Université catholique de Louvain
Faculté des Sciences Economiques
Sociales et Politiques
Département des Sciences Economiques

Center for Operations Research and Econometrics

Rent seeking and hubs: towards a New Economic Geography of Sub-Saharan Africa

Alain Pholo Bala
Juin 2009

Thèse présentée en vue de l'obtention du grade
de docteur en sciences économiques

Composition du jury:

Jacques-François Thisse (co-promoteur)	Université Catholique de Louvain
Dominique Peeters (co-promoteur)	Université Catholique de Louvain
Luc Bauwens	Université Catholique de Louvain
Kristian Behrens	Université du Québec à Montréal
Luisito Bertinelli	Université du Luxembourg

Contents

Acknowledgements

Several people have contributed to the completion of this thesis. First and foremost, I would express my gratitude to my two supervisors, Jacques-François Thisse and Dominique Peeters. Jacques-François Thisse kindly accepted to be my supervisor and provided crucial insights about most of the issues addressed in this thesis. His great scientific inspiration was decisive to put this thesis on the track. Dominique Peeters was the first to show interest in my research project. He constantly assisted me with his guidance and helped to solve many practical problems such as funding. I am also deeply thankful to Pierre Dehez whose decisive support allowed me to undertake graduate studies in Belgium.

Several persons contributed to specific parts of this thesis. I am deeply indebted to Luc Bauwens who constantly advised me in the elaboration of the second chapter and put his strong expertise in contribution to help me solving tricky issues. I am also very thankful to Kristian Behrens who is my coauthor for the third chapter. He owns large stakes of this part of the thesis and contributed to improve the other chapters. I am gratefull to Luisito Bertinelli as well. He helped me constituting my database and provided me valuable advices and remarks.

I spent all these years at CORE, a very stimulating place for innovative scientific research. I am very thankful to all its administrative staff for its kind collaboration.

My Graduate studies at UCL were the occasion to meet wonderful people. I would like to mention in particular Youyou Baende, Michel Tenikue, Jean-Baptiste Ntagoma, André Nyembwe, Raouf Boucekkine, Sven Wauters, Fausto Galli, Eric Rengifo, Eve Ramaekers, Isabelle Thomas, Claire Dujardin, Fernanda Estevan, Jean-François Maystadt, Kenmei Tsubota, Tanja Mlinar, and Filippo Calciano for their friendship.

I am also grateful to Jean-Pierre Miankodila and Lambert Nieme for their persistent encouragements. Finally, I am strongly indebted to Viviane, my wife. I can not thank her enough for her support, confidence and patience.

Chapter 1

General Introduction

> "*The truth of the matter is that African cities are part of the cause and a major symptom of the economic crisis that has enveloped the continent* [...] *Their inability to serve as veritable engines of growth and structural transformation as cities in other societies is certainly a serious element in the present equation. This inability is a function of their historical background, their colonial evolution, and their transitional characteristics within changing modes of production*" (Mabogunje, 1994, p. xxi)

Sub-Saharan Africa urbanization process raises concerns. The poor economic performances of that region[1], has fueled the strong prejudice that its cities have not been an impulse for economic growth. Yet, this bias has not triggered a corresponding interest in the literature. Africa is hardly covered by the few formal urban studies focusing on developing countries. Those studies address mostly urbanization issues in Latin America or in Asia. The fact that Africa is generally depicted as the least urbanized region and as deprived of significant agglomerations economies may give some rationale to this lack of interest.

Yet, Sub-Saharan Africa urbanization provides an interesting story for investigation. It is a dynamic process as evidenced by the fact that this region features the highest urban growth rates in world. A persistently rising demography of Africa biggest cities and a massive population redistribution from rural to urban areas fuel this urban spurt. Such a high urbanization speed reflects the major social and economic changes arising in those countries in a context of increasing trade integration and technological development.

This thesis is a modest attempt to uncover the main features and the underlying drives of Sub Saharan urbanization process. In the second chapter we begin by describing the

[1]Seventy percent of the one billion people living in countries caught in development traps, the so-called 'bottom billion' (Collier, 2007), are from Africa.

stylized facts of urbanization in that region. Sub-Saharan Africa's urban pattern features urban bias and urban primacy. Therefore, spatial distribution in that region is skewed. Locational advantages and political effects have induced a bias in favor of political capitals and ports. Second nature determinants, i.e. agglomeration economies arising through the interplay of increasing returns to scale, while constituting a major urbanization drive, seem to be less noticeable in that part of the World.

Thus, literature seems to emphasize rent-seeking and hubs determinants at the expense of agglomeration economies strenghening the prejudice against Sub-Saharan Africa biggest cities. In the third chapter we question that preconception. To do so we perform an empirical analysis investigating the relationship between urban concentration and economic growth. Henderson (2003) performed a similar study and found that economic growth follows an inverted U-shape pattern in terms of urban concentration and that there is an optimal degree of urban concentration in terms of maximizing economic growth. In our analysis we relax the assumption that the economic growth rate is a concave function of urban concentration. To allow such flexibility, we model economic growth using a semi-parametric function, with the non-parametric term depending on urban concentration. Such an approach has the advantage of allowing subsequently to check out if urbanization patterns differ across group of countries by performing tests of equality of non-parametric functions of different subsamples.

A major challenge of this study is the explicit handling of the data missingness problem. Most of empirical papers in economics handle missing data by listwise deletion i.e. by deleting any observation having at least one missing datum. This approach has faced a lot of criticism by the statistics literature. Indeed, apart from the fact that it implies the loss of all the information conveyed in the observations having missing data, the estimates obtained with such a method have been proved to be biased if the data remaining after deletion do not constitute a random sample of the overall database. In order to avoid such a shortcoming, we implement a two-step multiple imputation Bayesian algorithm. Our results outline heterogeneity between regions: urban concentration has a positive impact on economic growth in Europe, poverty traps prevail for Latin America and Asia, while Africa non-parametric curve differs, depending of the measure of urban concentration considered.

Such empirical findings pointing out the specificity of the urbanization patterns in different regions, prompt us to better characterize the underlying drives of Sub-Saharan Africa urbanization process. This was the challenge of our theoretical investigation which is twofold. On one hand we have attempted in the fourth chapter to explain urban agglomeration in Sub-Saharan Africa by both economic and political factors as done in Robert-Nicoud and

Sbergami (2004). However, our political process differs significantly from the one they used. Robert-Nicoud and Sbergami (2004) use a probabilistic voting model which leads to a more even distribution of economic activities than the market mechanism. As democracy doesn't describe adequately politics in Sub-Saharan Africa, we use rather a different political process which consists in agents deciding on whether or not to enter a political elite in order to extract rents to maximize their own welfare. We find that this rent-seeking behavior fuels the formation of large urban agglomerations in developing countries, via mechanisms of interregional income transfers. Such a finding may serve to explain a seemingly paradoxical aspect of urban development in SSA: agglomeration despite high trade costs. Indeed, the NEG literature roughly predicts a positive and monotone relationship between freeness of trade and the degree of agglomeration, whereas most SSA countries are characterized by high values of interregional trade costs.

On another hand we rely on international trade and hub effects to explain the formation of agglomerations. Allowing for locational advantages through hubs is crucial. Indeed, without any spatial heterogeneity it is more likely to obtain the opposite result that trade integration fosters spatial dispersion since it undermines backward and forward linkages triggered by import-substitution (Krugman and Elizondo, 1996). With international trade and locational advantages we are thus able to explain the evidence of the persistence and the fostering of huge agglomerations in coastal locations of Sub-Saharan Africa. According to the two models developed in fifth chapter, we find that openness is likely to trigger agglomeration in the hub especially when transport costs are low. This result is consistent with Weber's theory of location (Beckmann and Thisse, 1986) which states that in in a star-shaped network without any dominant location, entry points are the optimal locations. Those results shed light on agglomeration processes in SSA. Indeed, SSA is characterized by the stability of spatial concentration of economic activities along coastal locations. This persistence of the location of several of its biggest cities in hubs is quite appealing, especially in the context of increasing trade integration facilitated by the increasing efficiency of transport technologies and by the general decrease of tariffs.

Chapter 2

Cities in Africa

2.1 Africa's urban growth

During the last fifty years, cities of the developing world have experienced an outstanding demographic growth. While in 1950, some 309 millions inhabitants were living in cities of developing countries, i.e. 42% of a total urban population evaluated at 732 millions, by 2005 the world's urban population had reached 3.15 billions, with 71.5% (2.25 billions) of inhabitants living in the cities of developing countries (United Nations, 2006). In 2005, cities in less-developed countries (LDCs) comprised 12 of the 15 largest, with a combined population in excess of 277 millions. Despite of a challenging and somewhat hostile economic environment, changing demographic trends characterized by the persistence of high fecondity rates coupled with mortality rates declines, have allowed cities of developing countries to keep an outstanding demographic dynamism.

Nowhere is this dilemma more visible than in the congested cities of Sub-Saharan Africa, where projections of urban population growth remain the highest in spite of the poor economic performance characterizing countries of this region. As Table 2.2 shows since 1950 Sub-Saharan Africa urban growth rates are the highest of the world. More precisely, it is the only region in the world that still has an urban growth rate in excess of 3% after 1995.[1] While current rate of population growth in two of the world's largest cities, Tokyo and New York, is about 1% per annum it rises to more than 6% per annum in many African cities, including Nairobi, Lagos, and Lusaka. African cities have even uniformly higher growth rates and more rapidly expanding numbers than many cities in Asia and Latin America, although their absolute numbers of inhabitants are lower. Table 2.1 provides figures illus-

[1] Previously it was even greater than 4%. From 1995 urban growth rates have dropped below 3% in other parts of the world.

trating that contrast. From 1975 to 2000 African cities of 1 to 5 millions inhabitants grow the fastest. Likewise African cities from 5 to 10 millions inhabitants expand more rapidly than other regions megalopolises except Asian cities of more than 10 millions inhabitants.

Lagos in Nigeria is in Africa the best example of a massive and fast demographic growth. It did not even appear on the list of the thirty biggest cities in the World until occupying the twenty-seventh position in 1995, and is expected to jump to the eleventh position by 2015.

Such an urban population growth raises serious concern about the capacity of African cities to deal economically, environmentally, and politically with acute concentrations of people. Whereas cities are expected to offer the cost-reducing advantages of agglomeration economies and economies of scale and proximity as well as numerous economic and social externalities (pool of skilled workers, cheap transport, social and cultural amenities), this outstanding urban's population growth induces an inflation of the social costs of a progressive overloading of housing and social services, of increased crime, pollution, and congestion to such an extent that they may gradually outweigh the expected urban advantages.

The prolific growth of huge slums has even triggered that concern. Today shanty town settlements represent more than one-third of the urban population in all developing countries; in many cases they account for 60 percent or more of the urban total as shown in Table 2.3. In Sub-Saharan Africa, this proportion is even higher. While in the late 1980s, 72 of every 100 new households established in urban areas of developing countries were located in shanties and slums, in Africa, the statistic reached 92 households out of every 100 (Todaro, 1997).

Table 2.1: Distribution of the total population of major areas by type of settlement and size of urban settlement, 1975, 2000 and 2015

Major area	Type of settlement and number of inhabitants of urban settlement	Total population (millions)			Percentage distribution			Growth rate (percentage)	
		1975	2000	2015	1975	2000	2015	1975-2000	2000-15
Africa	Total population	406	794	1 110	100.0	100.0	100.0	2.68	2.24
	Urban Population	102	295	503	25.2	37.2	45.3	4.23	3.56
	10 million or more	0	0	27	0.0	0.0	2.5	..	..
	5 million to 10 million	6	23	21	1.5	2.9	1.9	5.35	-0.63
	1 million to 5 million	12	64	123	3.0	8.1	11.1	6.67	4.35
	500,000 to 1 million	14	26	43	3.5	3.3	3.9	2.50	3.31
	Fewer than 500,000	70	181	288	17.3	22.9	25.9	3.80	3.08
	Rural areas	304	498	607	74.8	62.8	54.7	1.98	1.31
Asia...............................	Total population	2,397	3,672	4,371	100.0	100.0	100.0	1.71	1.16
	Urban Population	592	1 376	2 005	24.7	37.5	45.9	3.37	2.51
	10 million or more	31	136	214	1.3	3.7	4.9	5.90	3.01
	5 million to 10 million	47	87	162	1.9	2.4	3.7	2.50	4.12
	1 million to 5 million	147	318	466	6.2	8.7	10.7	3.08	2.55
	500,000 to 1 million	78	150	193	3.3	4.1	4.4	2.61	1.72
	Fewer than 500,000	289	684	969	12.0	18.6	22.2	3.45	2.32
	Rural areas	1,805	2,297	2,366	75.3	62.5	54.1	0.96	0.20
Europe	Total population	676	727	705	100.0	100.0	100.0	0.29	-0.21
	Urban Population	455	534	538	67.3	73.4	76.3	0.64	0.04
	10 million or more	0	0	0	0.0	0.0	0.0	..	..
	5 million to 10 million	37	32	32	5.4	4.4	4.6	-0.52	0.01
	1 million to 5 million	82	110	117	12.2	15.1	16.5	1.15	0.41
	500,000 to 1 million	46	49	48	6.8	6.7	6.8	0.22	-0.14
	Fewer than 500,000	290	344	341	42.9	47.2	48.4	0.67	-0.05
	Rural areas	221	193	167	32.7	26.6	23.7	-0.54	-0.98
Latin America and the Caribbean	Total population	322	519	630	100.0	100.0	100.0	1.91	1.30
	Urban Population	198	391	507	61.4	75.4	80.5	2.73	1.73
	10 million or more	21	59	66	6.5	11.3	10.5	4.11	0.82
	5 million to 10 million	17	20	36	5.3	3.8	5.6	0.56	3.94
	1 million to 5 million	32	86	139	10.1	16.5	22.1	3.90	3.23
	500,000 to 1 million	18	39	45	5.5	7.5	7.2	3.14	1.00
	Fewer than 500,000	109	188	221	34.0	36.3	35.0	2.17	1.07
	Rural areas	124	127	123	38.6	24.6	19.5	0.10	-0.25
Northern America.........	Total population	243	314	356	100.0	100.0	100.0	1.02	0.83
	Urban Population	180	243	288	73.8	77.4	81.1	1.21	1.14
	10 million or more	16	30	32	6.5	9.5	9.1	2.54	0.53
	5 million to 10 million	16	7	13	6.4	2.2	3.7	-3.23	4.28
	1 million to 5 million	52	85	100	21.3	27.0	28.2	1.96	1.14
	500,000 to 1 million	17	26	23	7.0	8.4	6.5	1.73	-0.81
	Fewer than 500,000	79	95	119	32.6	30.3	33.4	0.73	1.50
	Rural areas	64	71	67	26.2	22.6	18.9	0.44	-0.37

Source: United Nations, 2002.

Table 2.2: Average annual rate of change of urban population 1950-2005

	Period										
Region	50-55	55-60	60-65	65-70	70-75	75-80	80-85	85-90	90-95	95-00	00-05
World	3.01	3.09	3.08	2.76	2.63	2.72	2.66	2.70	2.33	2.18	2.04
Dev Regions	2.37	2.31	2.13	1.79	1.48	1.21	0.94	0.96	0.75	0.58	0.54
Less Dev. Reg.	3.84	4.04	4.14	3.74	3.68	3.93	3.87	3.76	3.17	2.93	2.67
SSA	4.38	4.61	4.83	4.66	4.40	4.46	4.37	4.28	4.42	3.93	3.61
Asia	3.57	3.72	3.80	3.45	3.39	3.74	3.78	3.78	3.09	2.88	2.61
Europe	2.11	2.09	2.06	1.64	1.52	1.19	0.80	0.78	0.37	0.14	0.13
South Am.	4.42	4.27	4.29	3.94	3.74	3.53	3.05	2.74	2.35	2.16	1.98
North Am.	2.66	2.61	2.04	1.60	0.98	1.00	1.23	1.24	1.57	1.51	1.37

Source: United Nations (2006)

Those downsides in urban development in Africa prompt to better characterize the urbanization process in that region. This is done in the next section where we describe some significant features of Africa urbanization process.

2.2 Urbanization in Sub-Saharan Africa: some stylized facts

In this section we describe some stylized facts that seem recurrent in Sub-Saharan Africa Urbanization process. It is not intended here to present them as specific to that region. We just outline that they are particularly vivid in that part of the world. Urban development in Sub-Saharan Africa is characterized by a strong urban bias, by urban primacy, by a significant role played by hub and political effects and by a remarkable development of the informal sector.

2.2.1 Urban bias

Since political independence, economic development policies in most of Africa were underpinned mostly on import-substitution strategies. With their emphasis on industrial modernization, technological sophistication, modern education and metropolitan growth, such strategies created a substantial geographic imbalance in economic and noneconomic opportunities between rural and urban areas and therefore contributed significantly to the

Table 2.3: Residents of slums and squatter settlements as a percentage of urban population, by region and city

Region/city	**Slum dwellers as percentage of city population**
Latin America	
Bogota,Colombia	60
Mexico City,Mexico	46
Caracas,Venezuela	42
Middle East and Africa	
Addis Ababa, Ethiopia	79
Casablanca, Morocco	70
Ankara, Turkey	60
Cairo, Egypt	60
Kinshasa, Zaire	60
Asia	
Calcutta, India	67
Manila, Philippines	35
Seoul, South Korea	29
Jakarta, Indonesia	26

Source: Population Crisis Committee (1983), World Population Growth and Global Security, Report No. 13, p.2, as cited by Todaro (1997).

steadily accelerating influx of rural migrants into urban areas (Todaro, 1997; Mabogunje, 1994). Some of the instruments by which they induced such a favoritism toward urban areas are exchange rate, tariff and taxes policies that penalized rural dwellers more than urban inhabitants (Montgomery *et al.*, 2004).

However, those price distortions are not the only expression of urban bias. Expenditure biases have played an important role as well (Lipton, 1983). Instead of opting to invest sufficiently in interregional transport and telecommunications, national governments have rather chosen to undertake other public investments in infrastructure and social spending in primate cities. Therefore, public employment have expanded in capitals well beyond the efficient level (Becker *et al.*, 1994).

The prevalence of such strong urban bias is somewhat unexpected from countries where most of the population lives in the countryside. It may have been partly induced by a belief largely popularized by Kuznet's hypothesis that at the outset of development process the rising of inequalities may be beneficial in terms of economic growth. With respect to a urban context such an hypothesis has been translated in the assumption that urban concentration is helpful to conserve on physical infrastructure capital (transport and telecommunications) and managerial resources and to enhance information spillovers and knowledge accumulation when the economy is 'information deficient' (Henderson, 2003). This may have fed the pro-urban prejudice that the traditional rural sector should provide the resources that should allow the modern urban sector to develop. But such 'dispositional urban bias' (Lipton, 1983) cannot explain the all story. In a subsequent development we will show that African Political Economy may provide additional clues to grasp the assymetric spatial allocation of resources.

2.2.2 Urban primacy

Because of urban bias, the urbanization process in Sub-Saharan Africa leads to a major population redistribution between rural and urban areas. However, among urban centers themselves, this redistribution showed a strong preference for very large cities and metropolitan areas (Mabogunje, 1994). This depicts SSA countries as characterized by *urban primacy*. Urban primacy can be measured in a number of ways including the proportion of urban population leaving in the largest city or as the ratio of the largest city population to the population of the second largest city. The largest city is perceived as primate if it has a population bigger than twice the population of the second largest city. The contrary of urban primacy is *rank size regularity* which implies that, in a given national urban hierarchical system, the population of any given town is inversely proportional to its rank in the hierarchy, the underlying assumption being that countries respecting the rank size rule have a well balanced and well integrated system of cities, whereas those characterized by primate city size exhibit a lopsided, and loosely integrated urban system (Aryeetey-Attoh, 1997). Table 2.4 illustrates clearly the prevalence of primacy in Sub-Saharan Africa. There is only few countries in that region showing up levels of primacy below 20%: South Africa (11.1%) and Nigeria (17.1%) are among those few. While we may expect South Africa, with its high level of economic development, to be characterized by a more balanced urban system Nigeria departs from the typical pattern of the region because its large population is hardly compatible with the existence of just one primate city. We see also that while primacy is typical of small countries like Gambia, Burundi, Rwanda and Togo, where the capital cities dominate the landscape; it is also a characteristic of larger countries like Gabon, Angola, Congo, Ivory Coast, Senegal.

As described previously biggest cities in Africa expands very fast. The percentage of urban population living in Sub-Saharan Africa cities of more than 500,000 inhabitants rose from 6 to 41 between 1960 and 1980, and the number of such centers increased from 3 to 28 (World Bank, 1989 quoted by Mabogunje, 1994). Moreover, the number of African agglomerations with at least one million inhabitants had increased from 2 in 1950 to 21 in 1990 (Max Miller and Singh, 1994). Smaller cities in Africa grow at a much slower pace. Figure 1 shows clearly that between 1975 and 2000 smaller cities in Africa had a much lower growth rate.[2] In fact Africa lacks networks of small and middle-sized cities (Aase, 2003). Thus, instead of a progressive and step-by-step migration from small towns to cities and then from cities to metropolitan centers, there is a brutal one-stage migration from rural background to a life in crowded and huge cities and metropolitan areas.

With an average of primacy equal to 34.97%, SSA occupies the second rank among regions after Latin America, and has a primacy that is significantly higher than urban concentration in Europe. Moreover, as table 2.5 shows, in statistical terms, there is no significant difference between the extent of primacy in South America and in Africa.[3]

So despite the fact that population in African cities is lower than in many Latin American cities and that urbanization rates are lower in Africa, the 'small' urban populations in Africa shows up a distribution as skewed as South America urban population distribution reflecting strong imbalances in the spatial repartition of economic resources. Therefore, although urban populations in Africa are quite smaller in absolute terms, strong urban concentration is likely to be crucial feature of African urban development.[4]

The notion of primacy may be extended beyond population size to reflect the fact that African primate cities drain a disproportionate amount of social, cultural, economic and administrative resources. Dar es Salaam, for instance, accounts for a disproportionate share of Tanzania's manufacturing jobs (50.3% in 1978), value added manufacturing (56.9% in 1974), and social welfare expenses (Aryeetey-Attoh, 1997). We will expand further on that when describing political factors lying behing urban concentration in SSA.

[2]Figures related to the period between 2000 and 2015 are merely estimations.

[3]Urban primacy is measured here by the percentage of urban population living in the country's largest city. This data is taken from the UN Populations Division "World Urbanization Prospects Population Database: The 2003 Revision" for the year 2000, available online at `http://esa.un.org/unup/`.

[4]Literature is quite divergent about the prevalence of primacy in Sub-Saharan Africa. While Mabogunje (1994) and Aryeetey-Attoh (1997) assert that urban systems in that region are primate, Becker *et al.* (1994) and Kessides (2005) consider that urban concentration is not significantly higher than international standards. Statistical evidence provided by table 2.5 comforts the first claim.

Table 2.4: Urban Primacy in Sub-Saharan Africa

Country	% Urban Population in Largest city 1980	1990	2000
West Africa			
Benin		12.0	31.5
Burkina Faso	41	50.6	38.3
Cote d'Ivoire	34	44.5	44.3
Gambia		100.0	100.0
Ghana	35	27.1	19.4
Guinea	80	76.1	46.6
Liberia		56.6	43.2
Mali	24	37.4	31.0
Mauritania	39	80.3	39.3
Niger	31	38.5	34.0
Nigeria	17	20.4	17.1
Senegal	65	55.9	44.2
Sierra Leone	47	51.7	49.5
Togo	60	56.1	52.6
Central Africa			
Cameroon	21	21.9	22.4
Central African Rep	36	50.8	45.6
Chad	39	43.6	42.6
Congo	56	68.6	54.4
Gabon	56		59.7
Zaire	28	33.1	32.2
East and NE Africa			
Burundi		82.5	67.4
Ethiopia	37	29.2	25.5
Kenya	57	27.0	20.3
Rwanda	54		62.5
Somalia	13	12.4	36.5
Tanzania	50	25.4	18.8
Uganda	52	35.2	39.3
Southern Africa			
Angola	64	62.6	56.6
Bostwana		36.2	23.0
Madagascar	36	23.6	36.0
Malawi	19		34.2
Mozambique	83	40.7	19.1
Namibia		30.3	40.5
South Africa	13	12.4	11.1
Swaziland		19.2	28.9
Zambia	35	23.5	35.7
Zimbabwe	50	31.8	32.6

Source: For 1980 and 1990 data United Nations (1991), *United Nations Statistical Yearbook*, Washington, DC: United Nations; World Bank (1990), *World Development Report*, New York: Oxford University Press; Rondenelli (1983), *Secondary Cities in Developing Countries, Beverly Hills: Sage* as cited by Aryeetey-Attoh (1997). For 2000 data Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat (2004). *World Urbanization Prospects: The 2003 Revision.* Data Set Name (POP/DB/WUP/Rev.2003/Data set number/File number), data set in digital form.

