**. 2005. Working paper, avaiable at http://www4.fe.uc.pt/spatial/doc/lecture7.pdf.

LESAGE, James P.; PACE, R. Kelley. SPATIAL ECONOMETRIC MODELING OF ORIGIN-DESTINATION FLOWS*. **Journal of Regional Science**, v. 48, n. 5, p. 941-967, 2008.

LUCAS JR, Robert E. On the mechanics of economic development. **Journal of monetary economics**, v. 22, n. 1, p. 3-42, 1988.

MATSUYAMA, Kiminori; TAKAHASHI, Takaaki. Self-defeating regional concentration. **The Review of Economic Studies**, v. 65, n. 2, p. 211-234, 1998.

MENEZES, Tatiane; AZZONI, C.; MOREIRA, C. Diferenças em gastos com Aluguel entre Estados, Tipos de Área e níveis de Renda Familiar no Brasil. **Gasto e consumo das famílias brasileiras contemporâneas. Brasília: IPEA**, v. 2, p. 329-346, 2007.

MILLER, Paul; MULVEY, Charles; MARTIN, Nick. What do twins studies reveal about the economic returns to education? A comparison of Australian and US findings. **The American Economic Review**, v. 85, n. 3, p. 586-599, 1995.

MILLER, Ronald E.; BLAIR, Peter D. **Input-output analysis: foundations and extensions**. Cambridge University Press, 2009.

MINCER, J. **Family Migration Decisions**. The Journal of Political Economy, v. 86, n. 5, p. 749-773, 1978

MCT - Ministério da Ciência e Tecnologia. **Second Brazilian Inventory of Emissions and Anthropogenic Removals of Greenhouse Gases**. 2010.

MORELLO, T. F. (2012). Coletânea de textos sobre a Reforma do Código Florestal Brasileiro. Revista Visão Agrícola, v. 10, ano 7, jan.-abr. 2012.

MORTON, Douglas C. et al. Cropland expansion changes deforestation dynamics in the southern Brazilian Amazon. **Proceedings of the National Academy of Sciences**, v. 103, n. 39, p. 14637-14641, 2006.

NEMESIS **Núcleo de Estudos e Modelos Espaciais Sistêmicos**. Available at www.nemesis.org.br

PARETO, Vilfredo. **Cours d'economie politique**. Librairie Droz, 1964.

PNUD - Programa das nações unidas para o desenvolvimento**. Atlas do desenvolvimento humano no Brasil. 2013.**

PRODES – INPE. **Instituto Nacional de Pesquisa Espacial**. Disponível em www.inpe.gov.br/prodes .

PROJECT CATALYST. **Climate Works Foundation**, 2011. Available at SSRN: http://www.project-catalyst.info/.

REES, William; WACKERNAGEL, Mathis. Urban ecological footprints: why cities cannot be sustainable–and why they are a key to sustainability. **The Urban Sociology Reader**, p. 157, 2012.

ROMER, Paul M. Endogenous technological change. **Journal of political Economy**, p. S71-S102, 1990.

ROMER, Paul M. **Human capital and growth: theory and evidence**. National Bureau of Economic Research, 1989.

ROMER, Paul M. The origins of endogenous growth. **The Journal of Economic Perspectives**, v. 8, n. 1, p. 3-22, 1994.

SBPC. O código florestal e a ciência: contribuições para o diálogo. Grupo de Trabalho do Código Florestal. Sociedade Brasileira para o Progresso da Ciência, **Academia Brasileira de Ciências**. 2011. Disponível em: <http://www.sbpcnet.org.br/site/codigoflorestal/index.php>.

SILVA, José Antônio Aleixo da et al. **O Código Florestal ea ciência: Contribuições para o diálogo**. 2011.

SINGER, P.. **Migrações internas: considerações teóricas sobre o seu estado**. In: MOURA, H. A. (org.) Migração interna, textos selecionados. Fortaleza, BNB/ENTENE, p. 211-244, 722p, 1980

SJAASTAD, Larry A. The costs and returns of human migration. **The journal of political economy**, v. 70, n. 5, p. 80-93, 1962.

SOLOW, Robert M. A contribution to the theory of economic growth. **The quarterly journal of economics**, v. 70, n. 1, p. 65-94, 1956.

SPENCE, Michael. Job market signaling. **The quarterly journal of Economics**, v. 87, n. 3, p. 355-374, 1973.

Stata survey data reference manual: release 11. StataCorp LP, 2009.

TAYLOR, J. Edward. **Differential migration, networks, information and risk**. Migration and Development Program, Harvard University, 1984.

TODARO, Michael P. A model of labor migration and urban unemployment in less developed countries. **The American Economic Review**, v. 59, n. 1, p. 138-148, 1969.

VEIGA, J. E. (2002). **Cidades Imaginárias: o Brasil é menos urbano do que se imagina**. Campinas, SP: Autores Associados, 2002.

DEL CARMEN VERA-DIAZ, Maria; KAUFMANN, Robert K.; NEPSTAD, Daniel C. **The Environmental Impacts of Soybean Expansion and Infrastructure Development in Brazil's Amazon Basin**. Tufts University, 2009.

VON THÜNEN, Johann Heinrich. Der isolierte Staat in Beziehung auf Nationalökonomie und Landwirtschaft. **Gustav Fischer, Stuttgart (reprinted 1966)**, 1826.

WALKER, R. T. & HOMMA, A. K. O. (1996). "*Land use and land cover dynamics in the Brazilian Amazon: An overview*". Ecological Economics, 18(1):67–80.

WALKER, R. T.; MORAN, E.; ANSELIN, L. (2000). "*Deforestation and Cattle Ranching in the Brazilian Amazon: External Capital and Household Processes*". World Development, Vol. 28, No. 4, April 2000. Available at SSRN: http://ssrn.com/abstract=253176.

WOOLDRIDGE, Jeffrey M. **Econometric analysis of cross section and panel data**. The MIT press, 2002.

APPENDIX A.1: DATA AND DEFINITIONS ISSUES REGARDING THE GEOGRAPHIC DIVISION OF REGION 1: THE METROPOLITAN REGIONS OF AMAZON.

The five Metropolitan Regions chosen to define Region 1 in the first chapter of this thesis are the ones officially defined by IBGE as belonging within the Brazilian Amazon. However, two important issues arise with regard to the choice of these five areas to compose Region 1.

First, many researchers argue that there are more urban conglomerations within the Brazilian Amazon which could be considered metropolitan regions. However, as those are not officially declared by IBGE as Metropolitan Regions, they were not included in Region 1, and therefore, they belong to Region 2 in our analysis. The reason for that was to try to avoid several controversial hypothesis regarding definitions of metropolitan areas, which we wanted to avoid. Moreover, this choice, in fact, can be considered a conservative strategy, in terms of measuring the impacts of urbanization over deforestation.

Second, due to database issues, we were not able to build Region 1 exclusively with the municipalities which compose the 5 Metropolitan Regions considered. The reason for that is because our inter-regional Input Output tables can only be built at the level of IBGE's "Micro regions", and sometimes these Micro regions corresponded not only to the municipalities within the Amazonian Metropolitan Regions, but they also contained a few other municipalities which do not belong specifically to these Metropolitan conglomerates. Table A.1.1 shows the list of municipalities from the IBGE's Micro-Regions that actually compose Region 1 in our analysis, indicating whether they belong or not to the Metropolitan Regions according to the municipalities which compose these, also defined by IBGE. Also, due to reasons we discuss ahead, Table A.1.1 brings the share of urban population held by each city which composes our Metropolitan regions.

As Table A.1.1 shows, the municipalities which do not belong to Metropolitan Regions (as defined by IBGE), but constitute Region 1 by being part of the Micro-Regions used to build the Input-Output inter-regional system are usually less urbanized, in terms of their share of urban population, than are the ones which actually constitute the Metropolitan Regions. Thus, this means that our strategy of definition of the Amazon Metropolitan Regions (Region 1) can be considered conservative, since we are underestimating the rate of urbanization of Region 1, even though this urbanization rate is still 90.39%.

Furthermore, the population living within the municipalities which actually compose the 5 Metropolitan Regions of Amazon encompasses 94.68% of the total population from Region 1 as defined in our estimations. Thus, in terms of overestimating the total population living in the Metropolitan Region, the bias which we may underlie is only about only 5%.

Still, even with these two issues, our results show robust evidence that local urbanization seems to be exerting a relevant deforestation impact in Brazilian Amazon.