2.2.3 Hub and political effects

The skewed distribution of Africa population is further revealed by the coastal location of its largest cities. Indeed, because of the heavy import dependence of Sub-Saharan Africa economic strategies, port cities became the preferred location for industrial development (Mabogunje, 1994). Such a locational preference for coastal sites is not new, it holds since the beginning of colonization. At that time access to ports was vital for the outward shipment of raw materials back to the colonizing countries and the inward shipment

Table 2.5: Mean comparison tests of primacy between SSA and other regions

	% of urban population in the largest city		
SSA versus	Mean	Mean comparison test	p-value
Europe	24.56	$\mu_{SSA} > \mu_{EU}$	0.0005
Latin America	38.50	$\mu_{SSA} \neq \mu_{LA}$	0.3737
Asia	30.70	$\mu_{SSA} \neq \mu_{AS}$	0.1968

of manufactured goods. Consequently, the spatial structures of most African economies became strongly focused on a small number of port cities. From 1960 there has been a remarkable stability in the location of primate cities in SSA. Indeed, with the end of the colonial era, population redistribution toward these coastal cities did not cease. On the contrary, it increased as their cities retained and extended their dominance as the primary centers of economic activities (Kempe, 1996).

Because in most cases, ports were also the preferred location of colonial administrations especially in those countries without large European settler populations, these cities became at once the capital city, the port city, and the nascent industrial center of most African countries (Mabogunje, 1994; Christopher, 1994). This tendency proved to be very enduring even after the colonial period. Indeed, even after the independence in most cases, administrators of SSA countries continued to follow economic trends by locating capital cities at the hub of the post-colonial economy. Therefore, the majority of Sub-Saharan Africa Capitals is established on the coast at the point of entry and export, the initial location of capital economies being persistent with 28 of the 44 coastal colonial capitals in existence in 1900 retaining that status in 1991 (Christopher, 1994).

The location of central administration in those harbour cities induced a primate city development well beyond economic lines. Indeed, in all cases the extent of the bureaucracy often became the most significant aspect of the city's development. Being the locus of administration, capital cities are vested with an aura of sovereignty: there are the places from which political power is exercised over a state area. Where the city is also the main industrial and commercial center of the state, the government functions may occupy a distinct quarter. But in the absence of major industrial and commercial complexes in many African countries the economy of the capital remains heavily dependent upon government functions (Christopher, 1994).

The primate role of capital cities within the national urban hierarchy is further strengthened by the fact that a disproportionate part of the national budget is spent within them. That is where African Political Economy plays a role: Political leaders allocate an economically indefensible share of resources to primate cities because they fear the pressure of city elites (Lipton, 1983; Bairoch, 1988). Indeed, in non democratic countries spatial proximity is likely to increase political influence, the leadership is being more sensitive to the claims of the urban elites than to those of people living in the country side (Ades and Glaeser, 1995). This highlights the poor enforcement of property rights in those countries. New Institutional Economics explains the inefficient property rights enforcement prevailing in developing countries by the fact that in those countries rulers avoid to adopt rules opposed to the interest of the most powerful social groups (North, 1990).

Table 2.6 clearly reveals that SSA countries have the lowest non-corruption index. Thus, they are the more corrupted countries and are likely to show up the worst enforcement of property rights. They are also the nations characterized by the highest central governments expenditures.[5] This confirms the description frequently made that African countries are characterized by a strong interventionism in the economy. With the highest corruption level and the largest central governement expenditures on goods, services, and compensation of employees, SSA appears as the developing region where politics has the strongest impacts on regional imbalances and urban primacy.

The fact that capital cities are the place where the greatest share of public expenses are made explains why they are usually the subject of massive in-migration by those seeking economic and political opportunities not available elsewhere. However, because of the failure of their industrialization strategies, most of those primate cities are less the focus of manufacturing production than that of conspicuous and unwarranted consumption. But because of the economic recession characterizing that region the productive and consumptive capacity of African cities is not likely to be sustainable over the long run. Moreover, infrastructural facilities are barely maintained and the delivery of most urban services became erratic and subject to severe shortages (Mabogunje, 1994). As modern housing could no longer keep up with the pace of rural-urban migration, the expansion of those primate cities take mostly the form of shanty towns development. Furthermore, the shrinking of

[5] Non-corruption index is the "2005 Transparency International Corruption Perceptions Index". It is available online at http://www.transparency.org/policy and research /surveys indices/cpi/2005. Data on Central government expenditures on goods, services, and compensation of employees in % of central government expenditures in 2003 are from the "2005 World Development Indicators", available online at http://www.devdata.worldbank.org/wdi2005/cover.htm.

Table 2.6: Other mean comparison tests between SSA and other regions

SSA versus	Mean	Mean comparison test	p-value
	Non-corruption index		
SSA	2.77		
Europe	5.81	$\mu_{SSA} < \mu_{EU}$	0.0000
Latin America	3.39	$\mu_{SSA} < \mu_{LA}$	0.0139
Asia	3.40	$\mu_{SSA} < \mu_{AS}$	0.0134
SSA versus	Mean	Mean comparison test	p-value
	Central government expenditures		
SSA	60.27		
Europe	27.40	$\mu_{SSA} > \mu_{EU}$	0.0000
Latin America	41.94	$\mu_{SSA} > \mu_{LA}$	0.0025
Asia	48.79	$\mu_{SSA} > \mu_{AS}$	0.0473

economic opportunities in the formal economy of those countries often reduce migrants to the only employment alternatives remaining in the informal sector. Therefore, most of the economic dynamism of SSA relies on the shadow economy which appears as a crucial feature of African cities.

2.2.4 Urban informal sector

A common characteristic of African cities is the dualistic nature of their underlying economy. African cities are characterized by the coexistence of a formal sector including public and private sector enterprises that are officially recognized, nurtured, and regulated by the government and an 'shadow economy' gathering economic agents operating outside the mainstream of government activity, regulation, and benefits (Todaro, 1997; Aryeetey-Attoh, 1997). Escaping law enforcement is the mere motivation of the formation and the development of the informal sector. Indeed, according to Schneider (2004, 4-5), the informal sector includes all market-based legal production of goods and services that are deliberately concealed from public authorities for the following reasons:

1. to avoid payment of income, value added or other taxes,

2. to avoid payment of social security contributions,

3. to avoid having to meet certain legal labor market standards, such as minimum wages,

maximum working hours, safety standards, etc., and

4. to avoid complying with certain administrative procedures, such as completing statistical questionnaires or other administrative forms.

The existence of this unregulated, and mostly legal but unregistered informal sector was acknowledged in the early 1970s. Its development has been prompted by the rising inability of the formal sector to accommodate in-migrants searches of jobs. It is difficult to provide precise figures on the extent of informal activities because of definitional problems and inadequate methods of data collection. However, it is estimated that the informal sector grew by 6.7% a year between 1980 and 1989 and employed more than 60% of the workforce in Sub-Saharan cities in 1990. African cities differ in the proportions of informal workers: Charmes (1990), as quoted by Aryeetey-Attoh (1997), estimates a high of 73% in Burkina Faso; 65% in Kumasi (Ghana) and Niger; 50% in Lagos (Nigeria), Lome (Togo), and Senegal; 44% in Nairobi (Kenya) and Ivory Coast, and a low of 20% in Djibouti.

The informal sector is mostly composed of informal, small scale, and labor intensive firms. The contribution of those firms to Sub-Saharan Africa economies, while difficult to estimate appears as substantial. According to Schneider (2004, Table 5.4), the average size of its shadow economy expressed as a percentage of GDP is evaluated at 43.70% which is higher than the corresponding amount of any region of the World except Latin America.

Crucial features of the Sub-Saharan Africa urbanization process have just been described. But then the question arises about how and why Africa cities display such characteristics. Tackling that issue imply grasping the driving forces of the Sub-Saharan Africa urbanization process.

2.3 Determinants of Sub-Saharan Africa urbanization

Several theses are advanced to explain urbanization and the role of primate cities in developing countries. The first points out landscape heterogeneity to explain agglomerations formation. According to this thesis ‘First Nature’, that is spatial inequality of locations in terms of exogenously given features of different sites, is the main explanation of the emergence of urban concentrations. The second put forward economic factors to explain the formation of cities. According to this thesis scale economies and the various other efficiency gains associated with clustering activities are the primary impulse behind the formation of urban concentrations. This thesis raises the claim that equilibrium city size is reached by a subtle balance between agglomerations economies and centrifugal forces triggered by congestion, pollution and transport costs. This is the idea developed by Lösch, quoted by Fujita and Thisse (2002) p.7, when he stated that:

> "*We shall consider market areas that are not the results of any kind of natural and political inequalities but arise through the interplay of purely economic forces, some working toward concentration, and others toward dispersion. In the first group are the advantages of specialization and of large-scale production; in the second, those of shipping costs and of diversified production.*"

But it is likely that some cities experience a development triggered by other factors than those involved in the first and second nature. In fact another popular thesis focuses on politics as the main drive of city growth. Supporting that theoretical strand, Ades and Glaeser (1995) even made the radical claim that:

> "*Political factors even more than economic forces drive urban centralisation...*"

That school of thought relies on 'rent seeking' as its crucial theme: city dwellers are more influent in the political system, this enabling them to extract rents. Those rents increase economic opportunities in cities which induces further in-migration from the countryside.

All these theses are obviously not mutually exclusive; their underlying agglomeration forces operate to a different extent in different countries and cities. From an economic theory standpoint, they correspond to different stages of the perception and the integration of space in economic formalization. For a long time, space has been neglected by mainstream economics. As argued by Krugman this is probably because during a long period economics lacked a model embracing both increasing returns and imperfect competition, the two basic ingredients of the formation of the economics landscape. Indeed, according to Starrett's Spatial Impossibility Theorem, if space is homogeneous any competitive equilibrium in the presence of transport costs will feature only fully autarchic locations (Fujita and Thisse, 2002). Therefore, the only way to explain economically the emergence of economic agglomerations is to relax either the homogeneity hypothesis or the perfect competition assumption.

2.3.1 'First Nature' and agglomerations

Relaxing the spatial homogeneity hypothesis has been the first way to explain agglomerations formation and regional specialization. Spatial heterogeneity may imply either the unequal spatial repartition of natural resources, technologies or amenities or the a priori existence of hubs of transportation networks (ports, transshipping points).

Inequalities with respect to technologies is the main impulse of spatial differentiation in standard international trade theory. According to the Ricardian model some countries are supposed to have a comparative advantage in terms of production costs. Therefore, each area specialize in the production of the good for which its relative opportunity costs are the lowest. Conversely, to the previous model, no technological gap is allowed in the Heckscher and Ohlin neoclassical international trade model. In this setting spatial heterogeneity consists merely in unequal endowments in production factors and each country specializes in production of goods using intensively the most abundant local production factors (Combes *et al.*, 2006). Thus, according to trade theories, each region is supposed to have comparative advantages in the production of a certain set of goods because of differences in technological conditions (e.g., climate or soil) or immobility of some production factors (e.g., mineral deposits, or labor and capital confined within borders).

Trade theories have also served as modeling frameworks for hubs of transport networks. Heterogeneity in terms of accessibility is another major cause of spatial differentiation pointed out by the literature. In many countries, dominant cities have developed mostly at ports. This statement that we already made for Africa, holds for Asia as well (Fujita and Mori, 1996). While the preference of ports for cities development seems trivial for geographers as they represent the most convenient place for imports and exports, the economics modeling of port cities has remained as an unsettled question for a long time. The traditional modeling of port cities by location theorists and urban economists (cfr. Mills, 1972, Chapter 5; Schweizer and Varaiya, 1976; Schweizer and Varaiya, 1977) represents the geographical counterpart of the standard trade theory based on comparative advantages. In this context, the production activity of each region is organized surrounding high-quality ports for the convenience of exports of final products and imports of other regions' goods, which leads to the formation of port cities (Fujita and Mori, 1996).

'First Nature' is clearly appealing to grasp the location of heavy industries during the Industrial Revolution, because at that time the proximity of raw materials was a critical factor. It is pertinent to explain Florida attractiveness as well. It also explains why port cities throughout the World and especially in Sub-Saharan Africa became the preferred location for industrial development.

However, it falls short of providing a reasonable explanation of many other clusters of activities, which are less dependent on natural advantages (good examples are the metropolitan area of Tokyo or the Silicon Valley) (Fujita and Thisse, 2002). This argument emphasizes the limitation of the 'neoclassical port-city model'. Many port cities, especially those in developed countries, have continued to prosper despite that their initial advantage in terms of cheap water access has weakened long before. If they were 'neoclassical port cities', then they should have disappeared a long time ago with the decrepitude of their original advantage. Clearly, their continued prosperity indicates the prevalence of the 'lock-in effect' of some self-reinforcing agglomeration forces (Fujita and Mori, 1996).[6]

Craft and Mulatu (2005) perform an empirical test to check out for the relative merits of HO factor endowments and increasing returns during the Industrial Revolution in Great Britain. While they find out that the larger part of the story may rely on factor endowments, they also allege that the HO story may imply NEG overtones especially while the original advantages of the location lost their relevance.[7]

Moreover, while space is heterogeneous, it is futile to explain the marked inequality of development only on the basis of space being naturally heterogeneous. Indeed, as argued convincingly by Duranton and Puga (2004) the land on which Chicago has been built, for instance, is not all that different from other locations on the shore of Lake Michigan that have been more sparsely developed.

2.3.2 'Second Nature' and agglomerations

Because of the inability of first nature to provide a convincing explanation to some clusters of economic activities, a great deal of effort has been made to uncover, after controlling for first nature, the economic forces that induce agglomeration. Grasping this 'second nature' is the main objective of geographical economics. It begins with considering an initial situation in which space is homogeneous and production activities are equally distributed at all sites. Second nature arguments relax therefore the assumption of perfect competition. Indeed, once we abstract from the heterogeneity of the underlying space, without indivisibilities or increasing returns and in presence of transport costs only fully autarchic locations prevail

[6]As Fujita and Mori (1996) pertinently point out in the U.S. all the ten largest cities in 1920 (i.e., New York, Chicago, Philadelphia, Detroit, Cleveland, St. Louis, Boston, Baltimore, Los Angeles, and Buffalo) were developed as port cities, and most of them remain to be great cities today even though (water-based) ports have little importance for their leading economic activities today. The limitation of the 'neoclassical port-city model' has induced Fujita and Mori (1996) to propose an alternative modeling of port cities, which "represents a counterpart of the 'new trade theory' based on increasing returns".

[7]Indeed, Craft and Mulatu (2005, 512) indicates that the correct interpretation of the factor endowments variables may not always be straightforward. They showed in table 7 that the interaction *coal abundance*steam power use* is significant but *coal abundance*coal use* is not. The historical literature yields a possible explanation for this: proximity to coal per se mattered for the initial location of steam-powered industry but, as time passed, it was external economies of scale that sustained the attractiveness of the region.

and no agglomerations would emerge.

Scale economies are therefore a decisive argument explaining the formation of urban agglomerations. According to Marshall, they have three kinds of micro foundations. Agglomerations economies can arise from the diversity of intermediate goods, from the matching process on the labor market, and from knowledge spillovers. The first two micro foundations lie at the origin of the division of labor and provide a clear foundation for increasing returns in the aggregate.

Specifically increasing returns may arise in the final goods sector when the intermediate goods is described by a monopolistic competition model. Scale economies may further arise in a setting in which firms have fixed requirements for limited productive resources and the market structure is characterized by monopolistic competition. Models of geographical economics display either of those features. They exhibit monopolistic competition as a crucial characteristic. In most of those models several effects come into play to determine the spatial pattern of economic activity. One of them is the 'market access' effect which describes the tendency of monopolistic firms to locate their production in the big market and export to small markets. This 'market access' effect, or backward or demand linkage, is fairly intuitive: for any manufacturing (or service) activity facing transport costs on the goods that it sells, the most profitable location is, everything staying equal, to be close to the location with the largest mass of consumers.

Another effect is the 'cost-of-living' one which concerns the impact of firms' location on the local cost of living. Indeed, goods tend to be cheaper in the location with more industrial firms since consumers in that region will import a narrower range of products and therefore avoid more of the transportation costs. One more agglomeration force is generated by vertical linkages between upstream and downstream industries (Venables, 1996). Firms want good access to intermediate inputs of other firms that may be necessary for their own production.

The combination of those backward and forward linkages may induce a self reinforcing agglomeration process with a potential for 'circular or cumulative' causation. Manufacturers choose to produce in the largest city because of the concentration of demand and inputs, but there is a concentration of demand and inputs in that city in large part precisely because so many producers have chosen that site (Krugman and Elizondo, 1996). We may notice such a phenomenon in dense networks of firms in related industries engineering, electronics and even financial services, where firms that supply specialist financial skills locate near the big financial institutions, and these institutions benefit from access to the skills of the specialists (Overman and Venables, 2005). Yet, those agglomeration effects can be counterbalanced by a third force: the 'market crowding' effect which reflects the fact that, everything staying equal, imperfectly competitive firms may tend to locate in regions with few competitors.

The matching process on the labor market is another source of increasing returns. Imperfect competition in thick labor markets, such as those encountered in big cities allows for a reduction in average matching costs. Indeed, large pools of specialist workers and of the firms that use these skills benefit from better matching of skills with requirements, and also from risk sharing if there are firm or worker specific fluctuations in demand or supply. Incentives to acquire skills are greater if the skills are sought by several firms, so the worker is less likely to be subject to the monopsony power of a single employer. Knowledge spillovers are one more micro foundation of agglomeration economies. Productivity spillovers between firms may occur, as they are able to learn about and imitate the practices of other firms in the industry. Silicon valley is a concrete example where knowledge exchange, formal and informal, is quite widespread (Overman and Venables, 2005).

The final argument has to do with the provision of public goods. Provision of such goods is obviously subject to increasing returns. Taking advantage of those scales economies has been the main impetus of policies biased in favor of cities. Such an 'urban bias' in public expenditure and public goods provision has been advocated to induce an efficient allocation of resources (Arnott and Gersovitz, 1986).

What about the evidence of agglomerations economies in Sub-Saharan Africa? While evidence of agglomerations economies in developed nations is fairly well documented, this is not really the case for developing countries. Literature provide evidence of productivity effects of agglomerations in Asia (see Au and Henderson (2006) for China) and Latin America (see Overman and Venables (2005) for further references) but hardly for Africa.

As table 2.7 shows income levels in cities are higher than the country average level.[8] This observation holds for Africa as well, where cities income levels are 65% higher than average. This may indicate that increasing returns to scale and a higher productivity are at play in Africa cities.

Yet, some literature characterizes Sub-Saharan as less favored with respects to agglomeration economies. It seems that african manufacturing industries fail to realize agglomeration economies (Becker *et al.*, 1994). According to Collier (2006), Black Africa hasn't benefited from the stream of liberalization and the widening gap of labor costs that render relocation in low-income countries profitable to several industries. The chosen locations

[8] In this table HIC stands for Highly Industrialized Countries and LAC stands for the Latin America and the Caribbean Region.

Table 2.7: GDP and Cities Income Levels per capita

Region	GDP per Capita (in US $)	City Product (in US $)	Household Income (in US $)	Informal employment (in %)	Unemployment rate (in %)
Africa	441	729	1,637	54	23
Arab States	2,752	3,170	5,850	65	11
Asia Pacific	4,742	6,182	9,101	33	8
HIC	22,501	22,103	26,273	3	8
LAC	3,350	3,226	5,623	39	13
Transitional	2,541	2,905	3,591	21	9

Source: UN-HABITAT, 2001
(`http://ww2.unhabitat.org/programmes/guo/guo_analysis.asp`).

for relocation were mostly in Asia and not in Africa. While the factors that determined this relocation may have been only temporary, once Asia got ahead cumulative causation of agglomeration made it harder for Sub-Saharan Africa to catch up. Nowadays Africa has no substantial advantage over Asia in terms of labor costs while having large disadvantages in terms of agglomerations economies. African industry has a much lower total factor productivity and a much higher share of indirect costs than its counterparts in low income Asian countries (Kessides, 2005). The alleged failure of Sub-Saharan Africa industries to generate agglomeration economies gives some backing to the last thesis explaining cities development mostly by their ability to draw resources from the countryside.

2.3.3 Rent seeking

While the previous argument builds on real efficiency gains from the scale effects of urban agglomerations, an adverse claim states that urban-rural differentials arise merely from transfer payments: urban dwellers benefit not by enhancing productivity, but rather by extracting resources from the rest of society. Several arguments supporting this unproductive perception of cities have been developed, including Lipton's view of 'urban bias' (Lipton, 1983), Ades and Glaeser ' description of 'Romes without empires' (Ades and Glaeser, 1995) and Hoselitz's concept of 'parasitic city' (Hoselitz, 1955).

The main mechanism here is political influence. In several developing countries, many business formalities like starting a firm, hiring and firing workers, registering property, enforcing contracts, accessing capital and export/import markets, obtaining production rights, protecting investors and closing a business are subject to extensive regulation (World Bank, 2005 quoted by Overman and Venables, 2005; Henderson and Kuncoro, 1996). The granting of the necessary permits and licenses may be biased by some favoritism at the benefit of entrepreneurs living in the national capital, allowing central bureaucrats and politicians to extract rents (Henderson, 2003).

Moreover, the political power of the urban elites may induce the government to raise their real incomes at the expense of people living in the countryside (Ades and Glaeser, 1995). It may also allow city dwellers to benefit from 'biased' public expenditures programmes ranging from health through to transport (Overman and Venables, 2005). This further attracts workers to the city.

As explanations based on First nature, rent seeking arguments implies spatial heterogeneity. There is however a crucial difference between the two claims in the signification of regional heterogeneity: for first nature arguments heterogeneity of sites is exogenous and refer to the unequal distribution of resources or to differences in terms of accessibility while, for the rent seeking claim, it refers to asymmetry of the political power of cities dwellers compared to inhabitants living in the countryside and it is merely endogenous.[9]

Is there clear evidence that rent seeking prevails in explaining city growth in Sub-Saharan Africa? There is a lack of in depth studies on this issue due to carence of data. However, casual inspection of available evidence gives some credit to the rent-seeking story. Kinshasa for instance has been depicted as a centre of richnesses accumulation of the Congolese predatory state. The share of investments and public expenditures in the Congolese capital appears as disproportionate comparatively to its demographic weight. While Kinshasa gathered 9 per cent of the country's total population and 31 per cent of its urban population in 1984, it accounted for 42 per cent of the building sector, nearly 40 per cent of trade, one-third of direct taxes, and one-fifth of public services (in 1970). In 1975 it consumed 72 per cent of low-voltage power and accounted for 47 per cent of water sales. However, the declining ability of congolese state to generate and redistribute resources and the deliquescence of its administrative structures leads to question about the future of a megalopolis which had always based its growth on the power of the State (Piermay, 1997).

[9] In Chapter 4 we assume that the uneven distribution of political power is exogenous. But this is clearly an analytical simplification. A full model should yield microfoundations to the political elite clustering in capitals, which then induces rent seeking. We thank Kristian Behrens for this clarification.