Table A.1. 1 : IBGE's Micro-Regions and Metropolitan Regions of Brazilian Amazon

City	Metropolitan Region (MR)	Micro Region	Population in 2010	Urban Population in 2010	Share of Urban Population in 2010
Barcelos	-	13001	25,718	11,157	43.38%
Novo Airão	MR Manaus	13001	14,723	9,499	64.52%
Santa Isabel do Rio Negro	-	13001	18,146	6,856	37.78%
São Gabriel da Cachoeira	-	13001	37,896	19,054	50.28%
Autazes	-	13007	32,135	13,893	43.23%
Careiro	MR Manaus	13007	32,734	9,437	28.83%
Careiro da Várzea	-	13007	23,930	1,000	4.18%
Iranduba	MR Manaus	13007	40,781	28,979	71.06%
Manacapuru	MR Manaus	13007	85,141	60,174	70.68%
Manaquiri	-	13007	22,801	7,062	30.97%
Manaus	MR Manaus	13007	1,802,014	1,792,881	99.49%
Presidente Figueiredo	MR Manaus	13008	27,175	13,001	47.84%
Rio Preto da Eva	MR Manaus	13008	25,719	12,205	47.46%
Itacoatiara	MR Manaus	13009	86,839	58,157	66.97%
Itapiranga	-	13009	8,211	6,451	78.57%
Nova Olinda do Norte	-	13009	30,696	13,626	44.39%
Silves	-	13009	8,444	4,029	47.71%
Urucurituba	-	13009	17,837	10,448	58.57%
Ananindeua	MR Belém	15007	471,980	470,819	99.75%
Barcarena	-	15007	99,859	36,297	36.35%
Belém	MR Belém	15007	1,393,399	1,381,475	99.14%
Benevides	MR Belém	15007	51,651	28,912	55.98%
Marituba	MR Belém	15007	108,246	107,123	98.96%
Santa Bárbara do Pará	MR Belém	15007	17,141	5,458	31.84%
Serra do Navio	-	16003	4,380	2,575	58.79%
Pedra Branca do Amapari	-	16003	10,772	5,963	55.36%
Cutias	-	16003	4,696	2,442	52.00%
Ferreira Gomes	-	16003	5,802	4,175	71.96%
Itaubal	-	16003	4,265	1,754	41.13%
Macapá	MR Macapá	16003	398,204	381,214	95.73%
Porto Grande	-	16003	16,809	10,809	64.30%
Santana	MR Macapá	16003	101,262	99,111	97.88%
Paço do Lumiar	MR Grande São Luís	21002	105,121	78,811	74.97%
Raposa	MR Grande São Luís	21002	26,327	16,675	63.34%
São José de Ribamar	MR Grande São Luís	21002	163,045	37,709	23.13%
São Luís	MR Grande São Luís	21002	1,014,837	958,522	94.45%
Chapada dos Guimarães	-	51017	17,821	11,037	61.93%
Cuiabá	Cuiabá-Várzea Grande	51017	551,098	540,814	98.13%
Nossa Senhora do Livramento	-	51017	11,609	4,242	36.54%
Santo Antônio do Leverger	-	51017	18,463	7,160	38.78%
			6,937,727	**6,271,006**	**90.39%**

SOURCE: IBGE. Own Elaboration

APPENDIX A.2: Sectoral Aggregation Map

Table A.2. 1 : Sectoral Aggregation Map

Original Industry Number	Original Industry	Aggregated Industry Number	Aggregated Industry
1	Agriculture, forestry, extractive	1	Agriculture, forestry, extractive
2	Livestock and fisheries	2	Livestock and fisheries
3	Petroleum and Natural gas	3	Petroleum and Natural gas
4	Iron ore	4	Iron ore
5	Other quarrying industries	5	Other quarrying industries
6	Food and Beverage	6	Food and Beverage
7	Tobacco products	7	Tobacco products
8 9	Textiles Articles of apparel and accessories	8	Textiles, apparel and accessories
10	Leather products and footwear	9	Leather products and footwear
11	Wood products - exclusive furniture	10	Wood products - exclusive furniture
12	Pulp and paper products	11	Pulp and paper products
13	Newspapers, magazines, records	12	Newspapers, magazines, records
14	Petroleum refining	13	Petroleum refining
15	Alcohol	14	Alcohol
16 17	Chemicals Manufacture of resin and elastomers	15	Chemical products, resin and elastomers
18	Pharmaceutical	16	Pharmaceutical
19	Agrochemicals	17	Agrochemicals
20	Perfumery, hygiene and cleaning	18	Perfumery, hygiene and cleaning
21 22	Paints, varnishes, enamels and lacquers Various chemical products	19	Paints, varnishes, enamels, lacquers and other chemicals
23	Rubber and plastic	20	Rubber and plastic
24 25	Cement Other products of non-metallic minerals	21	Non-metallic minerals
26 27 28	Manufacture of steel and steel products Metallurgy of non-ferrous Metal products - except machinery and equipment	22	Steel and Metallurgy
29 30 31 32 33 34	Machinery and equipment, including maintenance and repairs Appliances Office machines and computer equipment Machinery, appliances and computer equipment Electronic material and communication equipment Medical and healthcare Equipment	23	Diverse Machinery and Appliances
35 36 37 38	Cars, small trucks and SUVs Trucks and buses Parts and accessories for motor vehicles Other transportation equipment	24	Vehicles and transport equipments
39	Furniture and products of diverse industries	25	Furniture and products of diverse industries
40	Electricity, gas, water, sewage and urban cleaning	26	Electricity, gas, water, sewage and urban cleaning
41	Construction	27	Construction
42	Trade	28	Trade
43	Transportation, storage and postal services	29	Transportation, storage and postal services
44 45 46 47	Information Services Financial intermediation and insurance Real estate services Maintenance and repair	30	Private Services except Food and Beverage
48	Accommodation services and meals	31	Lodging and Food services
49 50 51 52	Business services Private Education services Private Health services Other services	30	Private Services except Food and Beverage
53 54 55	Public education services Public health services Public administration and social security	32	Public Services

Source: Own Elaboration.

APPENDIX A.3: Estimating the Costs of Living per municipality

In order to estimate the costs of living for every municipality of Brazil, we followed a strategy based on recent literature about the subject. More specifically, several studies concerning the study and estimation of costs of living found evidence about the existence of a robust, positive and stable correlation between costs of living within one city (or region), and the average rent paid by households in this same municipality (or region). Azzoni et al. (2000) and Menezes et al. (2007), for example, provide a theoretical basis for this argument, and also find empirical evidence (using data for Brazilian States and Metropolitan Regions) corroborating that the average rent of one region may be used as a proxy for that region costs of living index.

Theoretically, the authors base their argument in a microeconomic perspective which reflects markets' equilibriums, supply and demand elasticities, and non-arbitrage conditions: as land is an asset which does not present any substitutes, and is spatially fixed, its supply curve elasticity tends to be very low. In turn, if the demand for land-use increases in a certain region, this immediately results in a local land price rise, and consequently, in an increase of the region's rental prices. Such increase in rental prices imply higher costs for industrial and tertiary activities within that region, which by its turn tend to end in increases of all final goods' and services' prices within that area. Therefore, it results in an increase of the whole region's costs of living. Empirically, the authors find robust evidence that, indeed, rental prices and costs of living are positively correlated: using data from Brazilian States for the year of 2003, the authors show that a simple cross-section regression between rental prices and costs of living result in a R^2 of 0.96, according to their estimations. Based on these results, we have chosen to try to estimate the average domiciliary rental for all Brazilian municipalities for a period prior to 2005, and use it as a proxy for the costs of living per municipality in that same period.

In order to do so, we have implemented a two step hedonic prices' Heckit (see Heckman, 1979) procedure mixing information from the IBGE's PNAD databases and the IBGE's 2000 demographic Census database, which we called a "Heckit Hedonic Rent Model", as described subsequently. The reason why we had to mix information for these two databases is the following: On the one hand, PNAD data contains the rental prices paid by each domicile as one of its available variables, along with other domiciles' characteristics. However, such information is only available at the Brazilian State level, since this survey does not contemplate all Brazilian municipalities in its sampling methodology. On the other hand, IBGE Census data is available for every municipality of Brazil, and contains many of the domiciles characteristics available at PNAD, however, the specific information about rental prices paid by households is not among the information set collected in the IBGE's Census surveys, even though the information regarding if the domicile is rented or not is available.

More specifically, the first step of this procedure consisted in firstly collecting microdata from the IBGE's PNAD (National Domiciliary Sample Research), which contains 2 sets of information necessary to implement it: 1) information about the monthly rental price paid by each domicile which composes its sample (at a national level); 2) information about these domiciles' characteristics, such as their number of dormitories and commodes, access to water treatment and energy, number of occupants, dummy variables indicating if the domicile is located or not within a metropolitan region, the Brazilian federal state in which the domicile is located at, etc. Given such information sets, the first step then consisted of estimating a hedonic prices regression of rental prices per domicile, that is, of regressing the value of the domiciliary rent paid monthly by each household against the vector of characteristics of his respective domicile.

The aim of this hedonic estimation is to consistently obtain the coefficients indicating how each of the hedonic characteristics of the domiciles affect their average rental price, to then proceed with the second step of the procedure. This second step consists of the following: as the domiciles characteristics selected in these hedonic regressions using PNAD microdata are also available on the IBGE's Census data (with the same specifications), we use the coefficients estimated in this first step regression to predict the rental values paid by each domicile of Census (which unfortunately, as already mentioned, are not collected in the Census survey) in every Brazilian municipality. Therefore, our estimations produce the estimated average rental values paid by households in each municipality of Brazil, according to how the characteristics of the domicile affect this rental prices, on average. Given these estimated rental prices per domicile, we finally obtain the rental prices per municipality by averaging these estimated rents per municipality. These average rental prices per municipality are the proxies for the costs of living which we use in our migration flows regressions.

Methodologically, in order to estimate this hedonic regression consistently, we once again have implemented a Heckit procedure to deal with a problem of sample-selection that arises when estimating how domiciles' characteristics may determine the rental price that households must pay. This sample-selection problem occurs because at the domicile level, rent is observed only for domiciles which actually have to pay rents, that is, for those domiciles which are not owned by the individual who lives in it. The problem arises because as we are interested in predicting the average rent per municipality (to use it as a proxy for that city's cost of living), we may say that we are interested in measuring not only how characteristics of rented domiciles may affect the value of this municipality's average rent, but also how characteristics of non-rented domiciles may affect the value of an hypothetical rent that would have to be paid in case this was a rented domicile. Or, in simple terms, the problem surges because the population of interest which we are interested to study concerns not only the domiciles which actually pay rents, but also the domiciles which do not have to pay it, since our real interest regards the relationship between rental prices and costs of living.