2.4 Conclusion and research perspectives

Sub-Saharan Africa's urban pattern is characterized by urban bias and urban primacy. The skewness of spatial distribution in Sub-Saharan Africa seems to be induced at least partly by hubs and political effects. Those determinants outline spatial heterogeneity in terms of either accessibility or political power. Agglomerations economies arising through the interplay of increasing returns to scale, while constituting a major urbanization drive, appear to be less acknowledged in Sub-Saharan Africa.

An interesting research agenda emerges from this description of Sub-Saharan Africa urbanization's stylized facts and determinants. As urban centralization is a crucial feature of Sub-Saharan Africa 's urban pattern, it is interesting to grasp rigorously why it occurs. To the best of our knowledge, up to now, NEG literature has not proposed satisfactory analytical frameworks explaining the impact of the aforementioned determinants on Sub-Saharan Africa urbanization. Therefore, models explaining the effects of these drivers on urban concentration are proposed on the fourth and the fifth chapters. This modeling assumes that agglomerations economies as well as hubs and political effects are pertinent. However, some crucial issues emerge in first place from the previous enumeration of the determinants of Sub-Saharan Africa urbanization: does rent seeking outplays agglomeration economies in explaining urban concentration in Sub-Saharan Africa? Does urban concentration foster economic growth in that region?

Those issues are quite controversial. Indeed, opposite claims have been raised with respect to Developing countries urbanization. The first states that developing countries are over urbanized with respect to their level of economic development and therefore gives more credit to the rent seeking story. Todaro and Bairoch reach this conclusion from different analytical frameworks. Bairoch claims that excessive urban concentration is due to, among other factors, rapid population growth which leads to rural crowding and stimulates rural to urban migration. In addition, he argues that artificially high urban wages pull a disproportionate part of the population to urban areas. While making a similar claim concerning urban wages, Todaro's analysis differs by its greater emphasis on economically inefficient migration caused by legally and socially determined minimum wage rates and migrants expectations (Todaro, 1969; Bairoch, 1988).

Conversely, to Todaro and Bairoch, Williamson (1987) asserts that there is no consistent evidence that developing countries are over urbanized, and that urbanization has outpaced industrialization in developing countries. According to him Third World city growth seems to obey to the classical rules of logistic curves tracing the diffusion of new technologies with urban population growth rates first sharpening and then slowing down. Williamson justifies the higher levels of those growth rates in Sub-Saharan Africa by the fact that overall rates of population growth were higher in that region both in cities and in the countryside.

Those two thesis differ on the crucial theme of the qualitative nature of the relationship between urban concentration and economic growth in developing countries. We tackle this issue more deeply in the next chapter where we investigate the relationship between urban concentration and economic growth and we check out whether there is any difference in the qualitative nature of this relationship across regions.

Bibliography

[1] Aase, A. (2003) *Urbanisering og byliv i Africa*. Trondheim, Mimeo. quoted by Wikipedia `http://en.wikipedia.org/wiki/Urbanization_in_Africa`.

[2] Ades, A. and E.L. Glaeser (1995) Trade and circuses: Explaining urban giants, *Quarterly Journal of Economics*, 110, 195-227.

[3] Arnott, R.J. and M. Gersovitz, (1986) Social Welfare Underpinnings of Urban Bias and Unemployment, *Economic Journal*, 96, 413-424.

[4] Aryeetey-Attoh, S. (1997) *The geography of Sub-Saharan Africa*. Upper Saddle River, Prentice-Hall.

[5] Au, C. and V. Henderson, (2006) Are Chinese Cities Too Small? *Review of Economic Studies*, 73, 549-576.

[6] Bairoch, P. (1985) *De Jericho à Mexico: Villes et économie dans l'histoire*. Paris: Gallimard. English translation *Cities and Economic Development. From the Dawn of History to the Present* (1988), London: Mansell Publishing Ltd.

[7] Becker, C.M., A.M., Hamer and A.R., Morisson (1994) *Beyond Urban Bias in Africa. Urbanization in an era of structural adjustment*. London: James Currey, Portsmouth: Heinemann.

[8] Christopher, A.J. (1994) Urbanization and National Capitals. In: Tarver, J.D. (ed.) *Urbanization in Africa: a handbook*. Westport: Greenwood Press , pp. 408-422.

[9] Collier, P. (2006) *Africa: Geography and Growth*. Mimeo.

[10] Combes, P.-P., T., Mayer and J.-F., Thisse (2006) *Économie géographique. L'intégration des régions et des nations*. Paris: Economica.

[11] Craft, N. and A. Mulatu (2005) What explains the location of industry in Britain, 1871-1931? *Journal of Economic Geography* 5, 499-518.

[12] Duranton, G. and D. Puga (2004), Micro-foundations of urban agglomerations economies. In: Henderson, J.V. and J.F. Thisse (eds.) *Handbook of Regional and Urban Economics, vol. 4*, North-Holland: Elsevier Science Publishers, pp. 2063-2117.

[13] Fujita, M. and T., Mori (1996) The role of ports in the making of major cities: Self-agglomeration and hub-effect, *Journal of Development Economics* 49, 93-120.

[14] Fujita, M. and J.-F., Thisse (2002) *Economics of Agglomeration. Cities, Industrial Location, and Regional Growth*. Cambridge: Cambridge University Press.

[15] Henderson, J.V. (2003) The Urbanization Process and Economic Growth:The So-What Question, *Journal of Economic Growth* 8, 47-71.

[16] Henderson, J.V. and A., Kuncoro (1996) Industrial Centralization in Indonesia, *The World Bank Economic Review* 10, 513-540.

[17] Hoselitz, B. F. (1955) Generative and Parasitic Cities, *Economic Development and Cultural Change* 3, 278-294.

[18] Kempe, R. H. (1996) *African Political Economy*. Cambridge, MA: M.E. Sharpe.

[19] Kessides, C. (2005) The Urban Transition in Sub-Saharan Africa: Implications for Economic Growth and Poverty Reduction, Africa Region Working Paper Series #97.

[20] Krugman, P. and R.L., Elizondo (1996) Trade policy and the Third World metropolis, *Journal of Development Economics* 49, 137-150.

[21] Lipton, M. (1983) *Why poor people stay poor: a study of urban bias in world development*. London: Temple Smith.

[22] Mabogunje, A.L. (1994) Introduction: Cities and Africa's Economic Recovery. In: Tarver, J.D. (ed.) *Urbanization in Africa: a handbook*. Westport: Greenwood Press , pp. xxi - xxxii.

[23] Max Miller, H. and R.N. Singh (1994) Urbanization during the Postcolonial Days. In: Tarver, J.D. (ed.) *Urbanization in Africa: a handbook*. Westport: Greenwood Press , pp. 408-422.

[24] Mills, E.S. (1972) *Studies in the structure of the urban economy* . Baltimore: Hopkins Press.

[25] Montgomery, M.R., R. Stren, B. Cohen, and H.E. Reed (2004) *Cities transformed. Demographic Change and Its Implications in the Developing World. Panel on Urban Population Dynamics*. Washington, D.C.: The National Academies Press.

[26] North, D.C. (1990) *Institutions, Institutional Change and Economic Performance*. Cambridge: Cambridge University Press.

[27] Overman, H.G. and A.J., Venables (2005) *Cities in the developing world*. Mimeo.

[28] Piermay, J.L. (1997) Kinshasa: A reprieved mega-city? In: Rakodi, C. (ed.) *The urban challenge in Africa: Growth and management of its large cities*. Tokyo - New York - Paris: United Nations University Press available online at `http://www.unu.edu/unupress/unupbooks/uu26ue/uu26ue00.htm`.

[29] Schneider, F. (2004) The size of the shadow economies of 145 countries all over the world: First results over the period 1999-2003, IZA Discussion Paper #1431.

[30] Schweizer, U. and P. Varaiya (1976) The spatial structure of production with a Leontief technology, *Regional Science and Urban Economics* 6, 231-251.

[31] Schweizer, U. and P. Varaiya (1977) The spatial structure of production with a Leontief technology-II: Substitute techniques, *Regional Science and Urban Economics* 7, 293-320.

[32] Todaro, M. P. (1969) A Model of Labor Migration and Urban Unemployment in Less Developped Countries, *American Economic Review* 59,138-148.

[33] Todaro, M. P. (1997) Urbanization, Unemployment and Migration in Africa: Theory and Policy, Policy Research Division Working Paper #104.

[34] United Nations (2002) *World Urbanization Prospects: The 2001 Revision. Data Tables and Highlights*. New York: United Nations Secretariat, Population Division.

[35] United Nations (2006) *World Urbanization Prospects: The 2005 Revision. Executive Summary, Fact Sheets, Data Tables*. New York: Department of Economic and Social Affairs, Population Division.

[36] Venables, A.J. (1996) Equilibrium Locations of Vertically Linked Industries, *International Economic Review* 37, 341-359.

[37] Williamson, J. G. (1987) Migration and Urbanization. In: Chenery, H. and T.N. Srinivasan (eds.) *Handbook of Development Economics, vol. 1*, North-Holland: Elsevier Science Publishers, pp. 424-465.

Chapter 3

Urban Concentration and Economic Growth: checking for specific regional effects

3.1 Introduction

Urbanization appears as closely linked to economic development. In any year the simple correlation coefficient between the urbanization rate and the log of per capita GDP is about 0.85 (Henderson, 2003). The intuition behind this stylized fact is clear. As economies develop, relative and absolute changes in demand increase the relative and absolute importance of the industrial and service sectors. These sectors are much less land intensive than the agricultural sector, and they allow easier substitution of non land for land inputs (Moomaw and Shatter, 1996). Therefore, in spite of urban land's high prices, firms of those sectors can cluster in urban areas to take advantage of Marshall's localized economies of scale (Henderson, 1974; Duranton and Puga, 2001; Henderson, 2003).

However, while urbanization is a universal phenomenon triggered by the sectorial shift from agriculture to industry and modern services, its speed seems to vary according to the level of development. The differential rates of growth between urban and rural population in 1950-80 show an inverted U pattern with middle-income countries having the highest rates while those of developed nations and low income countries are the lowest (Mazumdar, 1987). As middle-income countries are, on average, those that benefit from the highest growth rates and face the most drastic changes in their economic structures, this evidence would suggest that urbanization is related to economic development and industrialization (Yuki, 2007).

Yet, since several nations that urbanized the most rapidly, i.e. African nations and Latin American & Caribbean, grew relatively slowly (Mazumdar, 1987; Fay and Opal, 2000) the speed of urbanization seems not to be explained solely by economic growth (Yuki, 2007). This observation holds for urban concentration as well. This is the aspect of urban development that seems to interest economists the most (Henderson, 2003). This is also the one that has triggered the greatest deal of concern and controversy. In 2005, 15 of the 20 urban agglomerations of more than 10 millions inhabitants were located in developing countries. The surge of so many megalopolises in developing countries has been a subject of concern for international policy officials. For a long time international development agencies have suspected megacities of developing countries to be over populated and have considered their alleged 'overconcentration' as detrimental for economic growth.

Such a prejudice may be partly grounded on the analysis made by Todaro and Bairoch. From different analytical frameworks they arrived both at the same conclusion that there is 'excessive' urban concentration in developing countries. Bairoch argues that excessive urban concentration is due to, among other factors, rapid population growth which leads to rural crowding and stimulates rural to urban migration. In addition, he claims that artificially high urban wages pull a disproportionate part of the population to urban areas. While making a similar claim concerning urban wages, Todaro's analysis differs by its greater emphasis on economically inefficient migration caused by legally and socially determined minimum wage rates and migrants expectations (Moomaw and Shatter, 1996; Todaro, 1969; Bairoch, 1988).

There is however no unanimity regarding that issue. Challenging Todaro and Bairoch' claim, Williamson (1987) asserted that there is no evidence confirming that developing countries are overurbanized, and that urbanization has outpaced industrialization in developing countries. Mera (1973) claimed that the largest metropolitan areas in the world are likely to be less large than the optimum in terms of economic efficiency. Conversely, Ades and Glaeser (1995) found that both population share of the largest city and urbanization outside the main city have negative and significant effects on growth of GDP per capita, reaching the opposite conclusion that 'Large cities generate rent-seeking and instability, not long term economic growth'.

Economic literature firstly formalized the link between urban concentration and economic growth by the Williamson hypothesis. It states that economic development first increases and then decreases spatial concentration within a country, thus exhibiting a bell shaped relationship (Junius, 1999; Williamson, 1965; Alonso, 1980). At early stages of economic development, a country optimizes the use of its physical infrastructure and managerial resources by clustering them in primate and often coastal cities. Such spatial clustering favors information spillovers and knowledge accumulation when the economy is 'information deficient'. Nevertheless, at later stages of development process, deconcentration proceeds for the mere reason that the economy can sustain the spread of economic infrastructure and knowledge resources in the hinterland and because primate cities have become congested areas that are less efficient for economic agents (Henderson, 2003).

The bell shaped relationship has been confirmed by some empirical studies (El-Shakhs, 1972; Alonso, 1980; Wheaton and Shishido, 1981; Junius, 1999; Davis and Henderson, 2003). But it has also been contradicted by others. Richardson and Schwartz (1988) find no support of any link between primacy and economic growth. As Ades and Glaeser (1995), Mutlu (1989) and Moomaw and Shatter (1996) find a negative relationship between urban concentration and economic development. So alternative explanations focusing on non economic factors have been raised. One of them states that cities grow in a parallel way and that spatial concentration is unaffected by urbanization and economic development (Junius, 1999; Black and Henderson, 1999; Eaton and Eckstein, 1997), the distribution of urban population reflecting simply geography or historic shocks. Another hypothesis, supported by a large strand of the literature, outlines the importance of political institutions and policies in spatial concentration. Ades and Glaeser (1995) even asserts that 'political forces, even more than economic forces, drive urban centralization'. For political reasons a government may favor one or more cities over others, especially national capitals. Such a favoritism may take several forms: the government may underinvest in interregional transport and telecommunications favoring therefore consumers and producers in the national capital over those in the hinterland (Fujita *et al.*, 1999); it may impose restrictions in the capital and the export/import markets favoring firms located in the capital; finally it may allocate public services preferentially in the national capital.

One reason of the contradictions just outlined may be that countries are likely to show up a great deal of heterogeneity with respect to urbanization and growth patterns. Urbanization qualitative nature appears to vary across countries with on one hand countries experiencing urbanization accompanied by skill upgrading, industrialization, economic growth and the expansion of the urban formal sector and on the other nations experiencing simultaneously an urbanization without modernization , the expansion of the shadow economy and economic stagnation (Yuki, 2007). In the growth empirical literature, the objection has been raised that very different countries are unlikely to be drawn from a common surface as multiple regression assumes and evidence of widespread parameter heterogeneity has been provided (Temple, 1999).

The evidence that a substantially different pattern of urbanization prevails in Sub-Saharan Africa would then raise the interest of the quest for a way of modeling agglomeration economies more suitable for that part of the world. But does such an evidence exist? *Do developing and developed countries diverge with respect to urbanization patterns?* In order to obtain such an evidence, we analyze the relationship between economic growth and urban concentration. Therefore, our analysis is similar to Henderson's (2003). However, Henderson (2003) assumed that there exists an optimal level of urban concentration, and showed that deviation form that optimum may be very costly. Here, we relax the assumption that the economic growth rate is a concave function of urban concentration. Therefore, as Bertinelli and Strobl (2003), we model economic growth using a semi-parametric function, with the nonparametric term depending on urban concentration. However, contrary to them, we use a Differencing method to perform semi-parametric estimation. This method has the advantage of allowing us to check out subsequently if urbanization patterns differ across group of countries by performing tests of equality of non-parametric functions of different subsamples. Furthermore, our analysis differs from the one of Bertinelli and Strobl (2003) by the fact that we address explicitly the endogeneity problem.

Another major difference with the earlier literature is the way we handle the data missingness problem. Most of empirical papers in economics handle missing data by listwise deletion i.e. by deleting any observation having at least one missing datum. This approach has faced a lot of criticism by the statistics literature. Indeed, apart from the fact that it implies the loss of all the information conveyed in the observations having missing data, the estimates obtained with such a method have been proved to be biased if the data remaining after deletion is not a random sample of the overall database.[1] In order to avoid such a shortcoming, we implement a two-step multiple imputation algorithm which is convenient if data are missing at random (MAR).

The remainder of this chapter is organized as follows. Section 2 is devoted to the description of the estimation methodology. In this section we justify our specification choice, we present our basic estimation strategy - the Bayesian Semiparametric estimation - and tests of equality of nonparametric regression functions. Then we discuss endogeneity issues and methods for handling data missingness. Finally we describe our Estimation-Imputation algorithm. Section 3 presents the results obtained and Section 4 concludes.

[1]If the converse assertion were true data would be described as missing completely at random ($MCAR$).

3.2 Estimation Methodology

3.2.1 Specification

The Empirical Growth literature has for some time been dominated by papers with cross-country growth regressions. The formulation and the relevance of such regressions have been quickly subject to a rising skepticism. This approach was indeed prone to several shortcomings. The most important of them is that as cross-sectional regressions fail to control for individual heterogeneity, they face an omitted variables problem and thus yields biased estimates. The use of panel data allows to mitigate such an inconvenience. Indeed, in a panel data framework one may control for heterogeneity in the initial level of efficiency and thus ensure that coefficients will be unbiased. Secondly with panel data several lags of regressors may be used as instruments, alleviating therefore measurement errors and endogeneity biases (Temple, 1999; Magrini, 2004)

However, implementing traditional growth regression in a panel framework has some major drawbacks : rather than having exogenous technological change and population growth, they include determinants of population change leading away from the standard neo-classical framework. Moreover, estimation of growth models implies the additional complexities of dynamic panel data models (Henderson, 2003; Temple, 1999). So we estimate a Total Factor Productivity model. Total Factor Productivity models rely on a production function with two factors of production, physical capital K_{it} and labor L_{it}, and a variable reflecting the effects of technological progress A_{it}. Adopting the Cobb-Douglas specification, we get the following production function:

$$Y_i(t) = K_i(t)^{\alpha} A_i(t)(L_i(t))^{1-\alpha} \tag{3.1}$$

Linearizing, differencing and normalizing with respect to labor leads to the following specification:

$$ln\frac{Y_{i(t)}}{L_{i(t)}} - ln\frac{Y_{i(t-1)}}{L_{i(t-1)}} = \alpha[ln\frac{K_{i(t)}}{L_{i(t)}} - ln\frac{K_{i(t-1)}}{L_{i(t-1)}}] + ln\frac{A_{i(t)}}{A_{i(t-1)}} \tag{3.2}$$

Total factor productivity growth is modeled as a function of (i) education of the labor force, which captures the capacity of adopting new technologies (Grossman and Helpman, 1991; Durlauf and Quah, 1998),(ii) internal country considerations affecting efficiency and growth, like urban primacy (Henderson, 2003). We therefore assume the following functional form for productivity growth

$$ln(A_{i(t)}/A_{i(t-1)}) = z'_{i(t-1)}\delta + f(x_{i(t-1)}) \tag{3.3}$$

with x representing urban concentration, and z representing a row vector of control variables including: average years of high school and college as a proxy of education of the labor and time fixed effects.[2]

We consider the second term of the right hand side of equation (3.3) as a non-linear function of urban concentration. There are indeed good reasons to assume that it is not linear in urban concentration. Firstly there is the so-called Williamson effect according to which urban concentration should be high at first stages of development and then decrease as the economy develops. The Williamson effect implies therefore a bell shaped relationship between urban concentration and the level of economic development. Henderson (2003) has checked the validity of Williamson hypothesis by modeling $f(x)$ as a quadratic function of x. With such a function he was also able to verify that, as we may expect, urban concentration decreases with the population of a country and national geographic size and he has shown that there is an optimal urban concentration level from which departures could entail significant losses in terms of economic growth. Secondly as noticed by Bertinelli and Strobl (2003), urban concentration variables are bounded from above and below, and may even just affect growth differently near their bounds comparatively to mid-values.

While the intuition behind the Williamson effect is quite appealing, as stated previously there are compelling arguments suggesting other kinds of relationship. So we may wonder if the Williamson hypothesis is actually supported by the data. Thus, following Bertinelli and Strobl (2003) we will allow some flexibility in our specification by using a semiparametric specification, with the nonparametric term being a function of urban concentration.

3.2.2 Bayesian Semiparametric estimation

Because of the plurality of theoretical frameworks as well as diverging empirical results, we have no obvious functional form for the regression relationship between growth rates, urban concentration and other controls. Such a context gives backing to arguments of non-parametric econometricians who stress that the implications of economic theory is often non-parametric and propose semiparametric or non-parametric functional forms (Koop

[2]In this semi-parametric specification we do not control for countries fixed effects. Indeed, the differencing method we use for performing this estimation implies a reordering of the data matrix that undermines the implementation of efficient fixed effects estimation methods. Therefore, to estimate fixed effects we must add countries dummies to the matrix of regressors, which increases remarkably its size and the number of coefficients to estimate. Thus, we only control for fixed effects when we estimate the model without the non-parametric function as shown in table 3.2. This omission of fixed effects is likely to trigger endogeneity. We will present a solution to this problem when we will address endogeneity issues.

and Poirier, 2004; Yatchew, 1998). Therefore, we opt for working with the following semi-parametric specification:

$$\Delta ln\big(Y_{i(t)}/L_{i(t)}\big) = z'_{1i(t)}\beta + f(x_{i(t-1)}) + \epsilon_{i(t)}, \quad \epsilon_{i(t)} \sim \mathcal{N}(0,\sigma^2) \tag{3.4}$$

Where $z_{1i(t)} = \big(\Delta ln\big(K_{i(t)}/L_{i(t)}\big), z_{i(t-1)}\big)'$, and $\beta = (\alpha, \delta')'$.

However, nonparametric methods are not very popular in applied work for the mere reason that nonparametric regression techniques are theoretically more complex than the usual tool kit of linear and nonlinear parametric modeling methods, and that they are computationally intensive (Yatchew, 1998).

To avoid such shortcomings, we use a Bayesian Semiparametric model (Koop and Poirier, 2004). Based on the standard Normal linear regression model with natural conjugate prior for which standard analytical results are available, this model has the advantage to be computationally simple. Furthermore, by using such a method we avoid the criticism addressed to usual Bayesian methods that they incorporate prior information. Indeed, in the approach we are using, the only type of prior input required is one prior hyperparameter, η, which controls the degree of smoothness of $f(x)$.

Let's consider N as the number of cross-sectional units and T as the number of time periods for each of them. In Bayesian semiparametric estimation of (3.4) $f(x)$ plays the role of an intercept and observations have to be reordered so that $x_1 \leq \ldots \leq x_i \leq \ldots \leq x_{N(T-1)}$. Defining $y = \big(\Delta ln(Y_1/L_1), \ldots, \Delta ln\big(Y_{N(T-1)}/L_{N(T-1)}\big)\big)'$, $x = \big(x_1, x_2, \ldots, x_{N(T-1)}\big)'$, $Z = \big(z_{11}, \ldots, z_{1N(T-1)}\big)'$, and letting $\gamma = \big(f(x_1), \ldots, f(x_{N(T-1)})\big)'$, $W = (Z, I_{N(T-1)})$ and $\theta = (\beta', \gamma')$, we can rewrite (3.4) as:

$$y = W\theta + \epsilon, \tag{3.5}$$

In equation (3.5) there are more variables than observations. Therefore, additional information is needed to overcome the fact that $W'W$ is singular. Assuming that $f(x_{i(t-1)})$ is smooth we may use the following partially informative prior (Koop and Poirier, 2004):

$$D\gamma \sim \mathcal{N}_{N(T-1)-m}\left(0, \sigma^2 P_0^{-1}\right) \tag{3.6}$$

where $P_0^{-1} = diag\,(\eta\;\eta\;\ldots\;\eta)$ and D is a differencing matrix given by the following expression:

$$D_{(N(T-1)-m)\times N(T-1)} = \begin{bmatrix} d_0, d_1, d_2, ..d_m, 0, ..., 0 \\ 0, d_0, d_1, d_2, ..d_m, 0, ..., 0 \\ \vdots\ \vdots \\ 0, ...0, d_0, d_1, d_2, ..d_m, 0 \\ 0, ...0, d_0, d_1, d_2, ..d_m \\ 0, ..., 0 \\ \vdots\ \vdots \\ 0, ..., 0 \end{bmatrix} \tag{3.7}$$

with m the order of differencing, d_i with $0 \leq i \leq m$ a differencing weight. Combined with an Inverted Gamma prior on σ^2, the prior on $D\gamma$ is conjugate. It yields an integrable posterior. Indeed, the posterior precision $W'W + M_0$ is of full rank.[3] To estimate β we apply the differencing matrix to (3.5) which yields:

$$Dy = DZ\beta + Df(x) + D\epsilon \approx DZ\beta + D\epsilon \tag{3.8}$$

Application of OLS to (3.8) is straightforward.

Once $\hat{\beta}$ is obtained, we may obtain γ by the following extended regression:

$$\begin{pmatrix} y - Z\hat{\beta} \\ 0_{N(T-1)-m} \end{pmatrix} = \begin{pmatrix} I_{N(T-1)} \\ \left(P_0^{1/2}\right)' D \end{pmatrix} \gamma + \begin{pmatrix} \epsilon \\ \epsilon_0 \end{pmatrix} \quad with \quad \begin{pmatrix} \epsilon \\ \epsilon_0 \end{pmatrix} \sim \mathcal{N}\left(0, \sigma^2 I_{2N(T-1)-m}\right) \tag{3.9}$$

3.2.3 Tests of equality of regression functions

With such a nonparametric estimation framework we may perform tests of equality of nonparametric functions across subsamples. Several procedures to realize such tests have been designed. Most of them involve direct comparison of nonparametric estimates of regression curves or analysis of residuals from such regressions. Yatchew (1999) eases the test by proposing a procedure avoiding the computation of nonparametric regressions. Suppose that we have G subsamples of different sizes respectively N_1, N_2,..., N_G with $\sum_{i=1}^{G} N_i = N(T-1)$. Let's apply regression model (3.5) to each subsample separately, then we have for any subsample i we have:

$$y_i = Z_i\beta + f(x_i) + \epsilon_i, \tag{3.10}$$

[3]the prior precision M_0 being given by $M_0 = \begin{pmatrix} 0 & 0 \\ 0 & P_0 \end{pmatrix}$

where $y_i = (y_{i1} \dots y_{iN_i})'$, $Z_i = (z_{11}, \dots, z_{1N_i})'$, $x_i = (x_{i1}, \dots, x_{iN_i})'$, $\epsilon_i = (\epsilon_{i1}, \dots, \epsilon_{iN_i})'$ with $\epsilon_i \sim \mathcal{N}(0, \sigma^2 I_{N_i})$ and $i = 1, \dots, G$.

Supposing that data have been reordered so that within each subsample, the x' s are in increasing order and defining $y = (y_1', \dots, y_G')'$, $Z = (Z_1', \dots, Z_G')'$ and $y_{npar} = y - Z\beta$ then we may compute the 'within' estimator of σ^2 as:

$$s_w^2 = \frac{1}{N(T-1)} y_{npar}' D_{test}' D_{test} y_{npar}, \tag{3.11}$$

with D_{test} defined to be the following block diagonal matrix

$$D_{test} = \begin{bmatrix} D_1 & 0 & \dots & 0 \\ 0 & D_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & D_G \end{bmatrix} \tag{3.12}$$

with each bloc D_i of dimension $N_i \times N_i$ having the same structure as (3.7).

Let's define P_p as the 'pooled' permutation matrix that reorders the overall dataset so that the 'pooled' x are in increasing order. Then we define the 'pooled' variance estimator as:

$$s_P^2 = \frac{1}{N(T-1)} y_{npar}' P_p' D_{test}' D_{test} P_p y_{npar}, \tag{3.13}$$

Under the null hypothesis that all nonparametric regression functions are identical we may define the following test statistic

$$\Upsilon = (mN(T-1))^{1/2} \left(s_P^2 - s_w^2\right) = \frac{m^{1/2}}{(N(T-1))^{1/2}} y_{npar}' Q_\Upsilon y_{npar} \xrightarrow{D} \mathcal{N}\left(0, 2\pi_\Upsilon \sigma^4\right), \tag{3.14}$$

with $Q_\Upsilon = P_p' D_{test}' D_{test} P_p - D_{test}' D_{test}$ and $\hat{\pi}_\Upsilon = m\, tr(Q_\Upsilon Q_\Upsilon)/NT$ and supposing that $\hat{\pi}_\Upsilon \xrightarrow{P} \pi_\Upsilon > 0$.

Therefore, $\Upsilon / s_w^2 (2\, \hat{\pi}_\Upsilon)^{1/2} \xrightarrow{D} \mathcal{N}(0,1)$ and we would reject the null hypothesis for large positive values of the test statistic. Rejection of the null hypothesis suggests that the pattern of urbanization is different between groups of countries.

3.2.4 Data

Several variables are involved in estimation of equation (3.5) namely education of the labor force, urban concentration, capital and output growth rates. While measuring education is quite straightforward with average years of high school and college of population of at least 25 years old standing as a convincing proxy, things are more involved with respect to urban concentration. There have always been several measures of spatial concentration so that deciding what measure to opt for is an issue. The Hirschman-Herfindalh index and the Pareto parameter have been the first to be used. But there are available only for few years for a limited sample of countries. Therefore, they don't fit for panel data.

Urban primacy defined as the share of urban population living in the largest city, is conversely available over years for more countries. It is moreover closely correlated to the previous measures (Henderson, 2003) and thus has been used in many studies. But as explained in Bertinelli and Strobl (2003), this measure seems unsatisfactory when there are huge differences between country sizes. Indeed, small countries tend to gather the quasi-totality of their urban population in a single city. Moreover, there are often cities other than the largest city that account for large proportions of the urban population. Therefore, using urban primacy as a measure of urban concentration results in attributing low values to countries like India which has many large cities, but very large values for small countries. Furthermore, changes in urban primacy sometimes don't reflect changes in the total population. It has been noticed that while the share of the largest city often decreased as a consequence of increasing urbanization, urban concentration increased due to a more than proportional increase in medium and large agglomerations. All those shortcomings have induced Bertinelli and Strobl (2003) to adopt another measure of urban concentration: urban density defined as the share of the urban population living in cities larger than 750,000 inhabitants. For reason of completeness we will adopt both measures in this study. This will allow us to check whether the results are robust to changes in measures.

Estimation of a TFP model requires data on the capital stock. We use Dareshwar and Nehru (1993) data on the capital per capita along with their output per worker measure. Those measures are based on perpetual inventory methods and are in local currency units. To take into account variations of purchasing power across countries, those results were converted in PPP at 1987 exchange rate.

3.2.5 Endogeneity Issues

The use of Dareshwar and Nehru data raises endogeneity issues. Indeed, although their measurements are carefully done, there are likely to suffer from measurements errors.

Endogeneity issues are further compounded by the fact that in equation (3.4) contemporaneous shocks $\epsilon_{i(t)}$ potentially affect covariates at period t and even at period $t-1$. Indeed, perspectives of shocks in economic growth are likely to induce migration to the largest city, increasing therefore urban concentration (Henderson, 2003). Finally this endogeneity problem is also triggered by the fact that equation (3.4) implies a pooled estimation where individual heterogeneity is not controlled for. Since fixed effects are likely to be correlated to regressors this fuels the endogeneity bias.

Therefore, regressors are not strictly exogenous. In order to be able to identify the 'causal' effect of urban concentration on economic growth rather than simple correlations we have to explicitly address this endogeneity problem. To do so using values of covariates at $t-2$ and $t-3$ as instruments, we may implement instrumental variables techniques. Such a task may appear more involved in the context of semiparametric estimation since conventional instrumental variables techniques seem to be not directly transferable in a semiparametric framework. However, Yatchew (2003) presents an approach for handling endogeneity in nonparametric estimation. Using that approach, let's denote w as a vector of instruments for $x_{(t-1)}$ uncorrelated with $\epsilon_{i(t)}$, with w being given by $w_{(t-1)} = (x_{(t-2)}, x_{(t-3)})'$, then:

$$x_{(t-1)} = w'_{(t-1)}\pi + u_{(t-1)} \quad E\left(u_{(t-1)}|w_{(t-1)}\right) = 0 \quad E\left(\epsilon_{(t)}|w_{(t-1)}\right) = 0 \tag{3.15}$$

Suppose now that $E\left(\epsilon_{(t)}|x_{(t-1)}\right) = \rho u_{(t-1)}$ so that $\epsilon_{(t)} = \rho u_{(t-1)} + v_{(t)}$. We can thus rewrite equation (3.4) as:

$$\Delta ln\big(Y_{i(t)}/L_{i(t)}\big) = z'_{2i(t)}\beta_1 + f(x_{i(t-1)}) + v_{i(t)}, \quad v_{i(t)} \sim \mathcal{N}(0,\sigma^2) \tag{3.16}$$

with $z_{2i(t)} = \left(z'_{1i(t)}, u_{(t-1)}\right)'$, $\beta_1 = (\beta', \rho)'$, and $E\left(v_{(t)}|x_{(t-1)}, z_{2(t-1)}\right) = 0$.

After differencing equation (3.16), we may apply instrumental variables estimation to the parametric variables only, moving therefore from an endogeneity problem in a nonparametric estimation to an endogeneity problem in a parametric specification.

Rewriting (3.16) as

$$Dy = DZ_2\beta_1 + Df(x) + D\epsilon \approx DZ_2\beta + D\epsilon \tag{3.17}$$

where $y = \left(\Delta ln(Y_1/L_1), \ldots, \Delta ln\big(Y_{N(T-1)}/L_{N(T-1)}\big)\right)'$, $Z_2 = \left(z_{21}, \ldots, z_{2N(T-1)}\right)'$, $x = \left(x_1, x_2, \ldots, x_{N(T-1)}\right)'$,

we may rearrange (3.17) in order to distinguish regressors that are endogenous and those that are predetermined. Thus, we get:

$$Dy = DY_1\beta_2 + DZ_{11}\gamma + Df(x) + D\epsilon \approx DY_1\beta_2 + DZ_{11}\gamma + D\epsilon \tag{3.18}$$

where Y_1 is a $N(T-3)\times m_1$ data matrix gathering all endogenous regressors: stock of capital per capita, average years of schooling and u_{t-1}, the residual of the regression of x_{t-1}, on its instruments w_{t-1}, and Z_{11} is a $N(T-3)\times k_1$ data matrix collecting all predetermined regressors i.e. time dummies.

Equation (3.18) is only one of the structural equations of a static *simultaneous equations model* (SEM) containing as much equations as endogenous variables. We may estimate parameters of (3.18) in a 'limited information spirit', i.e. without explicit consideration to the restrictions pertaining to the remaining structural equations. To do so it is necessary to join to (3.18) the reduced form corresponding to endogenous variables that appear as regressors in (3.18)

$$DY_1 = DZ_1\Pi + DV_1 = DZ_{10}\Pi_0 + DZ_{11}\Pi_1 + DV_1 \tag{3.19}$$

where $Z_1 = (Z'_{10}, Z'_{11})'$ stands for the $N(T-3)\times k$ matrix gathering predetermined regressors as well as instruments of endogenous variables, gathered in data matrix Z_{10}.

Thus, we have to estimate the following system of equations

$$\begin{aligned} Dy &= DY_1\beta_2 + DZ_{11}\gamma + D\epsilon & (3.20)\\ DY_1 &= DZ_{10}\Pi_0 + DZ_{11}\Pi_1 + DV_1 & (3.21) \end{aligned}$$

3.2.6 Method for handling data missingness

The dataset we are working on is characterized by a significant rate of data missingness. From the 679 observations corresponding to 97 countries, 204, i.e. 30% observations have at least one missing value. Moreover, 3 of the 5 variables have missing observations.[4]

Table 3.1: List of variables with missing values

$N°$	Variable	Missing values	% Missing
1	economic growth rate	158	23
2	capital growth rate	158	23
3	total average year of schooling	117	17

The general way to handle missing data is to transform the incomplete dataset into a complete one. The usual practice to artificially create a complete data set implies either:

[4]Time dummies are not included.

throwing away cases with missing values (listwise deletion), or imputing, i.e., estimating and filling in, missing data using some ad hoc method like mean imputation, regression-based imputation, dummy variable adjustment, hot-decking.... Then one treats the altered data set as if the deleted cases had never been observed, or the imputed values had always been observed (Schimert *et al.*, 2001).

Listwise deletion and ad hoc methods can lead to misleading inferences because they either throw away or distort information in the data. Listwise deletion is the usual practice for handling data missingness in empirical research in economics. Yet, by throwing away information, listwise deletion may cause at best a significant loss of information, and they may even induce a severe selection bias if data are not *Missing Completely at Random* (MCAR), i.e. missing data are not a random sample of the complete dataset (Schafer, 1997; King *et al.*, 2001). Ad hoc imputation methods don't fix the problem. For instance imputing averages on a variable-by-variable basis preserves the observed sample means, but it distorts the covariance structure. On the other hand imputing predicted values from regression models inflates observed correlations. More generally ad hoc imputations methods by treating imputed data as if there were real fail to reflect any uncertainty due to missing data and thus produce biased standard errors, and p-values (Schafer, 1997).

Therefore, the appropriate way to handle missing data is then to rely on model-based imputation methods. In that category multiple imputation has a clear advantage over single imputation. Indeed, conversely to the latter, it yields inference reflecting sampling variability due to the missing values (Schimert *et al.*, 2001). Multiple imputation methods are unbiased if the missing-data mechanism is ignorable.[5]

Schafer (1997) compares model based imputation and listwise deletion by evaluating the performance of Maximum Likelihood (ML) estimates and listwise deletion (CC for Complete Case) estimates by simulation. To do so he consider a bivariate dataset with variable Y_1 completely observed for units $1, 2, ..., n$ and Y_2 observed only for units $1, 2, ..., n_1$ with $n_1 < n$ the number of observations.

He showed that the CC estimate is biased whenever $\rho \neq 0$ for the non-MCAR missingness mechanisms whereas the ML estimates are unbiased under all the missingness mechanisms.[6] Moreover, under the more restrictive MCAR, ML estimates have an advantage

[5]A missing-data mechanism is ignorable if it is *Missing at Random* (MAR), i.e the probability that datum is missing may depend on the datum itself but only through variables that are observed and if the parameters $\theta = vec(\mu, \Sigma)$ of the data model and the parameters ϕ of the missingness mechanism are distinct. Indeed, under ignorability, neither the model of the missingness mechanism nor the nuisance parameters ϕ are relevant for making inferences about θ (Schafer, 1997).

[6]ρ is the correlation coefficient.

over CC estimates whenever $\rho \neq 0$ because their variances are lower. The explanation for this low variance is that Y_1 becomes an increasingly valuable predictor of the missing values of Y_2 as ρ increases. Therefore, from considerations of bias, consistency and efficiency ML estimates are superior to CC estimates.

Model-based multiple imputation methods assume a statistical model for the distribution of data. The model that is the most widely used is the joint multivariate normal model. It requires special iterative computation tools to extract meaningful summaries like parameters estimates and standard errors. Those computation tools proceed generally in two steps. First, conditionally on the observed values and a starting value of θ - the parameters matrix -, missing values are imputed. Then, once missing values are imputed, a complete dataset is obtained from which θ can be computed. Given this value of θ we can reperform the first step and so on until the algorithm converges (Schafer, 1997). Two principal classes of algorithm are generally considered in the context of model-based multiple imputation methods: Data Augmentation (DA) and Expectation-Maximization (EM) algorithm and extensions.

In this chapter we implement a Data Augmentation algorithm. But we can't use it as such. In fact variables in the regression model (3.5) differ merely by the fact that some of them are independent variables and one is a dependent variable. Depending on the kind of variable we are dealing with, different distributions should be considered. While the standard multivariate model may hold for the group of independent variables, it is more logic to impute the dependent variable according to distribution implied by the regression model (3.5), i.e considering that:

$$y \sim \mathcal{N}\left(W\theta, \sigma^2 I_{N(T-1)}\right) \tag{3.22}$$

But then the imputation process becomes much more involved since the dependent variables should be imputed conditionally on the independent variables that have also to be imputed. To perform such a two-step imputation we implement a Gibbs-Sampler algorithm.

3.2.7 Estimation-Imputation algorithm

As indicated above our algorithm is divided in two steps: firstly imputation of missing covariates and secondly estimation of parameters and imputation of the dependent variable.

Imputation of covariates by a data augmentation algorithm

The Data augmentation algorithm is merely a Gibbs sampler implying two steps that are performed iteratively: an I-step and a P-step.[7] The I-step simulates missing values of covariates, given observed values of covariates and values of parameters of the data matrix computed at the preceding iteration

$$X_{mis}^{(t+1)} \sim P\left(X_{mis}|X_{obs}, \theta^{(t)}\right), \tag{3.23}$$

while the P-step simulates values of the data matrix parameters at the current iteration given observed values of covariates and missing values of covariates computed at the current iteration

$$\theta^{(t+1)} \sim P\left(\theta|X_{obs}, X_{mis}^{(t+1)}\right). \tag{3.24}$$

Bayesian estimation and imputation of missing values of the dependent variable

Without any endogeneity concern, imputation of the dependent variable is once again a Gibbs Sampler involving two steps: firstly imputation of the missing values of the dependent variable according to the regression model and conditionally on observed values of the regressand, on values of the regressors and on parameters of the regression model computed at the previous iteration:

$$y_{mis}^{t+1}|y_{obs}, X, \beta^t, \sigma_t^2 \sim \mathcal{N}(X\beta^t, \sigma_t^2) \tag{3.25}$$

and secondly draws of the parameters from a Normal-Inverted Gamma distribution conditionally on observed values of the dependent variable, on values of the dependent variable imputed at the current step, and on the regressors

$$\beta^{t+1}, \sigma_{t+1}^2|y_{obs}, X, y_{mis} \sim NIG\left(\hat{\beta}, X'X, s, n-k-2\right) \tag{3.26}$$

which implies firstly drawing σ^2 from an Inverted Gamma distribution and then conditionally on σ^2 drawing β from a Normal distribution

$$\beta^{t+1}|\sigma_{t+1}^2 \sim \mathcal{N}\left(\hat{\beta}, \sigma_t^2\left(X'X\right)^{-1}\right) \tag{3.27}$$

$$\sigma_{t+1}^2 \sim IG_2\left(n-k-2, s\right) \tag{3.28}$$

With endogenous regressors things are more involved. There is a huge amount of literature on Bayesian limited information estimation of SEM (Dreze and Richard, 1983; Zellner

[7] A detailed description of this algorithm is provided in Appendix C. The algorithm pseudocode is described in Appendix E.

et al., 1988; Bauwens and Van Dijk, 1990). Dreze and Richard (1983) provided results on exact Bayesian analysis of SEM and showed that for a specific choice of prior the posterior distribution of parameters is a poly-t density. The problem is that such a distribution is generally not analytically tractable. It must be integrated numerically to obtain moments, marginal distribution , etc. Moreover, it does not have simple forms from which draws of structural parameters can be made easily.

Zellner *et al.* (1988) provide an alternative approach to deal with Bayesian limited information estimation which avoids such shortcomings by allowing direct Monte Carlo simulation. Following their approach we estimate structural coefficients of equation (3.18) by a Gibbs Sampler. In the first step this algorithm draws the vector of structural parameters from a multivariate student conditionally on data and on the reduced form parameters. Then on a second step it draws reduced form parameters conditionally on data and on structural form coefficients. Let's rewrite (3.20) as

$$Dy = W_2\delta_2 + D\epsilon. \tag{3.29}$$

The first step draws the vector of structural form parameters δ_2 from $p(\delta_2|\Pi, D)$ which is a multivariate student density having as parameters

$$\tilde{\delta_2} = \left(W_2' M_{V1} W_2\right)^{-1} W_2' M_{V1} Dy \tag{3.30}$$

$$s_1^2 = \frac{1}{\nu_1}\left(Dy - W_2\tilde{\delta_2}\right)' M_{V1}\left(Dy - W_2\tilde{\delta_2}\right) \tag{3.31}$$

with $\nu_1 = N(T-3) - (m_1 + k_1)$ the number of degrees of freedom and

$$M_{V_1} = I_{N(T-3)\times N(T-3)} - DV_1(V_1' D' D V_1)^{-1} V_1' D'$$

The second step draws the matrix of reduced form parameters Π from $p(\Pi|\delta_2, D)$ which is a matricvariate student density having as formula

$$p\left(\Pi|\tilde{\delta_2}, D\right) = f_{MT}^{k\times m_1}\left(\Pi|\tilde{\Pi}, Y_1'\tilde{M}_\epsilon Y_1, Z' M_\epsilon Z, N(T-3) - k\right), \tag{3.32}$$

where $\tilde{\Pi} = \left(Z' M_\epsilon Z\right)^{-1} Z' M_\epsilon Y_1$ and

$$\tilde{M}_\epsilon = M_\epsilon - M_\epsilon Z\left(Z' M_\epsilon Z\right)^{-1} Z' M_\epsilon,\ M_\epsilon = I_{N(T-3)\times N(T-3)} - D\epsilon\left(\epsilon' D' D\epsilon\right)^{-1}\epsilon' D'$$

We estimate parameters at each step by the means of each conditional distributions considered in the previous approach rather than drawing from those distributions. This procedure corresponds merely to performing 2SLS estimation at each iteration.

3.3 Results

3.3.1 Convergence of the missing data imputation algorithm

To ensure convergence of the overall algorithm we ran 2000 iterations.[8] Several convergence checks were necessary to select a number of iterations allowing us to obtain reliable imputations of regressors missing data. Figures 3.8 and 3.9 in Appendix D depict results for various diagnosis of convergence for respectively 2000 and 5000 iterations. As we see even for 2000 iterations autocorrelation between draws dies out very quickly, the graphs of the standardized CUMSUM statistics (convergence graphs) converge smoothly to zero, and plots of draws do not show any long run tendency. Diagnosis of convergence for 5000 iterations confirm that convergence is definitely achieved after 2000 iterations since no excursions away from zero are observed after that number of draws. Therefore, we opt to fix the number of iterations of the imputation algorithm to 2000.

3.3.2 Estimation results

Basic productivity model

We start our presentation of the results by showing the outcomes of the estimation of a basic productivity model excluding any urban concentration variable. In such a model the only argument in the productivity growth function $ln(A_{i(t)}/A_{i(t-1)})$ is education of the labor force. Table 3.2 presents those baseline results.

Table 3.2: Estimation of a basic productivity growth equation. dependent variable is $\Delta ln\big(Y_{(t)}/L_{(t)}\big)$.

	(1) Fixed effects	(2) IV
$\Delta ln\big(K_{(t)}/L_{(t)}\big)$	0.5034^{**}	0.2393^{**}
Human capital	0.0006	0.0183^{**}
Years effects	Yes	Yes
N[countries]	582[97]	388[97]

The multiple imputation method implemented allows a higher data coverage than the one obtained with listwise deletion methods. As only 68 countries have complete information, casewise deletion methods would perform estimation only on a much lower number

[8] In Appendix D we describe the criteria used to assess convergence.

of observations. Henderson (2003) estimates that basic productivity model from only 482 observations corresponding to 82 countries. Results of Fixed Effects estimation procedures yield estimates of elasticity of capital that are much higher than capital coefficient reported by the literature. Furthermore, coefficient of the human capital proxy is non significant. While the literature indicates that results on the education variable are non robust (Temple, 1999), those obtained on capital elasticity are clearly questionable. Hall and Jones (1999) assumes 0.33 as a capital coefficient and estimates provided by Henderson *et al.* (2001) from work on Korea lie in the range of 0.37-0.39. IV method obtains different results. With an estimate of elasticity of capital of 0.239, IV result is even lower than what is generally accepted by the literature. Moreover, the estimate of human capital coefficient (0.0183) has become significant.