A drastic example helps to understand the reason why we must also consider non-rented domiciles as part of our interest population: suppose that one specific municipality does not hold any rented domicile within its borders. This does not mean that the cost of living within that city is zero, as we know that this would be impossible to occur under any capitalist economic regime. However, in case we did not include the non-rented domiciles as our population of interest, we would be wrongly implying that the costs of living in that hypothetical city would be zero. Thus, our population of interest when treating rent as a proxy for cost of living must be the all domiciles within each municipality, and as we only observe rental prices for domiciles which are actually rented, the sample selection problem arises (see Wooldridge, 2002).

Therefore, in order to correctly estimate rent as the proxy for costs of living, we must also predict the hypothetical rents that would be paid by actually non-rented domiciles, according to its characteristics. Particularly in our case, this means to input the coefficients attained on the first step hedonic regression into the characteristics of all domiciles, rented or not. Statistically, this corresponds to implement a Heckit procedure which deals with this sample selection matter.

Specifically, such Heckit procedure which have been applied consists in estimating the hedonic regression in two stages, with the first being the estimation of the probability of the domicile to be rented through a probit model, and the second being the hedonic regression itself, but with the inclusion of the normalized predicted values of the probability estimated in the first stage (the inverse Mills ratio). As Heckman (1978) argues, to obtain consistent estimates, it is necessary that the first stage contains at least one significant variable which must not be included in the second stage, so that the sample selection criterion is at least partially explained exogenously in the Heckit first stage. Bearing that in mind, we have included three groups of variables only in the first stage of this procedure, which are: the highest individual's income within each domicile (among the incomes of all its occupants), the age of the eldest member of each domicile, and the ethnic group to which the household's head belongs to. The choice of these variables were based on the argument that individuals tend to choose between buying or renting their places of living according to their level of income, since individuals that are richer are more likely to be the owners of the domicile he lives in. In this sense, the ethnic group to which the individual belongs to also may be associated to this individual income condition (according to statistic researches on socio-economic aspects of different ethnic groups), as well as this individual's age. Also, these variables were chosen because they are among the few available variables in both PNAD and Census data which are not already used in the Heckit Hedonic Rent Model second stage implemented here.

As a final regard concerning the databases used in this specific costs of living estimation procedure, it is important to point out that we used the PNAD database referring to the year of 2004, and the Census data of the year 2000. The reasons we have chosen PNAD in 2004 instead of one whose year of reference would be closer to the year 2000 are two: 1) as previously explained, our migration data regards migration flows from 2005 to 2010. Therefore, coefficients from the Heckit Hedonic Rent Model may reflect conjectural economic conditions more closely related to the initial year of migration considered in this study by the choice of using the 2004 PNAD database; 2) From the year of 2004 and forth, PNAD sampling was reformulated in order to include more municipalities of the Amazon region than it included in its databases prior to 2003. As the study of this specific region is exactly the main objective of our study, this seemed a good reason to use PNAD databases referring to 2004.

Results from the Heckit Hedonic Rent Model are displayed in table A.3.1. Estimation was made by a Maximum Likelihood estimator, in which both stages are simultaneously estimated, and coefficient variances were adjusted linearly for large survey data procedures.

Table A.3. 1: Heckit Hedonic Rent Model

Explanatory Variables	*Second Stage (dependant variable: household rent)*	*First Stage (dependant variable: probability of household rental)*
nº of household members	-0.072***	-0.115***
squared nº of household members	0.003***	0.009***
nº of dormitories	0.197***	-0.062***
squared nº of dormitories	-0.015**	0.000***
nº of commodes	0.254***	-0.125***
squared nº of commodes	-0.003***	0.001
access[1] to sanitary treatment	0.112	0.866***
access[1] to piped water	0.253***	0.267***
access[1] to eletricity	0.185	0.665***
Metropolitan region dummy[2]	0.600***	0.600***
state dummy - Acre	0.301***	-0.303***
state dummy - Amazonas	0.253*	-0.131
state dummy - Roraima	0.171**	-0.221*
state dummy - Pará	-0.216***	0.100
state dummy - Amapá	0.553***	-0.352***
state dummy - Tocantins	-0.210**	0.247**
state dummy - Maranhão	-0.299***	0.040
state dummy - Piauí	-0.668***	0.093
state dummy - Ceará	-0.815***	0.260***
state dummy - Rio Grande do Norte	-0.452***	0.230***
state dummy - Paraíba	-0.635***	0.099
state dummy - Pernambuco	-0.716***	0.195***
state dummy - Alagoas	-0.523***	0.313***
state dummy - Sergipe	-0.522***	0.022
state dummy - Bahia	-0.538***	-0.007
state dummy - Minas Gerais	-0.455***	0.076
state dummy - Espírito Santo	-0.280***	0.005
state dummy - Rio de Janeiro	0.010	-0.057
state dummy - São Paulo	-0.011	-0.005
state dummy - Paraná	-0.372***	0.042
state dummy - Santa Catarina	0.013	-0.026
state dummy - Rio Grande do Sul	-0.163***	-0.156**
state dummy - Mato Grosso do Sul	-0.162***	0.211***
state dummy - Mato Grosso	0.102	0.106
state dummy - Goiás	-0.265***	0.204***
state dummy - Distrito Federal	-0.228***	0.293***
household head ethnic group -	-	0.050
household head ethnic group - black	-	-0.054
household head ethnic group - oriental	-	0.051
household head ethnic group - mullatto	-	-0.048
eldest household member age	-	-0.022***
squared eldest household member age	-	0.000*
domiciliar nominal income	-	0.000***
constant	3.616***	-1.566***

notes: *** p<0.01, ** p<0.05, * p<0.1. 1 - the variable assumes the value 1 if access the domicile has access, and 0 otherwise, 2 - the dummy assumes the value 1 if the domicile is located within a metropolitan region.

SOURCE: Own elaboration

As we can observe in table A.3.1, regarding the first stage results, the higher the number of family households living within the domicile, the lower the probability of this house to be rented. Moreover, bigger houses tend to be owned by the individuals who actually live in them, as rental probability falls with the number of dormitories and commodes. Also, domiciles located at Metropolitan Regions are more leaned to be rented than houses located outside these areas. This is also an expected result, once metropolitan regions generally tend to present a higher scarcity of domiciles' availability of their real estate market, which causes real estate prices to be higher within metropolitan areas. Moreover, we find that domiciles were more likely to be owned by younger household members living within them, as the probability of rental decreases with the age of the elder household living within them.

Finally, higher levels of domiciliary income tend to increase rental probability. This result was unexpected and it is difficult to interpret, as individuals with higher income are in general more capable of purchasing their own houses. However, it is possible that this result might be reflecting preferences of richer individuals in respect to the real estate markets as an investment decision, when compared to other actives available in this sense: it is possible that these richer individuals prefer to invest their savings in other assets rather than buying a property, as it would be in the case that the real estate market perspectives are low. Another possibility would be to interpret this result from a spatial economics modeling perspective: individuals with higher income usually live next to their workplaces, in order to reduce transport costs, since they are more able to choose their living location (due to their higher level of income). As these individuals earn more than the average worker, we may infer that labor productivity is also higher in these areas. As a von Thünen inspired model would predict, economic activities that present higher productivity tend to be located in central areas, in order to benefit from closeness to central markets. Thus, real estate prices nearby these areas are probably higher than other areas, as land tend to be more valuable next to where central markets are located. These higher real estate prices, in turn, discourage them to purchase domiciles within these areas, and encourage them to rent those specific households.

Second stage results show that, as expected, rental prices tend to be higher the bigger the size of the domicile, measured by the number of dormitories and commodes it pursues. Furthermore, as expected, better access to water treatment within the region where the domicile is located at increases its rental price. Moreover, houses placed within metropolitan areas are also more expensive in terms of rental, as expected by spatial economic models which emphasize that areas which concentrate economic activities tend to present higher land values. Another result is that rental prices drop as the number of individuals living within the domicile increases. Possibly, this may be reflecting real estate pricing structure in peripheral suburban areas, where large (and sometimes more than one) families with very low income live, in a typical situation caused by high inequality, which is also a typical characteristic of Brazilian cities.