By looking at table 3.3, we can notice the gap between estimates obtained with the listwise deletion method and earlier estimates obtained with imputation. For OLS estimation, listwise deletion implies higher estimates for both independant variables. For instrumental variables estimation, it entails a higher estimate of the elasticity of capital and a lower estimate of the human capital regression coefficient. Moreover, table 3.3 clearly indicates the significant loss of degrees of freedom resulting from the use of casewise deletion method: only 69 countries provide relevant observations for estimation. This outlines the payoff yielded by the use of imputation methods.

Table 3.3: Estimation of a basic productivity growth equation with the listwise deletion method. Dependent variable is $\Delta ln(Y_{(t)}/L_{(t)})$.

	(1) Fixed effects	(2) IV
$\Delta ln(K_{(t)}/L_{(t)})$	0.5163^{**}	0.2729^{**}
Human capital	0.0035	0.0161^{**}
Years effects	Yes	Yes
N[countries]	406[69]	268[69]

Productivity model with primacy

Parametric estimates

Table 3.4 presents results of a productivity model including primacy as the urban concentration measure. As before OLS estimation procedure provides a capital coefficient that is

Table 3.4: Estimation of a productivity growth equation with primacy. Dependent variable is $\Delta ln(Y_{(t)}/L_{(t)})$.

	(1) OLS	(2) IV
$\Delta ln(K_{(t)}/L_{(t)})$	0.5221^{**}	0.3402^{**}
Human capital	0.0169^{**}	0.0169^{**}
Years effects	Yes	Yes
N[countries]	582[97]	388[97]

much higher than commonly found in the literature. Conversely, to the previous estimation the human capital coefficient is now significant. Yet, its value is much weaker than the one obtained by Henderson' (2003) GMM estimation.

Instrumental variables estimation yields more appealing results for the parametric coefficients. The elasticity of capital is now in line with the literature. Our results are even closer to literature than estimates provided by Henderson *et al.* (2001) from work on Korea which lie in the range of 0.37-0.39 and are thus higher. Since OLS and IV procedures yield equivalent values of the human capital coefficient, the former does not seem to suffer from a significant endogeneity bias.

Non-parametric curves

For OLS procedure, estimation of the non-parametric regression curve for the overall sample provides a so irregular pattern that no clear lessons can be drawn from it. Conversely, when we consider sub-samples constituted from geographical or developmental criteria, procedures provide interesting results. Sub-samples non-parametric results are reliable only with imputation. Indeed, as table 3.12 shows, regions such as Asia and Sub-Saharan Africa have a significant share of observations with missing data (respectively 32.92% and 48.45%). The resulting reduction of sample sizes would preclude meaningful non-parametric estimation.[9]

The pattern shown by the non-parametric curve estimated for Europe (Figure 3.1) is quite irregular. However, except for values of primacy below 0.2 where GDP decreases with primacy or lying in the range 0.23-0.39 where the curve is flat, the curve exhibits a

[9]Table 3.12 shows that Europe has a significant share of missing data as well (28.57 %). This is caused by the poor information obtained from countries of Eastern Europe for the GDP per capita and the capital per worker variables.

positive slope. This slope is even the sharpest for the 0.20-0.23 and the 0.39-0.46 ranges. This curve suggests that except for low values of primacy, GDP is globally an increasing function of primacy. So while low values of primacy appear to be detrimental to economic growth, urban concentration seems to be associated to economic growth for most of its variation range. A result that recalls findings of Wheaton and Shishido (1981) with a sample gathering countries from all regions.

Countries from other continents exhibit a very different picture as shown by non-parametric curves drawn from other regions of the World. Figure 3.1 also shows the non-parametric curve obtained from Asia. This curve is irregular and exhibits a visible urban concentration trap.

Latin America non-parametric curve appears as the most irregular, indicating that it is difficult to draw out a smooth non-parametric regression function from this subsample (Figure 3.2). This non-parametric curve declines sharply after a level of primacy higher than 0.30. Latin America has been for a long time the region where primacy concerns are the highest. This result indicates that the huge urban concentrations prevailing in that region are detrimental to economic growth confirming the prejudice that most of countries of that region are overconcentrated.

Africa provides an interesting picture. It is the only continent that exhibits the bell shaped predicted by Williamson hypothesis. Economic growth increases sharply with primacy until a maximum of about 0.38 is reached then it declines till a primacy level equal to 0.55. After that value it exhibits a globally flat pattern. This picture suggests that African countries that have levels of primacy below that maximum have not exhausted agglomerations economies yielded by their primate cities. Conversely, African countries with levels of primacy greater than that maximum may be expected to have congestion, pollution and other centrifugal forces overwhelming agglomeration economies.

Results from tests of equality of non-parametric regressions functions estimated by OLS procedures signal heterogeneity in the overall sample, confirming the impression yielded by the previous pictures (Table 3.5). Deeper inspection indicates however that this heterogeneity lies mostly in the opposition between developing countries and developed ones, and between Europe and other regions of the World. For regions pertaining to the developing World there is no evidence of significantly different patterns. So there is some gap between the result of those statistical tests and the impression provided by the pictures.

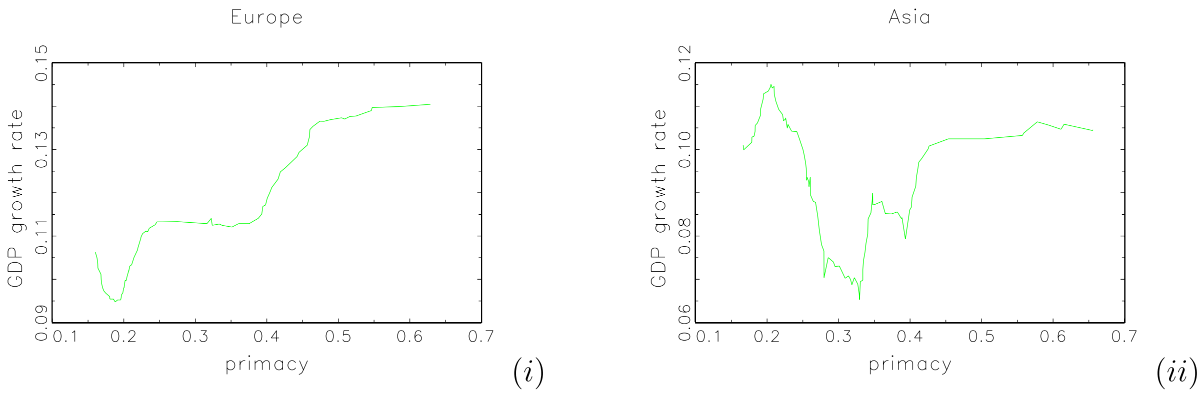

Figure 3.1: European and Asian non-parametric curves. OLS procedure

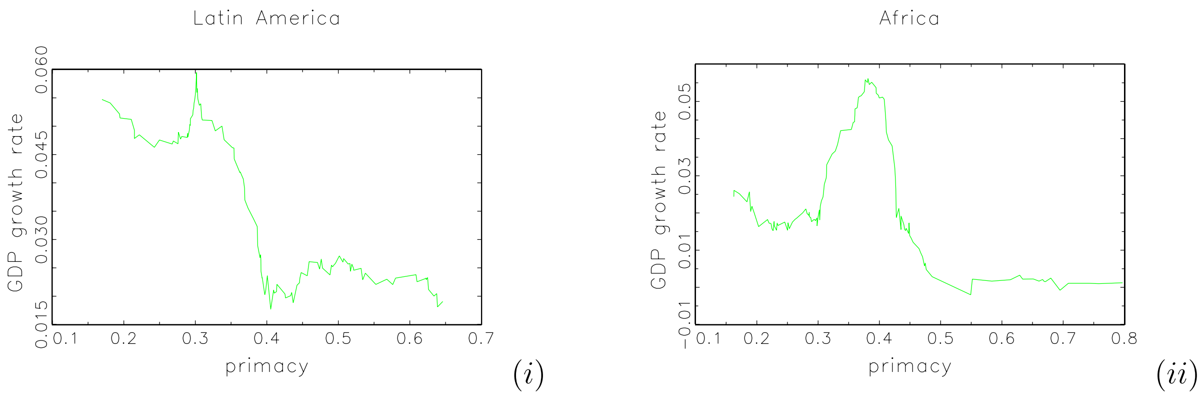

Figure 3.2: Latin America and African non-parametric curves. OLS procedure

Obviously IV procedures are expected to be the most reliable, yet they provide results that are very similar to the previous curves.[10] The non-parametric curve obtained from European countries is globally increasing (Figure 3.3) suggesting as before that urban development in that region is well balanced and efficient.

While 'overconcentration'[11] does not appear to be a pertinent description of Europe, the picture from the developing countries appears to be more mitigated. Inspection of non-parametric curves for Asia and Latin America uncovers as before the existence of urban concentration traps. Such trap is much more significant for Latin America where a significant range of urban primacy appears to be associated with negative economic growth.[12]

[10]Yet, because of the loss of degrees of freedom, the Europe non-parametric curves obtained for the instrumental variables procedure seem to be very imprecise. Thus, in Appendix F devoted to confidence intervals, we replace them by non-parametric curves estimated for developed countries.

[11]The term is understood here as the prevalence of excessive urban concentrations

[12]Argentina, Chile, Dominican Republic, Guatemala, Nicaragua, Peru and El Salvador are the countries of that region where primacy seems to have the most detrimental impact on economic growth.

Table 3.5: Tests of equality of non-parametric regression function

	P-value
Overall sample	0.004**
Developing VS Developed countries	0.000**
Europe VS Africa	0.000**
Asia VS Africa	0.323
Latin America VS Africa	0.433

This is consistent with the peculiar spatial distribution prevailing in Latin America. This continent is well known for the loose integration of several of its very vast regions. This has induced a spatial configuration with a strong urban concentration and sharp contrasts between on one hand fast growing metropolitan areas and vast abandoned rural regions. This result is also consistent with the failure of Latin America development strategy based on import substitution and comforts the idea that Latin America is the region of the world where primacy issues are the most involved.

Asia faces a similar urban concentration trap. Yet, the negative impact of primacy on its economic development seems less dramatic, most of the range of primacy variation being associated with a positive economic growth. The picture from Black Africa confirms previous results as well. Black Africa remains the only geographical area where the pattern of the relationship between growth and primacy is similar to the prediction of the Williamson hypothesis. Inspection of the African non-parametric curve indicates a maximum at about 37%. Thus, urban concentration of countries having a primacy below that level is not detrimental to economic growth. The lopsided spatial distribution of several African countries does not seem to be an obstacle to their economic development as they do not appear to have exhausted their agglomeration economies. Mali, Chad and Zimbabwe are the countries that are the closest to that optimum. Conversely, SSA countries with a primacy higher than 0.45 - like Angola, Congo, Guinea, Mozambique, Senegal - have the worst economic growth rates.

While differences appear from the comparison of the patterns of different regions, tests of equality of non-parametric functions don't signal any difference between the different non-parametric curve at the 5% signification level. This contrasts singularly with the result obtained with the OLS estimation procedures. This striking contrast may be caused by the loss of degrees of freedom implied by the use of lagged variables as instruments in IV

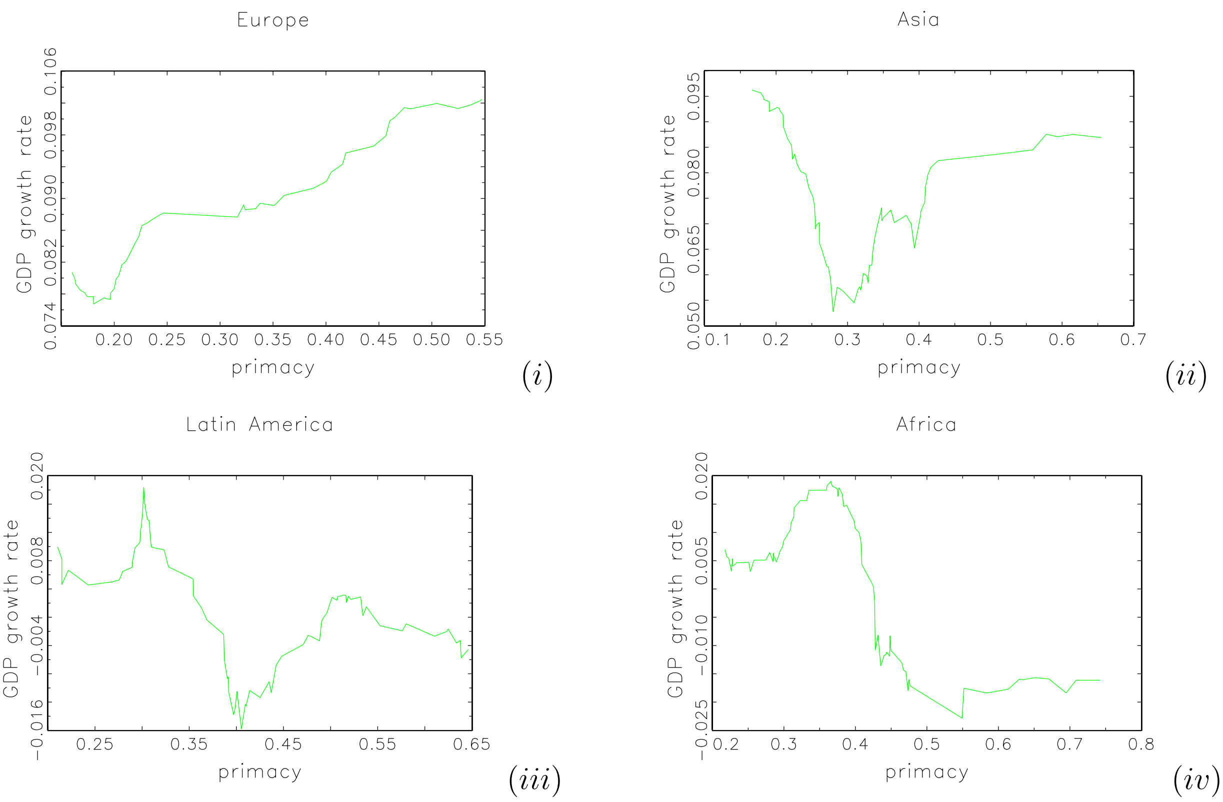

Figure 3.3: Non-parametric curves. IV procedure

estimation methods.

Table 3.6: Tests of equality of non-parametric regression function

	P-value
Overall sample	0.149
Developing VS Developed countries	0.346
Europe VS Africa	0.193
Asia VS Africa	0.123
Latin America VS Africa	0.365

Table 3.7: Estimation of a productivity growth equation with urban density. Dependent variable is $\Delta ln(Y_{(t)}/L_{(t)})$.

	(1) OLS	(2) IV
$\Delta ln(K_{(t)}/L_{(t)})$	0.5148^{**}	0.2802^{**}
Human capital	0.0192^{**}	0.0207^{**}
Years effects	Yes	Yes
N[countries]	582[97]	388[97]

Productivity model with urban density

Parametric estimates

Table 3.7 shows results of a productivity model including urban density as the urban concentration measure. As for primacy estimations, OLS estimation procedure yields a capital coefficient that is much higher than what is commonly assumed in the literature. Conversely to primacy estimations, OLS estimate of the human capital coefficient is significant. However, while its value is a little higher than the ones obtained by primacy estimation procedures, it is still lower than Henderson' (2003) GMM estimation.

Instrumental variables estimation with the urban density variable corrects the capital coefficient which reaches a value that is even lower than what is generally assumed by the literature. Yet, IV estimate of the human capital coefficient is only slightly higher than the OLS one, indicating as before that the human capital coefficient is quite unaffected by any endogeneity bias.

Non-parametric curves

From OLS estimation we find that the pattern exhibited by the non-parametric curve for Europe (Figure 3.4) is as previously globally increasing. The shape is very similar to the one exhibited by the primacy non-parametric curve obtained under OLS estimation. Indeed, except for values of urban density below 0.2 and for values lying in the 0.27-0.39 range where the curve is flat, the curve exhibits a positive slope which is the highest in the 0.20-0.27 and the 0.39-0.42 ranges. This confirms the previous result that in Europe urban concentration seems to be associated to economic growth for most of the its variation range.

The previous findings appear to be robust for Asia and Latin America as well. A urban density trap is clearly visible in the non-parametric curve drawn for Asia and Latin America

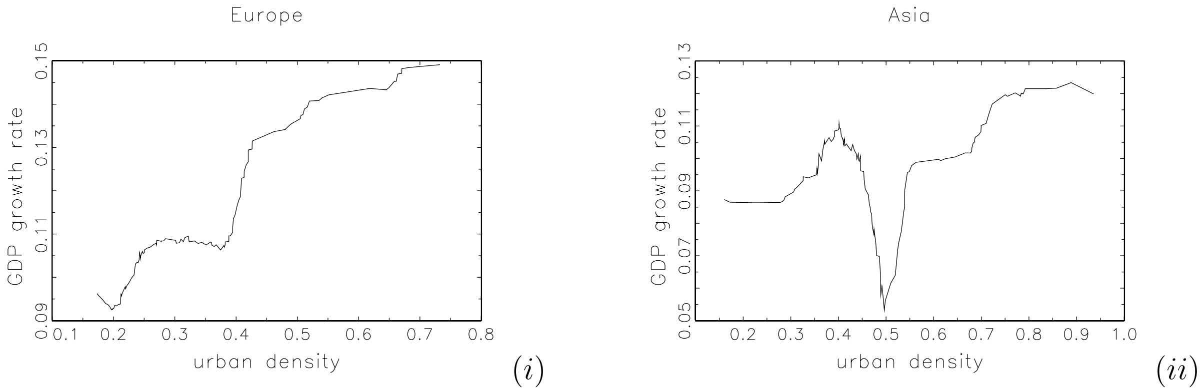

Figure 3.4: European and Asian non-parametric curves. OLS procedure

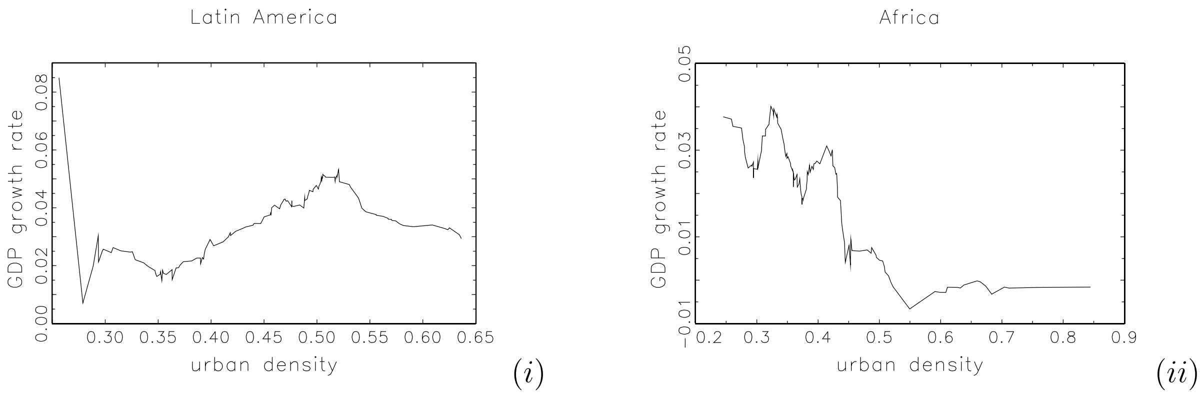

Figure 3.5: Latin America and African non-parametric curves. OLS procedure

in OLS estimation procedure. As previously, Latin America non-parametric curve is the least smooth. Results for Africa diverge from previous findings. While a inverted U-shape was clearly emerging for Africa in primacy non-parametric curves, this is clearly not the case for African OLS urban density non-parametric curve since it displays a succession of local maxima with a globally decreasing shape. This very irregular pattern does not ease any global interpretation of the impact of urban concentration of the economic growth of countries of this region of the world, urban density may raise or lower economic growth depending on the specific range of variation that we consider.

Tests of equality of non-parametric curves comfort at first sight this picture of diverging patterns across regions. While for the overall sample, heterogeneity in the non-parametric curves does not appear significant, tests of equality signal significant difference between developed and developing countries. However, conversely to differences suggested by the graphs of non-parametric curves, those tests fail to indicate any significant difference within the developing world between Africa and the Latin American and Asian continents.

Table 3.8: Tests of equality of non-parametric regression function

	P-value
Overall sample	0.099*
Developing VS Developed countries	0.054*
Europe VS Africa	0.007**
Asia VS Africa	0.114
Latin America VS Africa	0.142

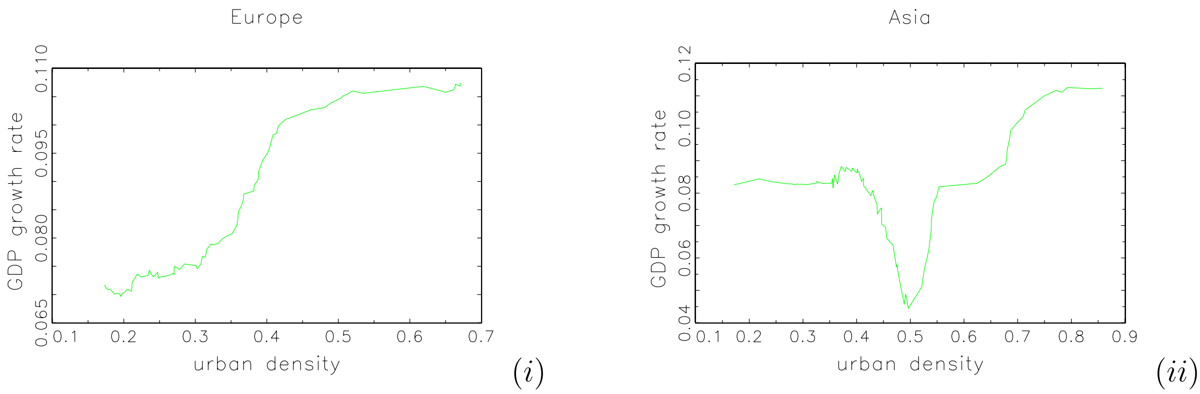

Figure 3.6: European and Asian non-parametric curves. IV procedure

Except for Africa IV estimation provides patterns of non-parametric curve that are very close to those obtained by OLS procedures. IV procedures even yield an Europe non-parametric curve that has an increasing shape for all the variation range of the urban density variable (Figure 3.6). Africa non-parametric curve exhibits a globally decreasing shape (Figure 3.7). This result suggests that countries of that region have not the economic infrastructure that may support a high share of population in big cities. Furthermore, as average urban density for the Black Africa amounts to about 0.39, such a non-parametric curve clearly indicates that Sub-Saharan Africa is overurbanized. But as this result diverges from previous findings with primacy non-parametric curves, this outlines more seriously that in some regions those two indicators of urban concentration are poorly correlated.

As before tests of equality of non-parametric regression curves yield contrasted results. While tests applied to the overall sample confirm the impression yields by the graphs of diverging patterns, they fail to confirm that African and Latin America non-parametric curves are significantly different.

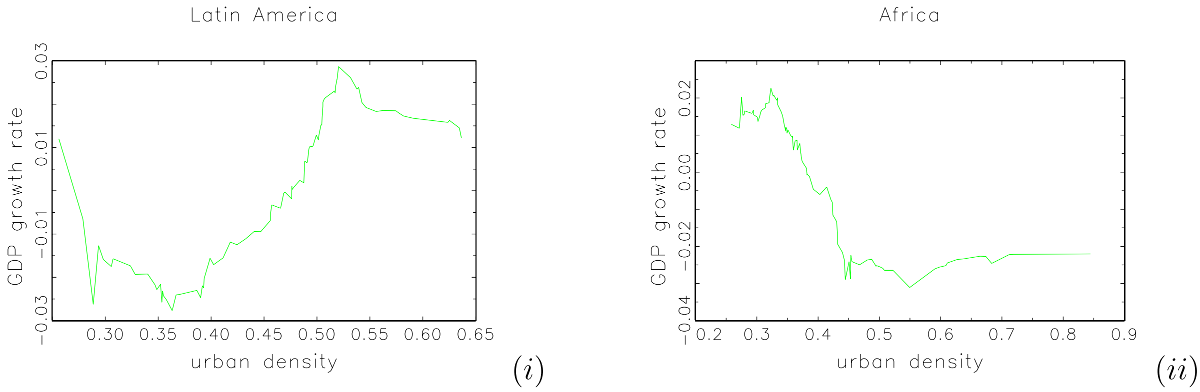

Figure 3.7: Latin America and African non-parametric curves. IV procedure

Table 3.9: Tests of equality of non-parametric regression function

	P-value
Overall sample	0.011**
Developing VS Developed countries	0.136
Europe VS Africa	0.021**
Asia VS Africa	0.031**
Latin America VS Africa	0.431

3.4 Conclusion

In a pretty old statement Hoselitz (1955) raised the claim that there is a contrast between urban development in developed nations and in developing countries: while in the former group there is an intimate connection between the economic demands for labor exerted by progressive accumulation of capital in urban industry and the growth of urban centers, this is not the case in many of the underdeveloped countries in Asia, Africa, and Latin America, where a number of other reasons seem to have induced rural-urban migration. That statement about the possible qualitative difference of the urban development in different groups of countries is at the heart of the issue addressed by this chapter.

Pointing out the heterogeneity between regions, our results back Hoselitz statement: urban concentration has a positive impact on economic growth in Europe, dummy traps prevails for Latin America and Asia, while Africa non-parametric curve differs depending of the measure of urban concentration considered. It exhibits a bell-shaped pattern with primacy but a curve with a decreasing slope when urban density is considered. In most cases tests of equality of non-parametric regression functions confirm the heterogeneity hypothesis.

Therefore, no general relationship between urban concentration and economic growth appears as credible. Any attempt to assess the impact of the spatial distribution on economic development should be addressed to groups of homogeneous countries. In this respect even the grouping we use for this empirical research may be improved. The poor smoothness of some of the non-parametric curves estimated, especially for Asia and Latin America may indicate that those groups are hardly homogeneous.

Moreover, Africa non-parametric curve does not seem to display a monotonically decreasing pattern with respect to urban primacy. Therefore, we may not discard the importance of agglomerations economies in Africa urbanization process. Those agglomerations economies will play an important role in the next chapters devoted to the modeling of the impact of respectively political factors and locational (dis)advantages on urban concentration.

3.5 Appendix A: Sample and Data Sources

The dataset consists of 97 countries.

Country list: Afghanistan, Algeria, Angola, Argentina, Australia, Austria, Belgium, Burkina Faso, Bangladesh, Bolivia, Brazil, Bulgaria, Canada, Chad, Chile, China, Côte d'Ivoire, Cameroon, Congo (Democratic Republic), Congo (Republic), Colombia, Costa Rica, Cuba, Czechoslovakia, Denmark, Dominican Republic, Ecuador, Egypt, El Salvador, Ethiopia, Finland, France, Ghana, Greece, Guatemala, Guinea, Haiti, Honduras, Hungary, India, Indonesia, Iran, Iraq, Ireland, Israel, Italy, Japan, Jordan, Kenya, Libya, Madagascar, Malaysia, Mali, Mexico, Morocco, Mozambique, Myanmar, Netherlands, New Zealand, Nicaragua, Nigeria, North Korea, Pakistan, Panama, Paraguay, Peru, Philippines, Poland, Portugal, Puerto Rico, Romania, Saudi Arabia, Senegal, Somalia, South Africa, South Korea, Spain, Sudan, Sweden, Switzerland, Syria, Tanzania, Thailand, Tunisia, Turquey, Uganda, Union of Soviet Socialist Republics, United Arab Emirates, United Kingdom, USA, Uruguay, Venezuela, Vietnam, West Germany, Yugoslavia, Zambia, Zimbabwe.

Table 3.10: Data Sources

Data	Sources
GDP/Capita Capital/Worker	Dhareshwar, A. and V. Nehru (1993), A New Database on Physical Capital Stock: Sources, Methodology and Results, Revista de Analisis Economico, 8, pp.37-59 http://www.worldbank.org/research/growth/ddnehdha.htm
Urban density Primacy	National urban population. UN World Urbanization Prospects CD-rom, File 10 POP/DB/WUP/Rev.2003/2/F10
Human capital	Measured by average schooling years in the male population aged 25 and over. Stems from Henderson (2003). Figures were obtained from Barro, R. and J.-W. Lee (2001) and from Census and survey figures primarily retrieved from UNESCO Statistical Yearbooks and UN Demographic Yearbooks. Remaining values are estimated using UNESCO school enrollment data and a perpetual inventory method. The data are not adjusted for quality of education day or or length of school year.

3.6 Appendix B: Descriptive statistics

Table 3.11: Descriptive statistics

Variables	Min	Max	Mean	Std Dev
Human capital	0.020	6.220	1.158	1.050
Primacy	0.044	0.797	0.312	0.149
Urban density	0.099	0.935	0.420	0.148
ln(Capital per worker)	6.994	11.706	9.622	1.176
ln(GDP per capita)	3.201	12.114	7.669	1.991

Table 3.12: Proportion of missing data by region

Region	Missing values	Sample size	%Missing
Asia	53	161	32.92
Europe	42	147	28.57
Latin America	14	147	9.52
North Africa	17	35	48.57
North America	0	14	0.00
Oceania	0	14	0.00
Sub-Saharan Africa	78	161	48.57

3.7 Appendix C: Imputation of covariates

For an arbitrary pattern of missing data parameter estimates cannot be obtained in closed form. Therefore, we resort to a data augmentation algorithm which implies iterative computations. To ease those computations, it is useful to group the rows of the covariates by their missingness pattern.

Following Schafer (1997), we may index the missingness patterns by $s = 1, 2, \ldots, S$, where S is the number of unique patterns prevailing in the covariates data matrix.[13] For a given data matrix X of dimension $n \times p$, with $n = NT$ let's define R as an $S \times p$ matrix of binary indicators with typical elements r_{sj}, where

$$r_{sj} = \begin{cases} 1 & \text{if } \mathrm{X_j} \text{ is observed in pattern s,} \\ 0 & \text{if } \mathrm{X_j} \text{ is missing in pattern s.} \end{cases} \tag{3.33}$$

Table 3.13: Matrix of missingness patterns associated with X.

	X_1	X_2	X_3	$\cdots$	X_p
patterns s=1	1	1	1		1
2	0	1	1		1
.	1	0	1		1
.	0	0	1		1
.	1	1	0		1
.	.	.	.		.
.	.	.	.		.
.	.	.	.		.
.	0	1	0		0
S	1	0	0		0

Table 3.13 shows the typical matrix R. For each missingness pattern s, let $\mathcal{O}(s)$ and $\mathcal{M}(s)$ denote the subsets of the columns labels $\{1, 2, \ldots, p\}$ corresponding to variables that are observed and missing, then we have respectively,

$$\begin{aligned} \mathcal{O}(s) &= \{j : \ r_{sj} = 1\} \\ \mathcal{M}(s) &= \{j : \ r_{sj} = 0\} \end{aligned} \tag{3.34}$$

Finally we denote by $\mathcal{I}(s)$ the subset of $\{1, 2, \ldots, n\}$ corresponding to rows of the data matrix exhibiting missingness pattern s.

[13] In this section our presentation will closely follow Schafer (1997)

The I-step

Since we assume that the rows of a data matrix are conditionally independent given θ simulation of (3.23) is carried out by drawing

$$x_{i(mis)}^{(t+1)} \sim P\left(x_{i(mis)}|x_{i(obs)}, \theta^{(t)}\right), \tag{3.35}$$

independently for $i = 1, 2, \ldots, n$. For a given row i in missingness pattern s the conditional distribution of $x_{i(mis)}$ given $x_{i(obs)}$ and θ is multivariate normal with means

$$E\left(x_{ij}|X_{obs}, \theta\right) = a_{0j} + \sum_{k \in \mathcal{O}(s)} a_{kj} x_{ik} \tag{3.36}$$

and covariances

$$Cov\left(x_{ij}, x_{ik}|X_{obs}, \theta\right) = a_{jk} \tag{3.37}$$

with $j, k \in \mathcal{M}(s)$, and a_{jk} denoting an element of the matrix

$$A = SWP\left[\mathcal{O}(s)\right]\theta \tag{3.38}$$

$SWP[\;]$ denotes the sweep operator. When applied to the parameters of a multivariate normal model it converts a variable from a response to a predictor. Considering $z \sim \mathcal{N}(\mu, \Sigma)$ a random vector of variables partitioned as $z^T = \left(z_1^T, z_2^T\right)$ with p_1 the length of z_1, $SWP[1, \ldots, p_1]\theta$ converts the parameters matrix θ

$$\theta = \begin{bmatrix} -1 & \mu^T \\ \mu & \Sigma \end{bmatrix} = \begin{bmatrix} -1 & \mu_1^T & \mu_2^T \\ \mu_1 & \Sigma_{11} & \Sigma_{12} \\ \mu_2 & \Sigma_{21} & \Sigma_{22} \end{bmatrix} \tag{3.39}$$

containing parameters of the marginal distributions of z_1 and z_2 to a matrix containing parameters of the conditional distribution of z_2 given z_1

$$SWP[1, \ldots, p_1]\theta = \begin{bmatrix} -1 - \mu_1^T\Sigma_{11}^{-1}\mu_1 & \mu_1^T\Sigma_{11}^{-1} & \mu_2^T - \mu_1^T\Sigma_{11}^{-1}\Sigma_{12} \\ \Sigma_{11}^{-1}\mu_1 & -\Sigma_{11}^{-1} & \Sigma_{11}^{-1}\Sigma_{12} \\ \mu_2 - \Sigma_{12}\Sigma_{11}^{-1}\mu_1 & \Sigma_{21}\Sigma_{11}^{-1} & \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \end{bmatrix} \tag{3.40}$$

$$= \begin{bmatrix} -1 - \mu_1^T\Sigma_{11}^{-1}\mu_1 & \mu_1^T\Sigma_{11}^{-1} & \alpha_{2.1}^T \\ \Sigma_{11}^{-1}\mu_1 & -\Sigma_{11}^{-1} & B_{2.1}^T \\ \alpha_{2.1} & B_{2.1} & \Sigma_{22.1} \end{bmatrix} \tag{3.41}$$

with

$$\alpha_{2.1} = \mu_2 - \Sigma_{12}\Sigma_{11}^{-1}\mu_1 \tag{3.42}$$

$$B_{2.1} = \Sigma_{21}\Sigma_{11}^{-1} \tag{3.43}$$

$$\Sigma_{22.1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \tag{3.44}$$

The P-step

Assuming that no prior information is available concerning θ, the complete data posterior is a normal-inverted Wishart distribution. Therefore, the P-step will consist merely of the following simulation

$$\mu|\Sigma, X \sim N(\bar{x}, n^{-1}\Sigma) \tag{3.45}$$

$$\Sigma|X \sim W^{-1}\left(n-1, (nS)^{-1}\right) \tag{3.46}$$

where $\bar{x}$ and S denote respectively sample mean and sample covariance matrix.

3.8 Appendix D: Convergence diagnosis

Each iteration of our overall estimation algorithm will consist in a completion of the imputation algorithm which will imply numerous iterations to achieve convergence and one iteration of each step of the algorithm for Bayesian estimation and imputation of missing values of the dependent variable. Therefore, if respectively N_I and N_E iterations are needed for the imputation and the Bayesian estimation algorithms the overall algorithm will perform $N_I \times N_E$ imputations and N_E parameter estimations.

To obtain reliable imputations or parameter estimates we have to run those Gibbs Sampler algorithms enough times to allow the algorithm to converge to the posterior distributions. Several procedures and statistics are available to assess convergence:

- Times series plot and autocorrelations: plotting iterates of components of θ is a quick and easy way to assess convergence. In case of fast convergence plots of iterates show no discernible trends; they resemble horizontal bands indicating a low ratio of noise to signal. For imputation algorithm it corresponds to situations where the fraction of missing data is moderate. Conversely, when the fraction of missing information is high, long-term trends and high serial correlation are likely to show up and the algorithm converges slowly. Another way to assess convergence is to investigate the relationship between iterates at time (t) and at time $(t+1)$. This may be done through the analysis of the autocorrelation function. If autocorrelations between draws die out very quickly, the convergence is fast. Conversely, if correlations are still high beyond 10 iterations draws display a high degree of serial dependence and convergence is slow.

- Geweke's test statistic: compares the estimate $\bar{g}_A$ of a posterior mean from the first draws with the estimate from the last draws $\bar{g}_B$. If the two subsamples (of size n_A and n_B) are well separated (i.e. there are many observations between them), they should be independent. The statistic, normally distributed if n is large and the chain has converged, is

$$Z = \frac{\bar{g}_A - \bar{g}_B}{nse_A^2 + nse_B^2}$$

where nse_A and nse_B represent numerical standard errors of each subsample

- Standardized CUMSUM statistic: the standardized CUMSUM for θ is:

$$CS_t = \left(\frac{1}{t} \sum_{i=1}^{t} \theta^t - m_\theta \right) / s_\theta$$

where m_θ and s_θ are the MC sample mean and standard deviation of the n draws. If the MCMC sampler converges, the graph of CS_t against t should converge smoothly to zero. On the contrary, long and regular excursions away from zero are an indication of the absence of convergence. A value of 0.05 for a CUMSUM after t draws means that the estimate of the posterior expectation diverges from the final estimate (after n draws) by 5 per cent in units of the final estimate of the posterior standard deviation.

Here are the graphics obtained for the autocorrelations, the standardized CUMSUM statistic (convergence graphs) and the sequence of draws for respectively 2000 and 5000 iterations:

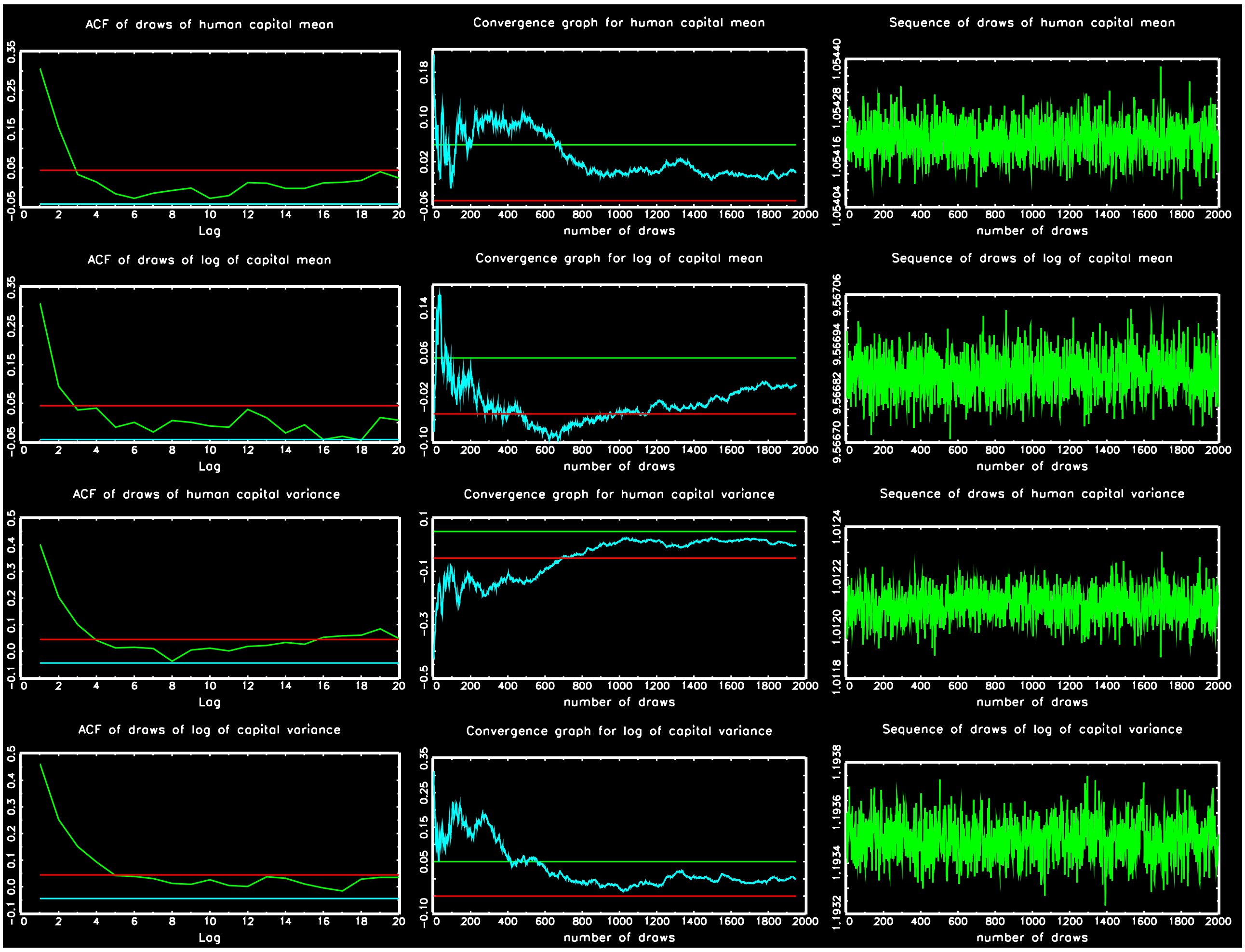

Figure 3.8: Convergence Diagnosis for 2000 iterations

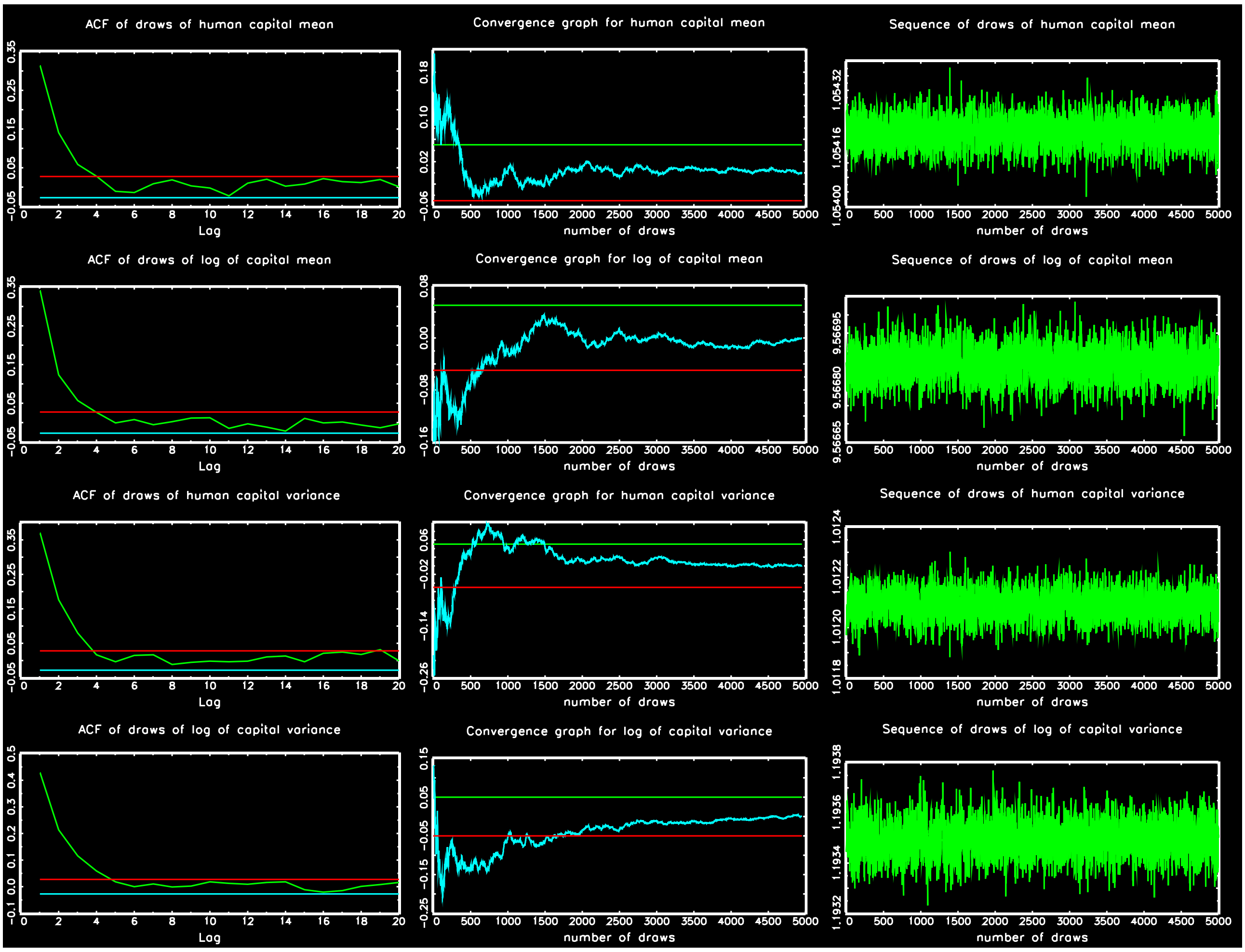

Figure 3.9: Convergence Diagnosis for 5000 iterations

3.9 Appendix E: Algorithm Pseudocode

In order to summarize our estimation procedure, we give hereafter our algorithm pseudocode, where D represents the data matrix. All the estimation and imputations procedures were implemented in Gauss.

for $i := 1$ to N_E do
  for $j := 1$ to N_I
    draw of $X_{mis}^{t+1} \sim P(X_{mis}/X_{obs}, \theta^t)$
    draw of $\theta^{t+1} \sim P(\theta/X_{obs}, X_{mis}^{t+1})$
  endfor

  if not endogeneity draw of $y_{mis}^{t+1}|y_{obs}, X, \beta^t, \sigma_t^2 \sim \mathcal{N}(X\beta^t, \sigma_t^2)$
    draw of $\beta^{t+1}, \sigma_{t+1}^2|y_{obs}, X, y_{mis} \sim NIG\left(\hat{\beta}, X'X, s, n-k-2\right)$

  else draw of $y_{mis}^{t+1}|y_{obs}, X, \delta_2^{t+1}, \sigma_t^2 \sim \mathcal{N}(X\delta_2^{t+1}, \sigma_t^2)$
    draw of $\delta_2^{t+1} \sim p(\delta_2|\Pi^t, D)$
    draw of $\Pi^{t+1} \sim p(\Pi|\delta_2^{t+1}, D)$
endfor