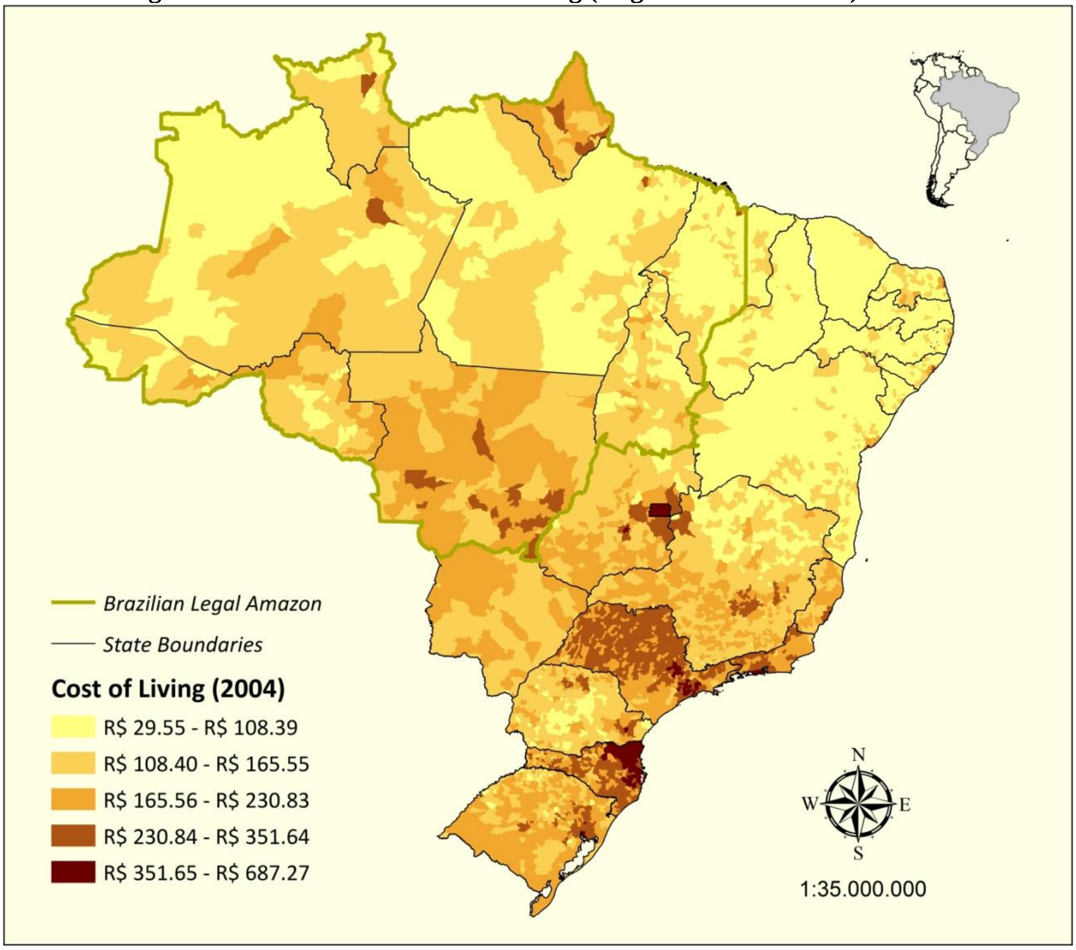
Given these coefficients' estimations, before having inputted them on census database for all Brazilian municipalities, as previously discussed, we have implemented a few empirical tests within the sample. These tests were made in order to evaluate if, in fact, the average estimated rents that this regression predicts exhibit or not a few properties that rental prices and cost of living tend to have, in general. As a first test, we have observed if predicted rents within the sample have resulted in non-negative estimated rents for all observations, since rental prices and costs of living are always positive. Results from this test shown that the average predicted rent within the sample was R$205.97, with a minimum of R$19.16 and a maximum of R$3,316.45. These values might be considered reasonable in terms of Brazilian municipalities.

Secondly, we have tried to measure if the estimated probabilities of domiciles to be rented (in the first stage) fitted well the real share of domiciles that were actually rented within the sample. In order to do so, we have compared the average predicted (from the Heckit first stage estimations) probability of any domicile within the sample to be considered as rented with the average estimated probability of one domicile to be rented, given that this domicile was actually rented. Results of this comparison showed an average probability of about 15% of any domicile to be considered rented regarding the full sample; and an average probability of 23% of one domicile to be considered rented by our model predictions, given that this domicile is actually rented. Thus, our estimation attributes higher probability of rental to domiciles that are actually rented, which may be considered as evidence that our predictions point to the right direction in this sense.

Thirdly, from a macroeconomic perspective, it is expected that costs of living may be positively correlated with nominal income, so that real income among different regions relatively converge in the long run. Thus, we have calculated the correlation between the predicted costs of living (rents) with the nominal domiciliary income within the sample. As results, we have found a positive covariance of 4.5, significant at 1% level. Moreover, after we have imputed these coefficients on variables from Census data and obtained the final predicted cost of living per domicile of Brazil, we have also calculated this covariance, and once again, found a positive and significant (at the 1% level) result.

Finally, once the coefficients were estimated as described above, we have imputed them on Census data, obtaining the predicted value of the rent that each domicile would pay according to their characteristics (estimated through PNAD data). Then, we have calculated the average value of this rent for each municipality, weighting this average by the total number of households which were actually rented in the year of 2000 (according to Census data). The final result is our proxy for the costs of living per municipality, which we sum up in Figure A.3.1.

Figure A.3. 1: Estimated Costs of Living (Avg. Household Rents) in 2004



SOURCE: Own Elaboration

As Figure A.3.1 shows, cities within the southern part of Brazil present the highest costs of living proxies of Brazil. This result was expected and is in accordance to the macroeconomic convergence perspective, in which regions which exhibit highest levels of nominal income are also the ones that might present the highest costs of living.

APPENDIX A.4: "Hierarchical Heckit" First Stage Table with the full set of explanatory variables

Table A.4. 1: "Hierarchical Heckit" First Stage: individual probability of migration in Brazil, explained by push factors and individual characteristics (showing with all explanatory variables)

"Hierarchical Heckit" First Stage: Individual Porbablility of Migrating

Explanatory Variable	*Probit Coefficient*	*t-stat*
Age in 2005	-0.044***	(-184.055)
Squared Age in 2005	0.001***	(101.234)
Gender	0.069***	(44.050)
Born in this city	-1.239***	(-627.555)
Etnic Group - Black (omitted: Caucasian)	-0.053***	(-19.429)
Etnic Group - Oriental (omitted: Caucasian)	-0.044***	(-6.606)
Etnic Group - Mullatto (omitted: Caucasian)	-0.061***	(-38.683)
Etnic Group - Indians (omitted: Caucasian)	-0.053***	(-4.413)
Population in the Origin City (2000)	-0.001***	(-45.669)
Labro Income variation bt	-0.020***	(-5.812)
Average Labor Income (Origin) (log)	-0.022***	(-4.372)
Avg. Cost of Living (Origin) (log)	0.070***	(13.840)
Cost of Living (Origin) (log) - Standard Deviation	-0.001***	(-8.253)
Gini Index (Origin)	1.150***	(59.046)
EAP / WAP ratio (Origin)	-0.018	(-1.069)
% of "Formal" Working force	-0.003	(-0.281)
% of Labor Force in Services and Commerce (Origin)	-0.490***	(-49.807)
IFDM Education Index (2005)	-0.302***	(-31.027)
% of Iliteracy (Origin)	-0.238***	(-13.223)
Number of Academics (2005) (Origin)	0.001***	(44.618)
% of Pop. w/ College Degree (Origin)	0.932***	(25.386)
% of Urban Population (origin)	0.099***	(12.201)
Pop. Density (Origin)	0.001***	(7.235)
Squared Pop. Density (Origin)	0.001***	(5.611)
% of Pop. aged 22 - 29 (Origin)	-1.321***	(-14.490)
% of Pop. aged 55 - 60 - "about to retire" (Origin)	5.232***	(15.463)
% of Pop. aged more than 60 - "retired" (Origin)	1.471***	(12.193)
Previous (1991 - 2000) Migration Inflow (Origin)	-1.667***	(-48.068)
Squared Previous (1991 - 2000) Migration Inflow (Origin)	2.653***	(37.094)
IFDM Health Index (2005)	-0.053***	(-4.374)
% of Pop. w/ Access to Sanitary Treatment (Origin)	-0.133***	(-32.915)
Constant term	-0.427*	(-1.803)
Agriculture, livestock, production forestry, fisheries and aquaculture (ommitted - "other industries")	0.393*	(1.679)
Extractive industries (ommitted - "other industries")	0.645***	(2.752)
General industry (ommitted - "other industries")	0.437*	(1.866)
Electricity and gas (ommitted - "other industries")	0.593**	(2.530)
Water, sewage, waste management activities and decontamination (ommitted - "other industries")	0.348	(1.484)
Construction (ommitted - "other industries")	0.502**	(2.146)
Trade, repair of automotive vehicles and motorcycles (ommitted - "other industries")	0.384	(1.640)
Transport, storage and mail (ommitted - "other industries")	0.374	(1.599)
Accommodation and food services (ommitted - "other industries")	0.458*	(1.955)
Communication (ommitted - "other industries")	0.494**	(2.110)
Financial activities, insurance and related services (ommitted - "other industries")	0.528**	(2.253)
Scientific and technical activities (ommitted - "other industries")	0.448*	(1.913)
Administrative services and related activities (ommitted - "other industries")	0.330	(1.411)
Public administration, national defence and social security (ommitted - "other industries")	0.463**	(1.977)
Education (ommitted - "other industries")	0.368	(1.574)
Human health and social services (ommitted - "other industries")	0.421*	(1.797)
Arts, culture, sport and recreation (ommitted - "other industries")	0.456*	(1.945)
Other service activities (ommitted - "other industries")	0.469**	(2.005)
Domestic services (ommitted - "other industries")	0.394*	(1.682)
Extraterritorial organizations and institutions (ommitted - "other industries")	0.333	(1.423)
Uninformed (ommitted - "other industries")	0.504**	(2.155)
Number of obs = 12,049,139 Number of Strata = 10,184	Population size = 113,222,203	

Note 1: *** p<0.01, ** p<0.05, * p<0.1

Note 2: All variables refer to the year of 2000, except when described in the variable's name.

Source: Own Elaboration

APPENDIX A.5. Spatial Econometric estimation of migration flows determinants (Hierarchical Heckits' second stage): regressions with only immigrants/emigrants of Amazon in the sample

In order to include a spatial feature in our estimations of the determinants of migration, since literature on the subject (reviewed previously) considers such feature as one possible important determinant of migration flows, we have also estimated spatial regressions with Amazon immigration and emigration flows as dependent variables in the second stage of our Hierarchical Heckit models. However, due to some computational and data problems which have emerged when implementing these spatial regressions, and since results proved to be very similar to the ones obtained by the non-spatial approach, we have chosen not to include such method and results in the main body of chapter 4. That being put, in this appendix we present the spatial econometrics method which we have implemented, and discuss these computational and data problems, also discussing the hypotheses and solutions which we have adopted to outline them. Also, at the end, we present the resulting regressions, and as these are very similar to the ones presented in chapter 4, we do not discuss them in detail here.

The first attempt to include a spatial dimension in our regressors was by introducing the distance between origins and destinations on our migration flows regressions. This was made by first calculating the great-circle distance between each origin and destination[73], and including it on equations 4.8 and 4.9, as the *D* variable. This has proved to be a good choice, as we will see later on, because the coefficients associated such distance variable was found significant and negative in all estimations, as it is to be expected, as we further discuss in the next section.

However, spatial econometric literature points that not only distance matters as a determinant of migration, but push-pull factors from origins' and destinations' neighbors might also play an important role, as previously discussed. Bearing that in mind, we followed the recent methodology developed by LeSage and Pace (2005 and 2008) in order to try to include these neighbor effects in our regressions, with a few restrictions that we discuss below.

[73] Such distance consider the coordinates of the economic centers of these municipalities. Specifically, we have used great-circle distances in such calculations, which follows the formula $D = r(\arccos(\sin\phi_1 \sin\phi_2 + \cos\phi_1 \cos\phi_2 \cos(\lambda_1 - \lambda_2)))$, where D is the distance between municipalities 1 and 2, with *r* being the earth's radius, (ϕ_1, ϕ_2) are the coordinates of municipality 1, in radians, (λ_1, λ_2)are the coordinates of municipality 2.

The starting point in order to implement this methodology is to try to estimate a spatial version of equation 4.7, including neighbors variables of both destinations and origins. In order to do so, we firstly have to build two spatial weights neighbors' matrices, which are: W_o for the origins' neighbors, and W_d for the destinations' neighbors. The interpretation of these matrices is slightly different from traditional weight matrices in spatial econometrics literature. More specifically, it is as follows: if we wish to calculate which are the average neighbors' migration flows towards each of the destinations, we must multiply W_d by the vector of flows, y , thus obtaining $W_d y$, which is called the spatial lag of the destinations. Supposing that there are three fictional municipalities, A, B and C, and that they are all neighbors between themselves, then the element in $W_d y$ which corresponds to the flow from A to B will represent an average of all flows towards A and C, as A and C are the neighbors of destination B. Similarly, considering the neighbors' weights matrix of the origins, the element in $W_o y$ which corresponds to the flow from A to B will represent an average of all flows originated in B and C, since B and C are the neighbors of A.[74]

Given these matrices, Lesage and Pace show that if we wish to include the spatial lags of both destinations and origins as explanatory variables, that is, if we wish to calculate a an origin-destination (OD) SAR[75] (spatial autoregressive model) (see Anselin, 1988 and Elhorst, 2005), then the transformation in equation 4.7 would result in the general form given by equation A.5.1:

[74] LeSage and Pace (2005 and 2008) argue that it is also possible to consider a third type of neighbors' weights matrix, which would be an interaction between the neighbor effects of the origins and destinations. This Matrix would be $W_m = W_d.W_o$, and as previously, the element in $W_m y$ which corresponds to the flow from A to B will represent an average of all flows coming from B and C along with the flows towards A and C, as B and C are the neighbors of origin A, and A and C are the neighbors of destination B. However, in this study, we have not includes such matrix, due to the high level of complications it would bring to the methodology: in terms of estimation, including it in regressions would lead to the necessity of using a modified version of traditional spatial cross-section estimators, which are not yet implemented in most statistical softwares; and mainly due to computational issues, because as we will see in subsequently, the W_o and W_d matrices which we have used will exhibit different sizes, which in turn precludes the multiplication between them. Not including W_m in the regressions means that we are assuming that interaction effects which mix destination and origin neighbor effects do not play an important role in our estimations. In simple terms, this means that we are assuming that all neighbor effects from origins and destinations are being captured separately. Given that we will even be forced to separate the origin neighbor effects from the destination neighbor effects due to reasons discuss subsequently, then omitting W_m do not seem to impose any additional restriction that were not already imposed by other reasons.

[75] For practical reasons of space, here we show only the general form of the SAR model, as in LeSage and Pace (2008). However, SEM or SAC models could also be implemented similarly, as the authors show.

$$y = \alpha + \Omega_o W_o y + \Omega_d W_d y + \beta_o X_o + \beta_d X_d + \gamma D + \rho_o \lambda_o + \rho_d \lambda_d + \varepsilon \quad \text{(A.5.1)}$$

where Ω_d is the coefficient of the destinations' spatial lag, which measures how much migration towards the neighbors of one destination municipality influence the migration flows towards this municipality; and Ω_o is the coefficient of the spatial lag of origins, which measures how much migration flows departing from one origin's neighbors influence the migration flows departing from this origin municipality.

If we were to follow the same kind of procedure implemented in chapter 4, the next natural step in this spatial estimation procedure would be to include the Brazilian Amazon interaction dummies in equation A.5.1, similarly to what have resulted in equations 4.8 and 4.9. However, this step was impossible to implement, due to computational difficulties that emerged because of the size of these spatial matrices, which we further explain here.

More specifically, the size of the weights matrix depends on the number of flows included in the regressions. If, for example, there were *n* municipalities in our sample, and every municipalities share a flow of migration with every other municipality, then there are going to be *n(n-1)* flows. Considering that the diagonal of the matrix will be zero, as it represents the neighbor relation between one flow with himself, and there are *n(n-1)* flows plus the zero diagonal to be represented by each weights matrix, then the size of this matrix would be (n^2 *x* n^2).

In practical terms, this size represents a difficulty in calculating these matrices for Brazil as a whole. This happens because in our sample, for Brazil, we had 253,038 positive migration flows. Thus, even without considering zero flows (due to the reasons given previously), these origin and destination weights matrices for Brazil as a whole would be of size 253,038 x 253,038, which means that each matrix would have more than 64 billion observations. Unfortunately, the calculation of matrices of this size proved to be extremely difficult and slow computationally.

Due to that, we had to impose restrictions on our spatial econometric strategy. The first set of restrictions were to estimate spatial regressions only with the migration flows involving the Amazon region in the sample, in order to reduce the required size of matrices. Specifically, between 2005 and 2010, immigration flows between origins municipalities located outside Brazilian Amazon and destinations municipalities located inside this region occurred between 12,610 different pairs of municipalities. At the same period, there were 11,703 flows of emigration departing from Brazilian Amazon cities towards the Rest of Brazil's municipalities. Thus, considering only immigration flows from the Rest of Brazil towards Amazon as the observations of *y* in our sample allows us to work with a W_d matrix of size 12,610 x 12,610. In a similar way, considering only emigration flows from Amazon towards other parts of Brazil as observations allows us to work with a W_o matrix of size 11,703 x 11,703. As these are much smaller matrices, we were able to compute them, and thus, these were the sizes of matrices we used in our regressions.

Therefore, further detailing, we were able to compute 2 different matrices for Brazilian Amazon flows: W_d as a contiguous 12,610 x 12,610 matrix, and W_o as a 11,703 x 11,703 contiguous matrix. Two explanations must be made to which regards the "contiguous" nature of these matrices. First, we choose to implement a contiguous matrix guided by LeSage and Pace (2007 and 2010), as this is the kind of matrix to which they develop robust estimators. Using a different specification of a neighbor matrix would require better enlightening of the econometric robustness in the following estimations which uses these matrices, which is not the goal of this study. Given this reason, we choose to work only with the type of spatial matrix already considered by these authors.

The second necessary explanation regards the definitions and interpretations of such contiguity, as these are slightly different for W_d or W_o. In the contiguous neighbor matrix W_d, for each of its w_d^{ij} elements, with *i* and *j* representing any pair of immigration flows whose origin municipality is located at the rest of Brazil and whose destination municipality is located inside Brazilian Amazon, $w_d^{ij} = 1$ if the flow *j* has a destination that shares a common border with the destination municipality of flow *i* (independently of the origins of both *j* or *i*, as this is a "destination" Matrix), and $w_d^{ij} = 0$ otherwise. By its turn, in the contiguous matrix W_o, we have that for each of the elements w_o^{ij}, with *i* and *j* representing any of the emigration flows whose origin municipality is located inside Brazilian Amazon and whose destination municipality is located at the rest of Brazil, $w_o^{ij} = 1$ if the flow *j* has an origin that is shares a border with the origin municipality of the flow *i* (independently of the destinations of both *j* or *i*, as this is a "origin" Matrix), and $w_o^{ij} = 0$ otherwise. Thus, in simple terms, W_d encompasses only neighborhood effects of destination cities within Brazilian

Amazon, whereas W_o encompasses only the neighbor effects of the origin municipalities within Brazilian Amazon.

At this point, it becomes clear that we were forced to divide migration flows regarding Brazilian Amazon into two separate groups, with one composed only by immigration flows towards Brazilian Amazon as the dependent variable, and another composed by only emigration flows departing from Brazilian Amazon as the dependent variable. However, as these flows have completely different origins and destinations for every observation, as what distinguishes these two sets is exactly the location of these origins and destinations, with no intersection, then it becomes impossible to include both $W_d y$ and $W_o y$ in a single equation. It is easy to comprehend why: since the W_d matrix we work with concerns only destinations within Amazon and origins outside this region, and W_o is defined as the exact opposite, then it is impossible that W_o could capture any origin neighbor effects of flows whose origin municipality is located outside Brazilian Amazon, which is exactly the case of all observations composed by $W_d y$, once W_d concerns only flows whose origin cities are located outside Brazilian Amazon.