3.10 Appendix F: Confidence intervals

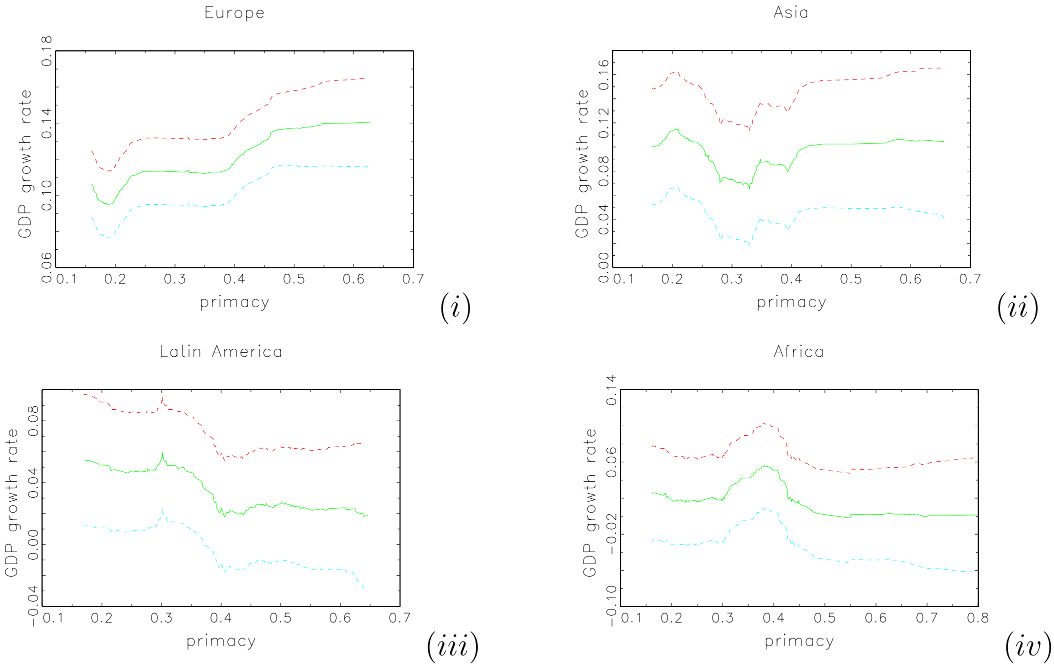

Figure 3.10: Non-parametric curves with primacy. OLS procedure

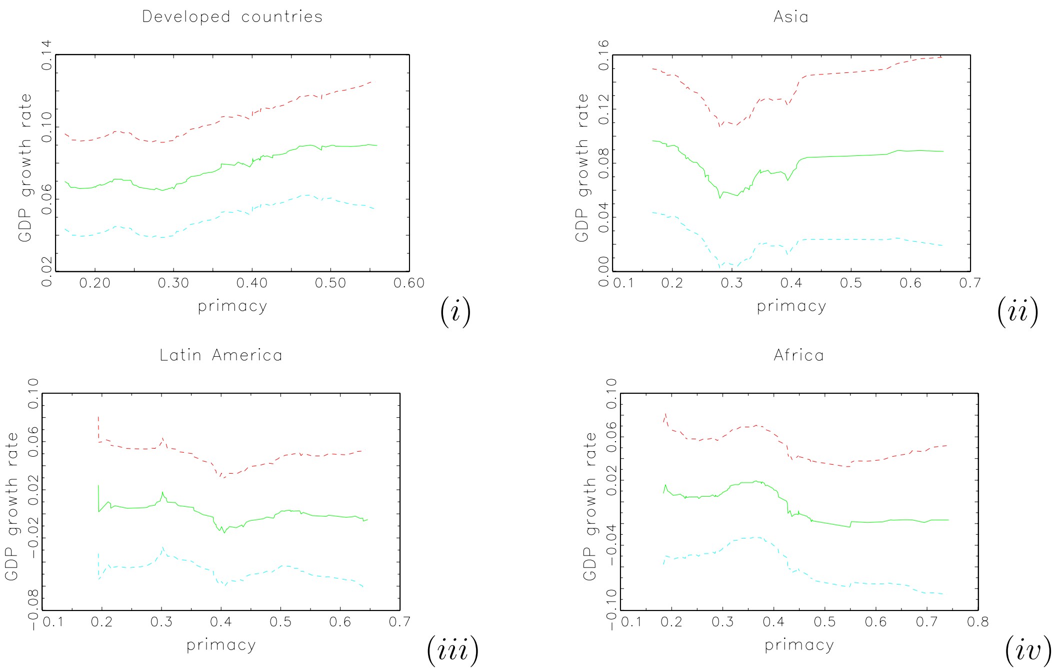

Figure 3.11: Non-parametric curves with primacy. IV procedure

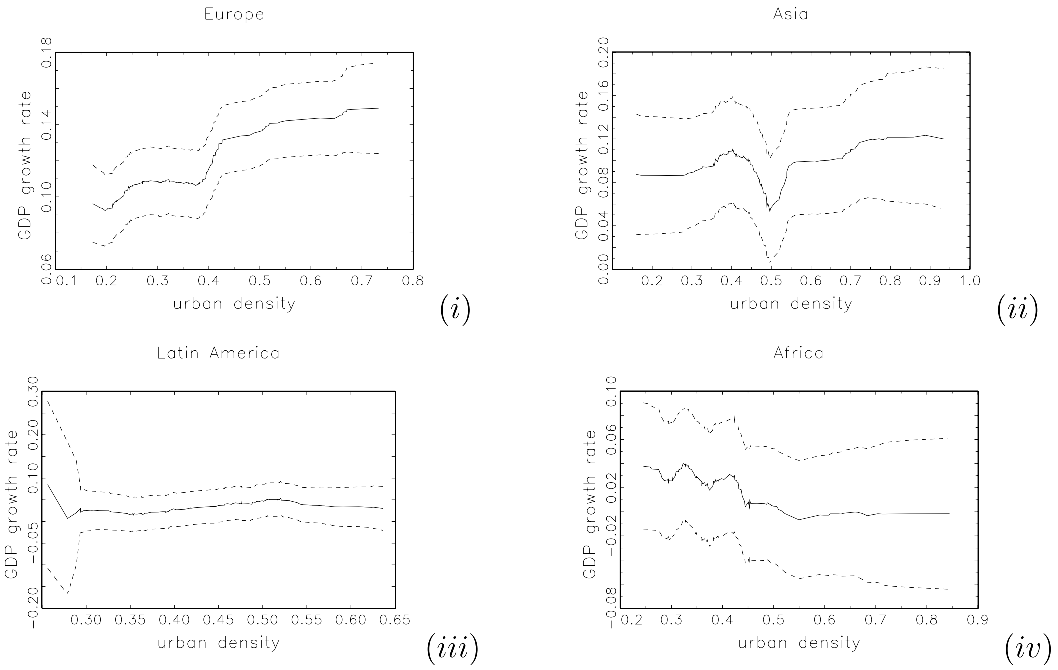

Figure 3.12: Non-parametric curves with urban density. OLS procedure

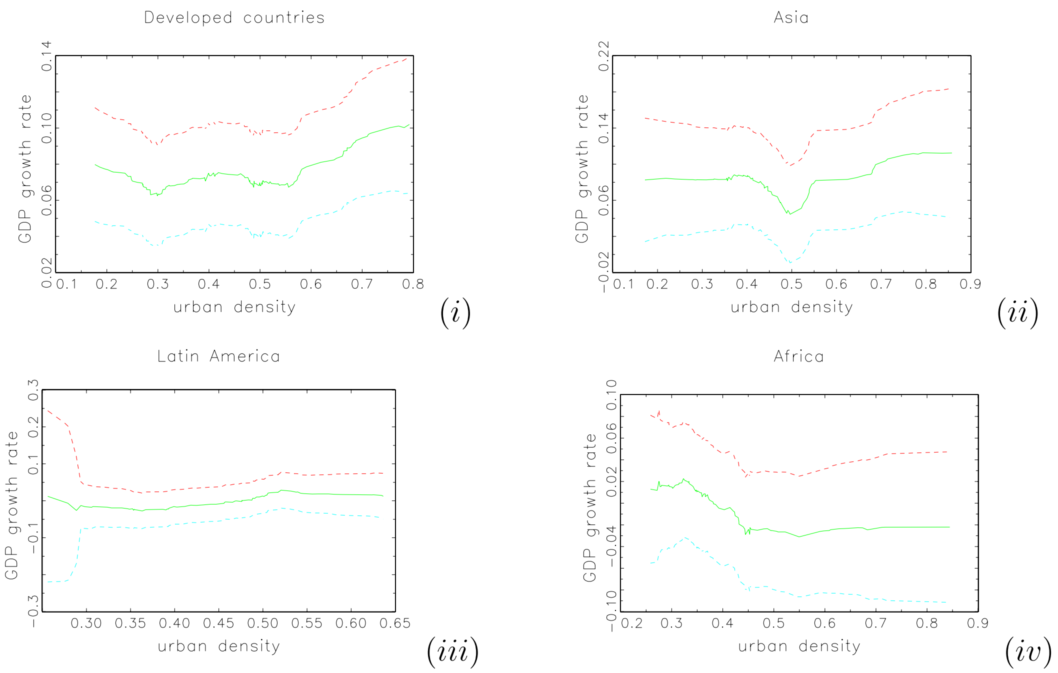

Figure 3.13: Non-parametric curves with urban density. IV procedure

Bibliography

[1] Ades, A. and E.L. Glaeser (1995) Trade and circuses: Explaining urban giants, *Quarterly Journal of Economics*, 110, 195-227.

[2] Alonso, W. (1980) Five Bell Shapes in Development, *Papers of the Regional Science Association*, 45, 5-16.

[3] Bairoch, P. (1985) *De Jericho à Mexico: Villes et économie dans l'histoire*. Paris: Gallimard. English translation *Cities and Economic Development. From the Dawn of History to the Present* (1988), London: Mansell Publishing Ltd.

[4] Barro, R. and J.-W. Lee (2001) International data on educational attainment: updates and implications, *Oxford Economic Papers* 3, 541-563.

[5] Bauwens, L. and H.K. Van Dijk (1990) Bayesian limited information analysis revisited. In: Gabszewicz, J.J., J.F Richard and L.A. Wolsey (Eds.) *Economic Decision Making: Games, Econometrics and Optimisation*, North-Holland, pp. 385-424.

[6] Bertinelli, L. and E. Strobl (2003) Urbanization, Urban Concentration and Economic Growth in Developing Countries, CORE Discussion Paper #2003/76.

[7] Black, D. and J.V. Henderson (1999) A Theory of Urban Growth, *Journal of Political Economy* 107, 252-284.

[8] Dareshwar, A. and V. Nehru (1993) A New Database on Physical Capital Stock: Sources, Methodology and Results, *Revista de Analisis Economico* 8, 37-59.

[9] Davis, J.C. and J.V. Henderson (2003) Evidence on the political economy of the urbanization process, *Journal of Urban Economics* 53, 98-125.

[10] Dreze, J.H. and J.F. Richard (1983) Bayesian Analysis of Simultaneous Equation Systems. In: Griliches Z., M.D. Intriligator (eds.) *Handbook of Econometrics*, North-Holland: Elsevier Science Publishers, pp. 517 - 598.

[11] Duranton, G. and D. Puga (2001) Nursery Cities, *American Economic Review* 91, 1457-1477.

[12] Durlauf, S. and D. Quah (1998) The New Empirics of Economic Growth, NBER Working Papers #6422.

[13] Eaton, J. and Z. Eckstein (1997) Cities and Growth: Theory and Evidence from France and Japan, *Regional Science & Urban Economics* 27, 443-474.

[14] El-Shakhs, S. (1972) Development, Primacy, and Systems of Cities, *Journal of Developing Areas* 7, 11-36.

[15] Fay, M. and C. Opal (2000) Urbanization without growth: A not so uncommon phenomenon, World Bank Working Paper #2412.

[16] Fujita, M., P.R. Krugman and A.J. Venables (1999) *The Spatial Economy: Cities, Regions, and International Trade*. Cambridge, MA: MIT Press.

[17] Grossman, G.M., and E. Helpman (1991) *Innovation and Growth in the Global Economy*. Cambridge, MA: MIT Press.

[18] Hall, R.E. and C.I. Jones (1999) Why Do Some Countries Produce So Much More Output Per Worker than Others, *Quarterly Journal of Economics* 83-116.

[19] Henderson, J.V. (1974) The Sizes and Types of Cities, *American Economic Review* 64, 640-656.

[20] Henderson, J.V., T. Lee, and Y.J. Lee (2001) Scale Externalities in Korea, *Journal of Urban Economics* 49, 479-504.

[21] Henderson, J. V. (2003) The Urbanization Process and Economic Growth:The So-What Question, *Journal of Economic Growth* 8, 47-71.

[22] Hoselitz, B. F. (1955) Generative and Parasitic Cities, *Economic Development and Cultural Change* 3, 278-294.

[23] Junius, K. (1999) Primacy and Economic Development: Bell shaped or parallel growth of cities? *Journal of Economic Development* 24, 1-22.

[24] King, G., J. Honaker, A. Joseph, and K. Scheve (2001) Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation, *American Political Science Review* 95, 49-69.

[25] Koop, G. and D.J. Poirier (2004) Bayesian variants of some classical semiparametric regression techniques, *Journal of Econometrics* 123, 259-282.

[26] Magrini, S. (2004) Regional (di)convergence. In: Henderson, J.V. and J.F. Thisse (eds.) *Handbook of Regional and Urban Economics, vol. 4*, North-Holland: Elsevier Science Publishers, pp. 2741-2796.

[27] Mazumdar, J. G. (1987) Rural-urban migration in developing countries. In: Mills, E.S. (Ed.) *Handbook of Regional and Urban Economics, vol. 2*, North-Holland: Elsevier Science Publishers, pp. 1097-1128.

[28] Mera, K. (1973) On the Urban Agglomeration and Economic Efficiency?, *Economic Development and Cultural Change* 21, 309-324.

[29] Moomaw, R.L. and A.M. Shatter (1996) Urbanization and economic development: a bias toward large cities?, *Journal of Urban Economics* 40, 13-37.

[30] Mutlu, S. (1989) Urban Concentration and Primacy Revisited: An Analysis and Some Policy Conclusions, *Economic Development and Cultural Change* 37, 611-639.

[31] Richardson, H.W. and G. Schwartz (1988) Economic Development, Population, and Primacy, *Regional Studies* 22, 467-475.

[32] Schafer, J.L. (1997) *Analysis of incomplete multivariate data*. Boca Raton: Chapman & Hall/CRC.

[33] Schimert, J., J.L. Schafer, T. Hesterberg, C. Fraley and D.B. Clarkson (2001) *Analysing data with missing values in S-Plus*. Seattle: Insightful.

[34] Temple, J. (1999) The New Growth Evidence, *Journal of Economic Literature* 37, 112-156.

[35] Todaro, M.P. (1969) A Model of Labor Migration and Urban Unemployment in Less Developed Countries, *American Economic Review* 59,138-148.

[36] Wheaton, W. and H. Shishido (1981) Urban Concentration, Agglomeration Economies, and the Level of Economic Development *Economic Development and Cultural Change* 30, 17-30.

[37] Williamson, J. G. (1965) Regional Inequality and the Process of National Development: A Description of the Patterns *Economic Development and Cultural Change* 13, 3-84.

[38] Williamson, J. G. (1987) Migration and Urbanization. In: Chenery, H. and T.N. Srinivasan (eds.) *Handbook of Development Economics, vol. 1*, North-Holland: Elsevier Science Publishers, pp. 424-465.

[39] Yatchew, A. (1998) Nonparametric Regression Techniques in Economics, *Journal of Economic Literature* 36, 669-721.

[40] Yatchew, A. (1999) An elementary nonparametric differencing test of equality of regression functions, *Economic Letters* 62, 271-278.

[41] Yatchew, A. (2003) *Semiparametric regression for the applied econometrician*. Cambridge: Cambridge university press.

[42] Yuki, K. (2007) Urbanization, informal sector, and development *Journal of Development Economics* 84, 76-103.

[43] Zellner, A., L. Bauwens and H.K. Van Dijk, (1988) Bayesian specification analysis and estimation of simultaneous equation models using Monte Carlo methods *Journal of Econometrics* 38, 39-72.

Chapter 4

Rent-seeking and urban primacy in Sub-Saharan Africa

"*It seems to be the thinking that Africans obtain an education in order to enter the government bureaucracy so as to be able to share the wealth of the nation rather than create wealth for the nation.*" (Bassey, 1999, p.106)

4.1 Introduction

There is a persistent debate about whether urbanization in developing countries favors economic development and growth. On the one hand, casual inspection of the available evidence suggests that, throughout history, cities played a lesser role in the economic development of the developing world than in that of now developed countries (Bairoch, 1988; De Long and Shleifer, 1993). This claim has been based on the observation that the urbanization process in most developing countries, and especially in sub-Saharan Africa (henceforth, SSA), seems to be one without much industrialization, without significant increases in agricultural productivity, but with a pronounced hypertrophia of the small tertiary sector, the government sector, and the shadow economy.[1] Put differently, urbanization does not seem

[0]This chapter has been realized in collaboration with Kristian Behrens. His contribution is gratefully acknowledged.

[1]According to World Bank data presented by Mbeki (2005, p.4), the percentage of the labor force working in the industrial sector in selected African countries changed as follows between 1970 and 1990: it stayed at 2% in Ethiopia; it increased from 5 to 7% in Kenya; it fell from 11 to 7% in Nigeria; and it rose from 9 to 16% in Gabon. At the same time, the corresponding changes in the service sector were all increasing: from 7 to 12% in Ethiopia; from 9 to 13% in Kenya, from 19 to 50% in Nigeria, and from 12 to 33% in Gabon. See Schneider (2004) for recent estimates of the size of African shadow economies.

to have allowed to reap the fruits of economic development and growth. On the other hand, one may argue that the real question to be asked in assessing the role of urbanization on growth is a counterfactual one: what would the current situation be like had the urbanization process not taken place? Several studies indeed suggest that the 'urbanization without industrialization' scenario has been exaggerated, that urbanization has led to some growth, even in SSA, and that it is not radically different from city growth in the past of developed nations (Williamson, 1988; Moomaw and Shatter, 1996).

Two fundamental, and highly intertwined, reasons for city growth in developing countries may be put forth: *politics and economics*. On the political side, the hypertrophia of the tertiary sector is likely to be driven by the concentration of political and economic power in cities. Many leaders in the developing world are quite sensitive to urban unrest, which threatens their political (and often physical) survival. The inflation of the tertiary sector, especially government administration, in a few big cities to accommodate the discontent and politically excluded part of the intellectual elite, is a convenient device for them to minimize the yearning for political changes. Ades and Glaeser (1995) claim that such political factors are the most important ones explaining urban centralisation and primacy in the developing world: politics directly affects urban concentration, because rent-seeking agents have to be spatially close to the political power. The rents they reap, together with the government's net transfer of resources from the country-side to the cities to run the state, further raise city population by attracting economic activities to the main centers of purchasing power.[2] The net result is that still more agents and activities are attracted by these transfers, which draws resources from the hinterland and makes cities in the developing world absorb a disproportionate fraction of overall economic activities. Eventually, the process becomes self-reinforcing and leads to the formation of very large urban cores as emphasized by the so-called 'new' economic geography (see Fujita *et al.*, 1999; Fujita and Thisse, 2002; Baldwin *et al.*, 2003).

The agglomeration process described in the above highlights that economical, political, and spatial factors should be jointly taken into consideration when trying to explain the formation of large urban agglomerations in the developing world. Stated differently, *political rent-seeking, rural-urban transfers, migration, increasing returns and geography* should be an important part of the whole story.[3] The main objective of this chapter is to present a new

[2]Similar mechanisms, leading to the formation of 'parasitic' cities due to the transfers of resources, have already been vividly depicted by Cantillon (1730, ch. I.V, in paragraph I.V.2) and Bairoch (1988). Note that, as shown by Dascher (2000), income transfers which violate Ohlson's principle of 'fiscal equivalence' also seem to play a role in explaining the growth of regional capitals in Germany.

[3]Increasing returns to scale are likely to be important in explaining agglomeration in SSA. Indeed,

economic geography (henceforth, NEG) framework that combines these ingredients to shed some light on agglomeration and urban primacy in developing countries. Although the NEG literature has been rapidly growing these last years there are, to the best of our knowledge, only few contributions dealing with political factors and issues specific to the developing world. Robert-Nicoud and Sbergami (2004) present a model that analyzes the impacts of political factors, increasing returns, and economic integration on agglomeration. Their main objective is to explain how and why the regional integration process of the European Union leads to important transfers of resources from the 'core regions' to the 'peripheral regions'.[4] Using a probabilistic voting model, the authors endogenize regional policy and transfers which are determined by the swing voters of the ideologically most homogeneous group. As the latter is predominantly located in the country-side, the peripheral regions can obtain transfers because of their relative political homogeneity. By contrast, the large region will keep the 'core' only if its relative economic size overcomes its political weakness due to ideological heterogeneity. Robert-Nicoud and Sbergami's (2004) main result, namely that the political process leads to a more even distribution of economic activities than the market mechanism, is quite opposite to what we seem to observe in many developing countries, namely transfers of resources from the hinterland to the cities, which increases agglomeration. Furthermore, a democratic process, as encapsulated in the probabilistic voting model, does not adequately characterize the political environment of most developing countries, especially in regions like SSA. We thus propose an alternative model in which the 'political process' consists in agents deciding on whether or not to enter a political elite in order to extract rents to maximize their own welfare. In such a setting, rent-seeking behavior leads to transfers from the country-side to the city, which shows that the nature of the political process matters for the direction of net transfers and the degree of agglomeration:

> "*When farmers form a majority of the population, they tend to subsidize the urban minority. When farmers form a minority, the urban majority subsidizes them.*"
> (Friedman, as quoted by Mbeki, 2005, p.5)

---

specialization of labor already prevailed in traditional cities of this region (see, e.g., Cocquery-Vidrovitch, 1993) and it is even more important in modern Black African cities, which experienced some industrialization and the development of a large informal service sector (Bairoch, 1988). Increasing returns and the importance of market size are, therefore, surely realistic features of Black African cities.

[4] The so-called 'Objective 1 regions' of the EU, defined as those with a per capita GDP of less than 75% of the EU average, benefit from Structural Cohesion Funds, which amount to a total of 18 billion Euro over the period 2000–2006 (for additional information see, e.g., `http://europa.eu.int/comm/regional_policy/objective1`).

Our model builds on the analytically solvable NEG model by Ottaviano and Forslid (2003) and features three types of agents: (i) immobile unskilled workers, who can choose to work in the formal sector or in the shadow economy; (ii) mobile skilled workers, who produce differentiated goods under monopolistic competition and increasing returns to scale; and (iii) an unproductive political elite, who taxes the other agents to maximize its own welfare. In accord with empirical evidence, this elite endogenously emerges from the skilled population when the expected rents are large enough to make it profitable.[5] Yet, the potential size of the elite constrains it in its tax choices, because an inflation of the elite leads to an erosion of the per capita tax shares and a reduction in the range of consumption goods through the diversion of productive resources (see, e.g., Krueger, 1974, for international trade considerations). Both features penalize the elite, since they reduce its income and the range of available consumption goods.[6]

Previewing our main results, we show that the presence of a tax-setting elite leads to interregional transfers of resources which raises the likelihood of agglomeration of production. Stated differently, agglomeration may take place because of rent-seeking, whereas the space-economy would be dispersed in the absence of such rent-seeking. In equilibrium, the unproductive elite will be larger the less differentiated the consumption goods are, and the higher the tax rates it sets are. The reason for the first effect is that when the returns to productive work are too low, rent-seeking constitutes a profitable alternative to production, whereas the second effect directly stems from higher rents. Focusing on the fully agglomerated equilibrium, we show that the equilibrium tax rate on skilled workers is zero, which allows to keep the size of the elite small. Rents are then extracted from the unskilled workers, provided that varieties are not too differentiated and that the expenditure share on manufacturing goods is small enough. When the latter two conditions are not met, there will be no elite formation in equilibrium since the returns to rent-seeking are too low when compared with the returns that can be secured in the productive sector.

The remainder of the chapter is organized as follows. Section 2 presents some stylized

---

[5]As Bassey (1999, p.3) put it: "the life chances of an individual in Africa for achieving political elite status are enormously enhanced if he or she belongs to, or can rise into, the upper level of the stratification system [...] education is about the only major determinant for moving into elite status in Africa." Higher education is at a premium in SSA since, according to Barro and Lee (2001, Table 3), it remains the region with the lowest percentage of highly educated people: 2.4% in 2000.