Thus, instead of estimating a single spatial equation as in A.5.1, the way we had to build W_d and W_o for different groups of migration flows have also forced us to estimate two separate equations (one using W_d, and another using W_o[76]). As a result, the equations that we have estimated in this spatial econometric approach are as in equations A.5.2 and A.5.3.

$$y^I = \alpha^I + \Omega_d W_d y^I + \beta_o^I X_o^I + \beta_d^I X_d^I + \gamma^I D^I + \rho_o^I \lambda_o^I + \rho_d^I \lambda_d^I + \varepsilon^I \quad \text{(A.5.2)}$$

$$y^E = \alpha^E + \Omega_o W_o y^E + \beta_o^E X_o^E + \beta_d^E X_d^E + \gamma^E D^E + \rho_o^E \lambda_o^E + \rho_d^E \lambda_d^E + \mu^E \quad \text{(A.5.3)}$$

where all variables are as defined previously, but with the index *I* referring to the sample composed only by the immigration flows from the Rest of Brazil to Brazilian Amazon, and

[76] In fact, mathematically, as W_d and W_o have different sizes, it would be impossible to include them both in a single spatial regression, as the number of rows of the matrix must be exactly equal to the number of observations in the sample. As a matter of fact, this is just another consequence of the fact that each of these matrices are built for completely different sets of observations.

the index E referring to the sample composed only by the emigration flows from Brazilian Amazon to the Rest of Brazil, and ε^{I} and μ^{E} are the random error terms, which are assumed to be independent of the explanatory variables on each equation.

One advantage of estimating two separate equations for immigration and emigration flows related to Amazon is that we are able to compare these two equations to each respective immigration and emigration equations 4.8 and 4.9. However, it is important to note that the interpretation of the coefficients in equations A.5.2 and A.5.3 is essentially different from the interpretation of the coefficients in 4.8 and 4.9. This is so because in equations 4.8 and 4.9, we are able to interpret the regression coefficients as the influence of each push or pull factor on the migrant’s decision of moving to (from) Brazilian Amazon instead of migrating to (from) a municipality located in the Rest of Brazil, since the sample is composed by migration flows all over the country. As equations A.5.2 and A.5.3 are estimated with the sample being composed only by immigration flows towards Amazon and emigration flows departing from Amazon, respectively, interpretation of its coefficients are as following: for equation A.5.2, each coefficient indicate the influence that its respective variable exerts on the migrant’s decision, once this immigrant has already decided to immigrate to one of the Brazilian Amazon’s municipalities. For equation A.5.3, each coefficient indicates the influence that its respective variable exerts on the migrant’s decision of migration, once this emigrant has already decided to move away from the Brazilian Amazon municipality which he lived in.

Further explaining this point, these interpretations are due to the fact that all variance of the endogenous and exogenous variables in equation A.5.2 occur among destination municipalities within Brazilian Amazon and origin cities within the Rest of Brazil. Thus, migration decision of moving, for example from one city in the Rest of Brazil to another city in this region are not being considered in the sample, in opposition to equations 4.8 and 4.9, in which all migration flows are considered and compared in groups. Thus, coefficients in equation A.5.2 represent how the Amazon destination cities’ characteristics influence the migrants’ decision on which city they might move to, but as only migration flows towards Amazon are considered, this concerns only the migrants whose decision to move towards Amazon as a whole was already taken. Therefore, summing up, interpretation of equation A.5.2 coefficients may be posed as a the following question: once the migrant has already decided to move towards the Amazon region (as a whole), how the characteristics of the available destination municipalities within this region will influence on his decision concerning to which specific city he will move to, and how characteristics at his origin (located in the Rest of Brazil) might have influenced him to leave his city within the Rest of Brazil? A similar interpretation for coefficients of equation A.5.3 may be given: once one migrant has already decided to leave Amazon, which destination cities' characteristics mostly influence him to decide to which specific city within the Rest of Brazil he will move to, and how much characteristics at his origin city within Brazilian Amazon may have influenced him to move away from this region?

Even though results from the estimation of these equations may not be considered directly comparable to the ones from equations 4.8 and 4.9, due to these restrictions we had to impose, we still chose to implement such spatial regressions, specially due to the given importance pointed out by literature (as seen in the previous sections) of including a spatial analysis when explaining migration flows. Moreover, in spite of posing the disadvantages already discussed here, these restrictions also bring a few advantages to our results, in general.

One first advantage is that as interpretation of equations 4.8 and 4.9 and the spatial equations A.5.2 and A.5.3 are different, they might be seen as complementary, as the following: results from estimation of equations 4.8 and 4.9 give us evidence about migrants behavior whether to move to (from) Brazilian Amazon o to (from) the Rest of Brazil, while results from our spatial econometrics regressions give us evidence on how the immigrants (emigrants) of Brazilian Amazon behave on average in response to characteristics among the several cities located within Brazilian Amazon (the Rest of Brazil).

A second advantage, of econometric order, is that LeSage and Pace (2005) show that when only one Weights matrix is inserted in each equation to be estimated, as in the cases of A.5.2 and A.5.3, then the more "traditional" spatial econometrics estimators can be applied, resulting in consistent estimators. Among these traditional estimators, we chose to implement the ones developed by Kelejian and Prucha (2007 and 2010), which consists of a Generalized Spatial Two Stages Least Squares (GS2SLS) in which innovations ε or μ are assumed to be iid (identically and independently distributed) and heteroskedastically distributed, with heteroskedasticity of any unknown form. Three were the reasons for the choice of these estimators: firstly, they do not rely on any distributional assumption of the error terms, which means that they may produce consistent estimations even if these errors are not normally distributed[77]. Secondly, these estimators are consistent even in the presence of heteroskedasticity[78]. Moreover, these estimators consistently estimate models which include the spatial lag Wy.

Specifically, we have estimated 4 different specifications of regressions for each of the equations A.5.2 and A.5.3. The first kind of regressions were estimated by the traditional non-spatial Ordinary Least Squares (OLS) estimator, and does not include any kind of spatial lags, which in terms of equations A.5.2 and A.5.3 is equivalent to assume that $\Omega_d = 0$ and $\Omega_d = 0$. The second kind of specification was the SAC (Spatial Autocorrelated Model) models, which include the spatial lags of the dependent variable as an explanatory variable (as in A.5.2 and A.5.3), and also a spatial dependence in the error term, in the form of equations A.5.4 and A.5.5.

$$\varepsilon = \kappa W_d \varepsilon + e \qquad \text{(A.5.4)}$$

$$\mu = \varphi W_o \mu + u \qquad \text{(A.5.5)}$$

Where u and e are the vectors of innovations (disturbances), κ is the coefficient term representing the spatial autocorrelation between the error terms of the Brazilian Amazon immigration equation A.5.2, and φ is the coefficient term representing the spatial autocorrelation between the error terms of the Brazilian Amazon emigration equation A.5.3. These two coefficients account for spatial neighbor effects that may not be observable between two neighbors' migration flows. For example, φ may be positive and significant if a non-included variable in one specific destination of a migration flow may also increase other migration flows towards the neighbors of this destination. As this variable was not included, this neighborhood effect is captured by the error term, and a spatial neighbor correlation may appear.

The third kind of specification which we have implemented in our spatial econometrics approach are the SEM models (spatial error autocorrelation model), which is a restricted version of the SAC model, with the restriction of assuming that all spatial correlation may be

[77] This is a specific advantage concerning the use of these GS2SLS estimators instead of the ones that use Maximum likelihood methods, as in Anselin (1988).

[78] This is another specific advantage of the use of GS2SLS estimators over the Maximum likelihood ones, because as Lee (2004) points out, heteroskedastic innovations may cause bias in ML estimations of spatial regressions.

captured by the error terms, that is, all spatial dependence is unobservable. Thus, it also assumes that equations A.5.4 and A.5.5 must be estimated together with A.5.2 and A.5.3, respectively, but with the prior that $\Omega_d = 0$ and $\Omega_d = 0$. Finally, the fourth kind of specification which we have implemented is the SAR (Spatial Autoregressive model) model, which is also a restricted version of the SAC model, and whose functional form is exactly as in equations A.5.2 and A.5.3, without the assumption of spatial dependence in the error terms.

The reasons why we have estimated these 4 sets of specifications are: by estimating and comparing the attained coefficients of these 4 sets altogether, we avoid making any assumptions on the form that spatial dependence assumes in our econometric modeling of migration flows, as in each of these models spatial correlation is included either in the error term (SEM), either as the spatial lag (SAR) of the dependent variable, or even in neither of them (OLS), or in both of them at the same time (SAC). Therefore, we are somehow capturing any forms of spatial dependence[79]. Secondly, as LeSage and Pace (2009) and Anselin (2003) point out, the interpretation of the coefficients' signs of estimations which include the spatial lag of the dependent variable among the regressors (SAR and SAC models) may not be straightforward if Ω_j differs from zero, *j* = *d*,*o*. This is so because an exogenous shock in one observation *i* of an explanatory variable will affect the dependent variable *y* directly (through the same observation *i*), which will be captured by its respective coefficient β. But it also affect *y* indirectly, as this initial change in *y* spreads through its *j* neighbors because of the spatial lag (through via Ω_j), which by its turn increases the values of the *y* observations that are neighbors of *i*, which in turn causes another change in observation *i* of y due to the change in its spatial neighbors. Therefore, if the spatial lags coefficients from SAC and SAR models are different from zero (significantly), then the effects of each explanatory variable in SAC and SAR models should not be interpreted directly, and it would become necessary to compute marginal effects (see Anselin, 2003, Abreu et al. 2004) in order to do such interpretation correctly. This need for marginal effects computation is specially the case when $\Omega_j < 0$, because in this case, direct and indirect effects go in opposite direction, with the indirect effects possibly annulling the direct one, or at least diminishing it.

---

[79] In fact, we also estimated regressions sets which included spatial lags of the explanatory variables. However, these spatial lags were not significant, and also did not change the signal or the significance of the other explanatory variables. Thus, we do not include these results or specifications in our estimations or methodology, as they do not interfere the evidence found by the implemented models.

If $\Omega_j > 0$, however, another possible way of interpreting the signs of the β coefficients in SAR and SAC models is possible, due to the fact that at least the sign of the all direct and indirect effect will be the same. This way is to simply estimate OLS and SEM models, and compare the attained coefficients of the explanatory variables with those from SAC and SAR estimations. If no significant difference is observed, them at least the signs of the SAC and SAR β coefficients may be considered consistent. In this case, as the indirect effect goes in the same direction of the direct effects, if we are interested in only analyzing the sign and significance of each explanatory variable's coefficient, then it becomes unnecessary to compute marginal effects, since these coefficients' signs may be interpreted directly (in terms of signs and significance, not in terms of their size) in SEM and OLS models. And as we will see in the results section, this is exactly the case in our estimations: we find that $\Omega_j > 0$, and we are mostly interested in analysing the signs and signifficance of each explanatory variable included in the model, not necessarily worring about estimating the precise size of their respective coefficients. Therefore, estimating all SAC, SAR, SEM and OLS models is justified in this study, and we are able to interpret their coefficients directly in our estimations.

Finally, results of the spatial regressions representing the estimations of equations A.5.2 and A.5.3 are, respectively, presented in tables A.5.1 and A.5.2. Once again, as these results are quite similar to the ones given by the estimation of equations 4.8 and 4.9, which are already discussed in chapter 4, we do not discuss the evidence brought by these tables here.

Table A.5. 1: "Hierarchical Heckit" Second Stage (spatial regression): sample composed only by Brazilian Amazon Immigration[80]

Explanatory Variable	*Heckit - OLS*	*Heckit - SAC - GS2SLS (Contiguous)*	*Heckit - SEM - GS2SLS (Contiguous)*	*Heckit - SAR - GS2SLS (Contiguous)*	*Variable Groiup*
Avg. Labor Income (Destination) (log)	3.776**	3.520**	3.858**	4.350***	Income and Employment
	(2.30)	(2.05)	(2.36)	(2.63)	
Avg. Labor Income (Origin) (log)	-6.551***	-5.251***	-6.478***	-5.290***	
	(-3.44)	(-2.90)	(-3.38)	(-2.81)	
Avg. Cost of Living (Destination) (log)	-8.397***	-8.517***	-8.436***	-8.948***	
	(-3.42)	(-3.46)	(-3.43)	(-3.57)	
Avg. Cost of Living (Origin) (log)	5.789***	4.870***	5.708***	4.778***	
	(3.05)	(2.79)	(3.00)	(2.64)	
***per capita* GDP (Destination) (log)**	-0.447	-0.613	-0.444	-0.516	
	(-0.56)	(-0.74)	(-0.56)	(-0.63)	
***per capita* GDP (Origin) (log)**	-1.153	-0.932	-1.135	-0.910	
	(-1.29)	(-1.02)	(-1.26)	(-1.01)	
EAP / WAP ratio (Destination)	11.415***	14.054***	11.674***	15.179***	
	(3.64)	(4.03)	(3.73)	(4.61)	
EAP / WAP ratio (Origin)	-7.279	-7.855	-7.242	-7.437	
	(-1.39)	(-1.53)	(-1.38)	(-1.43)	
Previous (1991 - 2000) Labor Income Growth (Destination	-0.563	-0.498	-0.545	-0.338	
	(-1.04)	(-0.93)	(-1.01)	(-0.62)	
Previous (1991 - 2000) Labor Income Growth (Origin)	3.580***	3.249***	3.545***	3.164***	
	(3.53)	(3.35)	(3.48)	(3.21)	
Unemployment rate (Destination)	-13.443**	-13.665**	-13.673***	-15.209***	
	(-2.54)	(-2.33)	(-2.58)	(-2.86)	
Unemployment rate (Origin)	-39.944***	-32.876***	-39.560***	-33.638***	
	(-4.59)	(-3.94)	(-4.53)	(-4.07)	
Gini Index (Origin)	2.214	1.214	2.218	1.626	Unequality
	(0.26)	(0.15)	(0.26)	(0.19)	
Gini Index (Destination)	-7.078	-5.545	-7.502	-10.430	
	(-1.11)	(-0.82)	(-1.18)	(-1.61)	
% of Pop. w/ Access to Sanitary Treatment (Origin)	2.764*	2.402	2.721*	2.248	
	(1.75)	(1.53)	(1.71)	(1.42)	
% of Pop. w/ Access to Sanitary Treatment (Destination)	10.839***	10.752***	10.838***	10.868***	
	(4.12)	(4.38)	(4.13)	(4.11)	
% of Labor Force in Services and Commerce (Origin)	6.408**	3.886	6.299**	4.160	Demographic & Urbanization
	(2.12)	(1.30)	(2.08)	(1.36)	
% of Labor Force in Services and Commerce (Destination)	20.634***	21.394***	20.776***	22.571***	
	(6.27)	(6.97)	(6.32)	(6.95)	
Pop. Density (Origin)	-0.000	-0.000	-0.000	-0.000	
	(-0.86)	(-0.63)	(-0.84)	(-0.56)	
Pop. Density (Destination)	0.014*	0.012	0.014*	0.011	
	(1.84)	(1.55)	(1.81)	(1.41)	
% of Pop. aged 22 - 29 (Origin)	270.991***	231.636***	268.904***	235.799***	
	(5.34)	(4.86)	(5.28)	(4.87)	
% of Pop. aged 55 - 60 - "about to retire" (Origin)	-165.749**	-105.856	-165.670**	-130.425*	
	(-2.17)	(-1.42)	(-2.17)	(-1.72)	
% of Pop. aged more than 60 - "retired" (Origin)	-48.176	-39.077	-47.718	-39.416	
	(-1.28)	(-1.06)	(-1.26)	(-1.05)	

SOURCE: Own Elaboration

[80] In this table we exhibit only the coefficients which proved to be signifficant, at least at the 10% level, in at least one among the 5 implemented estimators.

Table A.5.1 (cont.): "Hierarchical Heckit" Second Stage (spatial regression): sample composed only by Brazilian

Explanatory Variable	*Heckit - OLS*	*Heckit - SAC - GS2SLS (Contiguous)*	*Heckit - SEM - GS2SLS (Contiguous)*	*Heckit - SAR - GS2SLS (Contiguous)*	*Variable Groiup*
% of Pop. aged 22 - 29 (Destination)	16.151	26.128	17.737	34.673	Demographic & Urbanization
	(0.43)	(0.75)	(0.48)	(0.91)	
% of Pop. aged 55 - 60 - "about to retire" (Destination)	-316.327***	-325.170***	-318.369***	-338.970***	
	(-4.16)	(-4.26)	(-4.19)	(-4.34)	
% of Pop. aged more than 60 - "retired" (Destination)	-120.400***	-96.736***	-119.088***	-97.545***	
	(-3.50)	(-2.82)	(-3.47)	(-2.71)	
Previous (1991 - 2000) Migration Inflow (Destination)	-14.217***	-15.480***	-14.509***	-17.332***	
	(-3.26)	(-3.54)	(-3.33)	(-4.15)	
Previous (1991 - 2000) Migration Inflow (Origin)	-21.128***	-15.574**	-20.873***	-16.154**	
	(-2.73)	(-2.00)	(-2.69)	(-2.06)	
% of Pop. Born in the City (Origin)	17.779***	14.645**	17.666***	15.315***	
	(2.95)	(2.49)	(2.93)	(2.62)	
% of Pop. Born in the City (Destination)	12.850***	14.714***	13.194***	17.123***	
	(3.85)	(4.99)	(3.95)	(5.23)	
Pop. (Origin)	0.000**	0.000**	0.000**	0.000**	
	(2.40)	(2.15)	(2.36)	(2.01)	
Pop. (Destination)	0.000	0.000	0.000	0.000	
	(1.25)	(1.43)	(1.28)	(1.45)	
% of Iliteracy (Origin)	7.117	5.438	7.039	5.760	Education
	(1.07)	(0.87)	(1.05)	(0.88)	
% of Iliteracy (Destination)	3.909	7.403	4.386	10.056**	
	(0.92)	(1.60)	(1.03)	(2.22)	
% of Pop. w/ College Degree (Origin)	84.025***	62.877***	83.694***	69.635***	
	(4.66)	(3.77)	(4.61)	(4.00)	
% of Pop. w/ College Degree (Destination)	158.272**	164.826***	159.340**	172.900***	
	(2.46)	(2.64)	(2.48)	(2.66)	
Number of Academics (2005) (Origin)	-0.000**	-0.000*	-0.000**	-0.000*	
	(-2.03)	(-1.84)	(-1.99)	(-1.70)	
Number of Academics (2005) (Destination)	0.001	0.001	0.001	0.000	
	(0.67)	(0.66)	(0.66)	(0.53)	
Average Mills Ratio - "propensity to migrate" (Origin)	-30.294***	-24.135***	-30.041***	-25.184***	Selection
	(-5.02)	(-4.25)	(-4.97)	(-4.43)	
Average Mills Ratio - "propensity to migrate" (Destination)	-21.287***	-23.811***	-21.834***	-27.618***	
	(-5.64)	(-7.16)	(-5.80)	(-7.00)	
Constant term	148.105***	129.906***	147.960***	138.355***	
	(5.82)	(5.44)	(5.81)	(5.43)	
Centroid Distance (Origin - Destination) (log)	-6.139***	-4.917***	-6.126***	-5.383***	Spatial
	(-9.40)	(-9.00)	(-9.34)	(-9.03)	
Lambda	-	0.217***	-	0.135***	
	-	(8.47)	-	(4.98)	
Rho	-	-0.206***	0.017	-	
	-	(-7.31)	(1.56)	-	

Note 1: *** p<0.01, ** p<0.05, * p<0.1

Note 2: All variables refer to the year of 2000, except when described in the variable's name.

Note 3: Imigration to Amazon excludes migrantion flows within Amazonian municipalities.

Amazon Immigration[81]

SOURCE: Own Elaboration

[81] In this table we exhibit only the coefficients which proved to be signifficant, at least at the 10% level, in at least one among the 5 implemented estimators.

Table A.5. 2: "Hierarchical Heckit" Second Stage (spatial regression): sample composed only by Brazilian Amazon Emigration[82]

Explanatory Variable	*Heckit - OLS*	*Heckit - SAC - GS2SLS (Contiguous)*	*Heckit - SEM - GS2SLS (Contiguous)*	*Heckit - SAR - GS2SLS (Contiguous)*	*Variable Groiup*
Avg. Labor Income (Destination) (log)	-13.922*** (-5.71)	-11.069*** (-5.15)	-13.843*** (-5.65)	-11.149*** (-4.97)	Income and Empolyment
Avg. Labor Income (Origin) (log)	-1.237 (-0.54)	-1.085 (-0.48)	-1.261 (-0.55)	-1.344 (-0.58)	
Avg. Cost of Living (Destination) (log)	11.779*** (5.63)	9.699*** (5.35)	11.699*** (5.57)	9.554*** (5.04)	
Avg. Cost of Living (Origin) (log)	-2.360 (-1.11)	-0.816 (-0.39)	-2.237 (-1.05)	0.191 (0.08)	
per capita GDP (Destination) (log)	-1.296 (-1.57)	-0.927 (-1.11)	-1.282 (-1.54)	-0.883 (-1.05)	
per capita GDP (Origin) (log)	0.265 (0.37)	-0.446 (-0.58)	0.242 (0.34)	-0.463 (-0.52)	
EAP / WAP ratio (Destination)	10.678** (2.23)	8.253* (1.77)	10.652** (2.22)	8.919* (1.86)	
EAP / WAP ratio (Origin)	7.908 (1.39)	6.498 (1.08)	8.046 (1.41)	10.100* (1.67)	
Previous (1991 - 2000) Labor Income Growth (Destination)	1.488 (1.49)	0.844 (0.91)	1.466 (1.47)	0.799 (0.84)	
Previous (1991 - 2000) Labor Income Growth (Origin)	0.492 (0.52)	0.491 (0.54)	0.504 (0.54)	0.653 (0.67)	
Unemployment rate (Destination)	-39.853*** (-4.41)	-31.599*** (-3.52)	-39.573*** (-4.36)	-31.348*** (-3.50)	
Unemployment rate (Origin)	-18.982** (-2.35)	-15.585* (-1.85)	-19.255** (-2.38)	-22.666*** (-2.65)	
Gini Index (Origin)	3.123 (0.30)	6.835 (0.64)	3.109 (0.30)	4.260 (0.39)	Unequality
Gini Index (Destination)	4.843 (0.51)	4.564 (0.51)	4.699 (0.49)	2.619 (0.28)	
% of Pop. w/ Access to Sanitary Treatment (Origin)	11.339*** (2.98)	9.787*** (2.77)	11.297*** (2.97)	10.171*** (2.70)	
% of Pop. w/ Access to Sanitary Treatment (Destination)	3.281** (2.11)	2.706* (1.80)	3.265** (2.09)	2.679* (1.71)	
% of Labor Force in Services and Commerce (Origin)	20.624*** (4.07)	19.816*** (4.15)	20.719*** (4.10)	22.337*** (4.29)	Demographic & Urbanization
% of Labor Force in Services and Commerce (Destination)	17.247*** (4.31)	14.061*** (3.58)	17.279*** (4.30)	15.791*** (3.85)	
Pop. Density (Origin)	0.010* (1.65)	0.006 (1.04)	0.010 (1.61)	0.003 (0.57)	
Pop. Density (Destination)	-0.001 (-1.43)	-0.001 (-0.98)	-0.001 (-1.42)	-0.001 (-1.04)	
% of Pop. aged 22 - 29 (Origin)	13.514 (0.35)	39.674 (1.07)	14.381 (0.37)	40.192 (1.01)	
% of Pop. aged 55 - 60 - "about to retire" (Origin)	-8.321 (-0.06)	-42.209 (-0.31)	-10.023 (-0.07)	-50.745 (-0.35)	
% of Pop. aged more than 60 - "retired" (Origin)	-101.252 (-1.61)	-77.032 (-1.26)	-101.657 (-1.62)	-98.469 (-1.54)	

SOURCE: Own Elaboration

[82] In this table we exhibit only the coefficients which proved to be signifficant, at least at the 10% level, in at least one among the 5 implemented estimators.

Table A.5.2 (cont.): "Hierarchical Heckit" Second Stage (spatial regression): sample composed only by Brazilian Amazon Emigration[83]

Explanatory Variable	*Heckit - OLS*	*Heckit - SAC - GS2SLS (Contiguous)*	*Heckit - SEM - GS2SLS (Contiguous)*	*Heckit - SAR - GS2SLS (Contiguous)*	*Variable Groiup*
% of Pop. aged 22 - 29 (Destination)	411.505*** (7.81)	321.349*** (7.18)	409.253*** (7.73)	331.475*** (7.11)	Demographic & Urbanization
% of Pop. aged 55 - 60 - "about to retire" (Destination)	-53.951 (-0.70)	-40.086 (-0.54)	-53.662 (-0.69)	-38.000 (-0.49)	
% of Pop. aged more than 60 - "retired" (Destination)	-262.920*** (-5.50)	-212.558*** (-4.75)	-262.494*** (-5.46)	-229.371*** (-4.93)	
Previous (1991 - 2000) Migration Inflow (Destination)	-21.119*** (-2.86)	-17.782*** (-2.69)	-20.864*** (-2.81)	-15.575** (-2.16)	
Previous (1991 - 2000) Migration Inflow (Origin)	-2.073 (-0.47)	-2.906 (-0.68)	-2.164 (-0.49)	-4.215 (-0.92)	
% of Pop. Born in the City (Destination)	33.802*** (5.78)	26.807*** (5.00)	33.813*** (5.76)	30.088*** (5.14)	
% of Pop. Born in the City (Origin)	17.090*** (3.28)	17.213*** (3.30)	17.388*** (3.34)	22.746*** (3.91)	
Pop. (Origin)	0.000 (0.93)	0.000 (1.15)	0.000 (0.97)	0.000 (1.53)	
Pop. (Destination)	0.000 (1.55)	0.000 (1.64)	0.000 (1.53)	0.000 (1.51)	
% of Iliteracy (Origin)	23.646*** (3.62)	21.153*** (3.27)	23.854*** (3.64)	26.198*** (3.85)	Education
% of Iliteracy (Destination)	1.003 (0.13)	3.353 (0.45)	1.038 (0.14)	3.168 (0.41)	
% of Pop. w/ College Degree (Origin)	-10.765 (-0.25)	-4.272 (-0.11)	-10.086 (-0.24)	5.333 (0.12)	
% of Pop. w/ College Degree (Destination)	217.057*** (6.52)	176.152*** (6.41)	216.139*** (6.45)	182.203*** (6.09)	
Number of Academics (2005) (Origin)	0.000 (0.48)	0.000 (0.45)	0.000 (0.46)	0.000 (0.02)	
Number of Academics (2005) (Destination)	-0.000 (-0.94)	-0.000 (-1.16)	-0.000 (-0.93)	-0.000 (-0.99)	
Average Mills Ratio - "propensity to migrate" (Origin)	-27.093*** (-4.54)	-26.793*** (-4.50)	-27.487*** (-4.60)	-34.222*** (-5.03)	Selection
Average Mills Ratio - "propensity to migrate" (Destination)	-43.406*** (-7.60)	-33.990*** (-6.58)	-43.308*** (-7.55)	-36.763*** (-6.52)	
Constant term	145.810*** (5.30)	109.819*** (4.41)	145.759*** (5.29)	125.874*** (4.57)	
Centroid Distance (Origin - Destination) (log)	-6.108*** (-8.36)	-4.317*** (-6.50)	-6.095*** (-8.28)	-4.958*** (-6.73)	Spatial
Lambda	- -	0.298*** (7.03)	- -	0.191*** (3.91)	
Rho	- -	-0.289*** (-6.39)	0.022 (1.03)	- -	

Note 1: *** p<0.01, ** p<0.05, * p<0.1

Note 2: All variables refer to the year of 2000, except when described in the variable's name.

Note 3: Emigration from Amazon excludes migrantion flows within Amazonian municipalities.

SOURCE: Own Elaboration

[83] In this table we exhibit only the coefficients which proved to be signifficant, at least at the 10% level, in at least one among the 5 implemented estimators.