[6]Our model features 'love-of-variety', a characteristic of preferences often used in modeling developed economies. Yet, it is also appealing for investigating urbanization in the developing world since the fact that cities offer a larger assortment of consumption goods and services is one of the great incentives for rural-urban migration (Stahl, 1983, p.1).

facts about urban primacy, regional income transfers, and corruption. In Section 3, we develop the model and discuss the market outcome. Section 4 then investigates the spatial equilibrium and shows that rural-urban transfers make the emergence of agglomeration more likely. Section 5 deals with the issues of elite formation and tax setting, and derives the equilibrium taxes and the equilibrium size of the elite. Section 6 finally concludes.

4.2 Some stylized facts

To guide intuition, we start by isolating a few stylized facts on urban primacy based on a descriptive analysis of a 149 country sample.[7] This will allow us to establish results which suggest that rent-seeking and interregional income transfers are likely to be key drivers for urban primacy in SSA countries.

Table 4.1: Summary statistics (full sample)

Variable	Obs.	Min	Max	Mean	Std. dev.
% of urban population in the largest city	149	2.8	81	31.13	15.03
Ratio of largest to second-largest city	130	0	20.60	4.07	3.19
Food expenditure share	99	9.73	73.51	35.99	16.78
Non-corruption index	149	1.70	9.70	4.02	2.15
Top income tax rate	139	0	65.00	33.10	13.82
Top corporate tax rate	139	0	58.00	29.66	8.71
Size of the shadow economy (% GDP)	125	8.40	68.30	35.34	13.87
Central government expenditures	83	9	86	40.48	18.61
GDP per capita	133	91.06	48084.48	7096.11	11273.49

Table 4.1 provides some summary statistics for the full sample.[8] As one can see, there is a large degree of variation in all variables, including urban primacy as measured by either the share of the urban population living in the largest city or by the ratio of the largest to the second largest city.

[7]The full list of countries is provided in Appendix A, where we also give more detail on the different variables and data sources. Note that some variables are not available for some countries. In this case, we use smaller sample sizes.

[8]Note that the number of observations per variable varies due to missing observations.

Table 4.2: Comparing the full sample means with the SSA and OLM subsamples

Variable	Full sample mean	SSA subsample	OLM subsample
% of urban population in the largest city	31.13	34.97	31.39
Ratio of largest to second-largest city	4.07	4.31	4.43
Food expenditure share	35.99	47.34	46.22
Non-corruption index	4.02	2.78	2.94
Top income tax rate	33.10	39.52	29.39
Top corporate tax rate	29.66	34.15	29.49
Size of the shadow economy (% GDP)	35.34	43.70	40.10
Central government expenditures	40.48	60.27	45.23

Table 4.2 contrasts the full sample with the SSA subsample. At first sight, it suggests that there is a pattern specific to SSA countries. Indeed, these countries seem to have larger primate cities (as measured either by the urban population in the largest city or the ratio of the largest to the second-largest city), seem to be more corrupt, have a larger food expenditure share, and a bigger shadow economy than the remaining countries of the sample. Furthermore, top tax rates, both for personal and for corporate income, also seem to be higher in those countries. To get a clearer idea on the SSA specificity we also compare, in Table 4.2, the SSA subsample with the subsample of 'other low and lower middle income countries' (OLM) countries.[9] We see that the SSA and OLM samples are roughly similar in terms of the non-corruption index, the food expenditure shares, and the size of the shadow economies. Yet, SSA countries have higher top tax rates, higher central government expenditures on goods, services, and compensation of employees, and larger shadow economies. They are also characterized by a higher share of the urban population living in the largest city, though the gap is relatively small.

Since the full sample and the OLM subsample include countries from different continents (and also at various development stages), there is a great deal of heterogeneity. This prompts for a more detailed comparison between SSA and other geographical areas. According to the mean comparison test summarized in Table 4.3, SSA countries have a significantly larger urban primacy than European countries. However, *there is no significant difference, at the 5% significance level, between SSA and other developing regions in terms of urban primacy.* Put differently, urban primacy is not a criteria according to which SSA is

[9]The classification for OLM countries is that from the World Bank.

Table 4.3: Mean comparison tests between SSA and other regions

SSA versus	Mean comparison test	p-value
	% of urban population in the largest city	
Europe	$\mu_{SSA} > \mu_{EU}$	0.0005
Latin America	$\mu_{SSA} \neq \mu_{LA}$	0.3737
Asia	$\mu_{SSA} \neq \mu_{AS}$	0.1968
North Africa	$\mu_{SSA} \neq \mu_{NA}$	0.2032
SSA versus	Mean comparison test	p-value
	Non-corruption index	
Europe	$\mu_{SSA} < \mu_{EU}$	0.0000
Latin America	$\mu_{SSA} < \mu_{LA}$	0.0139
Asia	$\mu_{SSA} < \mu_{AS}$	0.0134
North Africa	$\mu_{SSA} < \mu_{NA}$	0.0466
SSA versus	Mean comparison test	p-value
	Central government expenditures	
Europe	$\mu_{SSA} > \mu_{EU}$	0.0000
Latin America	$\mu_{SSA} > \mu_{LA}$	0.0025
Asia	$\mu_{SSA} > \mu_{AS}$	0.0473
North Africa	$\mu_{SSA} \neq \mu_{NA}$	0.0872

likely to differ significantly from other developing regions. Yet, the mean comparison tests on the non-corruption index and on central government expenditures on goods, services and compensation of employees, reveal that SSA differs in other fundamental respects from the remaining developing countries in the sample. Indeed, given the significantly lower non-corruption index mean than in any other region of the world, SSA stands out as the region facing the highest degree of corruption. Furthermore, central government expenditures on goods, services and compensation of employees are significantly higher in SSA than in any other region of the world, expect North-Africa. If these expenditures are mostly financed by general taxes, yet spent locally in large cities under the form of administrative payrolls, they would lead to more rent-seeking and imply a regional redistribution of purchasing power that may drive the agglomeration of mobile firms and workers, thereby leading to the formation of very large urban centres.[10]

[10]Unfortunately, we have no information on how and where tax revenues are levied and on how and where they are ultimately spent in developing countries.

The results of the mean comparison tests are quite consistent with the results given by an analysis of the pairwise correlation matrix. As can be seen from Table 4.4, the non-corruption index and GDP per capita are quite naturally strongly and positively correlated, thus implying that highly corrupt countries rank among the poorest in the world. Furthermore, GDP per capita is significantly negatively correlated with central government expenditures on goods, services, and compensation of employees. In other words, countries with high values of central government expenditures are, therefore, among those with the weakest economies.

Table 4.4: Pairwise correlation matrix

	(1)	(2)	(3)	(4)	(5)	(6)
(1) % of urban population in largest city	1.00					
(2) Central government expenditures	0.4726*	1.00				
(3) Food expenditure share	0.1045	0.4712*	1.00			
(4) Non-corruption index	-0.1591*	-0.3989*	-0.7826*	1.00		
(5) Size of the shadow economy	0.2647*	0.2966*	0.6179*	-0.6969*	1.00	
(6) GDP per capita	-0.2110*	-0.5272*	-0.7408*	0.8785*	-0.6737*	1.00

Note: Correlations with * are significant at the 5% level.

Table 4.4 and Figure 1 reveal that urban primacy is significantly correlated with central government expenditures. Furthermore, as can be seen from Table 4.4, it is also significantly correlated with the non-corruption index. Put differently, corrupt countries with large central government expenditures on goods, services, and compensation of employees, generally have larger primate cities than the remaining countries. Hence, even if there is no specificity in observed urban primacy in SSA, this primacy may stem from other factors than in the rest of the developing world: high corruption which leads to a redistribution of purchasing power from the country-side to the cities, where it is spent by the elite. With the highest corruption level and the largest central government expenditures on goods, services, and compensation of employees, SSA may be the developing region where politics has the strongest impacts on regional imbalances and urban primacy.

4.3 The model

We now present an economic geography model that formalizes the interactions between the existence of an elite (corruption), the redistribution of purchasing power across regions

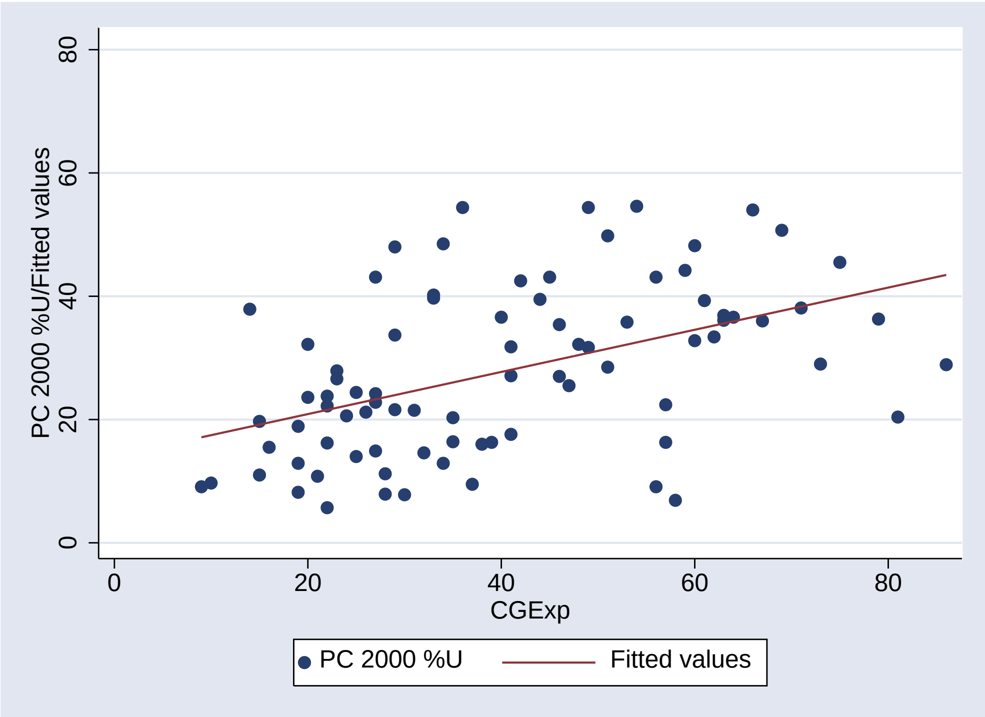

Figure 4.1: Central government expenditures (`CGExp`) and urban primacy (`PC2000%U`)

(interregional income transfers), and agglomeration (urban primacy). To do so, we build on the analytically solvable model of Ottaviano and Forslid (2003). We consider a country with two regions, labeled 1 and 2, respectively. Variables associated with each region will be subscripted accordingly. The population consists of an exogenously given mass $\overline{U}$ of unskilled and $\overline{S}$ of skilled workers, respectively. The skilled are geographically mobile, and they may be either production workers or part of an *unproductive elite*. The unskilled are geographically immobile, and always work in the production sector. Yet, contrary to the skilled workers, they may choose between working in the formal and in the informal sector (e.g., because they are harder to monitor and to tax than the skilled workers).[11] For simplicity, we assume that the unskilled are evenly distributed across the two regions, each of which accommodates a mass $\overline{U}/2$ of them. Stated differently, no region has an initial advantage in terms of the size of its immobile demand. All agents spend their incomes

[11]We assume that the unskilled cannot become skilled. Adding such a possibility is formally equivalent to endogenizing the wage elasticity of labor supply from agriculture to manufacting, as in Puga (1998). This in turn is equivalent to allowing the geographically mobile population to increase, which would only reinforce our agglomeration results. We also assume, for simplicity, that the total population is fixed. Yet, one should keep in mind that strong population growth and the associated pressure on scarce resources in the countryside are important factors which explain why large streams of migrants are 'pushed' from rural areas to cities. Adding such considerations would, again, reinforce our results.

locally only and work in their region of residence. Skilled workers are mobile across regions ('geographical mobility') and they may become part of the political elite ('social mobility'). In what follows, we denote by S the mass of skilled workers in the productive sector and by E the mass of unproductive skilled constituting the political elite. Note that both S and E are endogenously determined, with $S + E = \overline{S}$. Furthermore, we denote by $\lambda \in [0, 1]$ the share of productive skilled S living in region 1. We assume that the political elite is clustered into a historically determined center of power, which we henceforth refer to as the *capital* of the country (e.g., the historical capital).[12] Without loss of generality, we assume that region 1 is the capital. Our model may be viewed as a game with four stages:

1. the elite sets the tax rates t^U and t^S on unskilled and skilled;
2. skilled workers decide whether to enter the elite or not;
3. skilled production workers choose the region they live and work in;
4. firms maximize profits, and production and consumption takes place.

We solve the game by backward induction.[13]

4.3.1 Preferences

A representative consumer in region $i = 1, 2$ (whether he belongs to the skilled, to the unskilled, or to the elite) has Cobb-Douglas upper-tier preferences over agricultural and manufactured goods, with a CES sub-utility over a continuum of horizontally differentiated varieties. Formally, he solves the following consumption problem:

$$\max_{A_i,\, q_{ji}(\omega)} \quad A_i^{1-\mu} \left(\int_{\Omega_i} q_{ii}(\omega)^{\frac{\sigma-1}{\sigma}} \mathrm{d}\omega + \int_{\Omega_j} q_{ji}(\omega)^{\frac{\sigma-1}{\sigma}} \mathrm{d}\omega \right)^{\frac{\mu\sigma}{\sigma-1}}$$
$$\text{s.t.} \quad p^A A_i + \int_{\Omega_i} p_{ii}(\omega) q_{ii}(\omega) \mathrm{d}\omega + \int_{\Omega_j} p_{ji}(\omega) q_{ji}(\omega) \mathrm{d}\omega = y_i$$

where A_i is the consumption of agricultural good; $q_{ji}(\omega)$ and $p_{ji}(\omega)$ stand for the consumption and the price of variety ω in country i when it is produced in country j; Ω_i stands for

[12] We do not attempt to endogenously determine where this center is located. Although this is an interesting question, it is secondary to the aspects we are interested in.

[13] As always, timing of the stages is crucial to the game. One may, e.g., consider the case in which tax setting and the formation of the elite take place simultaneously. Yet, we believe that our timing is relevant since for agents to decide on tax rates they must already have chosen to be part of the elite. Furthermore, since our game is one-shot, we disregard the fact that the elite wants to remain elite in the long run, which would necessitate an analysis within a repeated games framework.

the set of varieties produced in country i, with measure n_i; y_i stands for the agent's income, which depends on the social group he belongs to (skilled, unskilled, or elite); $0 < \mu < 1$, and $\sigma > 1$ are parameters; and $p_i(\omega)$ is the price of variety ω in country i. Since preferences are homothetic, we obtain the following aggregate demand for firm ω in region i when it is located in region j:

$$D_{ji}(\omega) = \frac{p_{ji}(\omega)^{-\sigma}}{\mathbb{P}_i^{1-\sigma}} \mu Y_i \tag{4.1}$$

where Y_i is the total income of agents in region $i = 1, 2$ and $\mathbb{P}_i$ is the CES price aggregate. We assume that all varieties produced in each region are symmetric, which allows us to alleviate notation by dropping the variety index ω. The price aggregate $\mathbb{P}_i$ then reduces to

$$\mathbb{P}_i = \left(n_i p_{ii}^{1-\sigma} + n_j p_{ji}^{1-\sigma} \right)^{\frac{1}{1-\sigma}}. \tag{4.2}$$

4.3.2 Technology, taxes, and transportation

There are two factors of production, skilled and unskilled labor, and two sectors, manufacturing and agriculture. The agricultural sector produces a homogeneous good using unskilled labor only. We assume that this good is costlessly tradable across regions and we normalize, without loss of generality, the unit input coefficient in this sector to one. Perfect competition and costless trade then imply that the unskilled wages w^U are equalized across regions: $w_1^U = w_2^U = p^U = 1$, where the last equality comes from our choice of numéraire.[14] All unskilled workers are a priori free to work in the agricultural sector, and this sector is informal (untaxed) because hard to monitor.

Skilled workers face the choice to remain in the productive sector to earn a wage w_i there, or to get involved in politics and become part of the elite. The benefit of belonging to the elite is to participate in running the country. Among other things, we assume that it is the elite which determines the tax rates the different groups of agents face. To keep things simple, we suppose that the elite levies a *proportional income tax rate* t^S (resp., t^U) on the incomes of skilled (resp., of unskilled) workers. [15] While the skilled can evade taxation only by becoming part of the elite, the unskilled face the choice of working either in the

[14] Strictly speaking, such factor price equalization only holds when the mass of unskilled workers is large enough for some agricultural production to take place in both regions in equilibrium. In what follows, we assume that this condition holds. Formally, $\mu < \sigma/(2\sigma - 1)$ (see Forslid and Ottaviano, 2003).

[15] Since tax revenue is spent unproductively, we may interpret $1 - t^S$ and $1 - t^U$ as proxies for the 'degree of property rights enforcement' for skilled and unskilled workers, respectively (Ades, 1995).

formal manufacturing sector (hence paying taxes), or of working in the agricultural shadow economy (hence evading taxation). Put differently, the unskilled tax base is not perfectly inelastic but generally shrinks with the level t^U of taxation. This observation fits well with the empirical fact that all African countries have large shadow economies (i.e., fiscal evasion seems to be fairly easy due to the lack of enforcement). Recent estimates using a sample of 37 African countries indeed reveal that the mean average size of the shadow economy is about 43.2% of GDP (Schneider, 2004, Figure 3.1.1).

Manufacturing firms produce varieties of a horizontally differentiated consumption good using both skilled and unskilled labor. More precisely, each firm requires F units of skilled labor as a fixed input requirement and m units of unskilled labor per unit of output as a variable input requirement. The unskilled are willing to work in the formal sector if and only if their net income is equal to the one they can secure in the informal sector, which implies that firms have to pay $1/(1-t^U)$ per unit of unskilled labor. Put differently, taxing the unskilled raises firms' production costs because workers have the outside option of going into the shadow economy. Total production costs for producing a quantity Q in region $i = 1, 2$ are then given by

$$\mathrm{TC}_i(Q) = \frac{m}{1-t^U}Q + Fw_i,$$

where w_i stands for the skilled wage in region i. Given the fixed cost requirement, skilled labor market clearing then requires that the masses of firms in the two regions are as follows:

$$n_1 = \frac{\lambda S}{F} \qquad \text{and} \qquad n_2 = \frac{(1-\lambda)S}{F}. \tag{4.3}$$

To ship one unit of any variety between the two regions entails an iceberg trade cost of $\tau > 1$. Taking into account this resource waste effect, the profit of a representative firm in region i is given by

$$\Pi_i = \left(p_{ii} - \frac{m}{1-t^U}\right)D_{ii} + \left(p_{ij} - \frac{m\tau}{1-t^U}\right)D_{ij} - Fw_i,$$

where the demands are evaluated at (4.1). Because firms face an isoelastic demand, the profit-maximizing prices display a constant mark-up over marginal cost:

$$p_{ii}^* = \frac{\sigma m}{(\sigma-1)(1-t^U)} \qquad \text{and} \qquad p_{ij}^* = \frac{\sigma m\tau}{(\sigma-1)(1-t^U)}.$$

Substituting these prices into (4.2), letting $\phi \equiv \tau^{1-\sigma}$ stand for the freeness of trade between the two regions, and using the skilled labor market clearing conditions (4.3), the price indices can be expressed as follows:

$$\begin{aligned}
\mathbb{P}_1 &= \frac{\sigma m}{(\sigma-1)(1-t^U)}\left(\frac{S}{F}\right)^{\frac{1}{1-\sigma}}\left[\lambda+(1-\lambda)\phi\right]^{\frac{1}{1-\sigma}} \\
\mathbb{P}_2 &= \frac{\sigma m}{(\sigma-1)(1-t^U)}\left(\frac{S}{F}\right)^{\frac{1}{1-\sigma}}\left[\lambda\phi+(1-\lambda)\right]^{\frac{1}{1-\sigma}}
\end{aligned} \tag{4.4}$$

Note that, contrary to other NEG models (Krugman, 1991; Ottaviano *et al.*, 2002) in which the total mass of varieties is proportional to the exogenously fixed skilled population S, there are two distinct 'price index effects' in our model:

- for any given value of S, the price index in a region decreases with the share of firms located in that region ('regional market crowding effect'); and
- for any given distribution of firms, *the price indices in both regions decrease with the mass S of productive skilled workers* ('global market crowding effect').

This second effect, which implies that a smaller mass of productive skilled workers decreases welfare by reducing variety and increasing consumer prices, will be important in the subsequent analysis of the elite's behavior.

Product market clearing for each variety requires that a firm located in region i produces the total quantity

$$X_i = D_{ii} + \tau D_{ij} = \frac{\mu(\sigma-1)(1-t^U)}{m\sigma}\left(\frac{Y_i}{n_i+\phi n_j}+\frac{\phi Y_j}{\phi n_i+n_j}\right). \tag{4.5}$$

Since firms price above marginal cost, there exist pure operating profits which are competed away by firms' bidding for skilled labor. Therefore, in equilibrium the skilled wages absorb all operating profits:

$$w_i = \frac{mX_i}{F(\sigma-1)(1-t^U)} = \frac{\mu}{\sigma F}\left(\frac{Y_i}{n_i+\phi n_j}+\frac{\phi Y_j}{\phi n_i+n_j}\right), \tag{4.6}$$

which using (4.3), can finally be expressed as follows:

$$w_1 = \frac{\mu}{\sigma S}\left[\frac{Y_1}{\lambda+\phi(1-\lambda)}+\frac{\phi Y_2}{\phi\lambda+(1-\lambda)}\right] \tag{4.7}$$

$$w_2 = \frac{\mu}{\sigma S}\left[\frac{\phi Y_1}{\lambda+\phi(1-\lambda)}+\frac{Y_2}{\phi\lambda+(1-\lambda)}\right]. \tag{4.8}$$

4.3.3 Rent-seeking and elite formation

As stated before, skilled workers' only way to evade taxation is to become a member of the elite. When they do so, we assume that the skilled spend their time in an 'unproductive' way, which de facto reduces the amount of skilled labor available for the production of differentiated varieties. Although this is a strong assumption, it partly captures the fact that in many African countries political participation essentially stems from a rent-seeking motive, which is time intensive (e.g., because of lobbying) and therefore reduces the productive labor supply of agents involved in this kind of activities.[16] A skilled worker will become a member of the elite if and only if the rent he can secure from doing so exceeds the skilled wage he can secure in the productive sector.[17] Formally, a skilled worker will enter the elite if his rent r exceeds $(1-t^S)w_i$, where the rent satisfies

$$\begin{aligned} rE &\equiv m[n_1X_1+n_2X_2]\frac{t^U}{1-t^U}+S\left[\lambda w_1+(1-\lambda)w_2\right]t^S \\ &= S\left[(\sigma-1)t^U+t^S\right]\left[\lambda w_1+(1-\lambda)w_2\right], \end{aligned} \tag{4.9}$$

the right-hand side being the elite's total tax revenue. Note that we assume, for simplicity, that the elite *does spend all the tax revenue unproductively.* This captures the idea that "in those African countries in which corruption has become quite pervasive, the cost of public goods and services is highly inflated, usually to provide additional income for the individuals whose job it is to serve the public" (Mbaku, 2003, p.317). Adding some partial tax spending on the provision of public goods and services is not likely to weaken our results if this "social overhead" is largely allocated to cities and financed from general taxes instead of urban ones.

4.3.4 Market outcome

We first analyze the market outcome for any *given* allocation of skilled between the production sector and the elite, and for any given spatial distribution λ of skilled production workers across regions. Furthermore, t^U and t^S are considered as fixed at this stage.

Since the elite is, by assumption, fully agglomerated in region 1, the net aggregate

[16] We focus on unproductive rent-seeking only. See, e.g., Baumol (1990) for historical case studies of periods during which rent-seeking was even 'destructive'.

[17] There is a long-standing tradition which assumes that there are some fixed entry costs for becoming a member of the elite (see, e.g., Ades, 1995, for further references). For simplicity, we disregard these costs in this chapter. Adding them would reinforce our results but make the algebra more involved.

incomes accruing to the agents in both regions are given by

$$\begin{aligned} Y_1 &= \frac{\overline{U}}{2} + \lambda S w_1 (1 - t^S) + Er \\ &= \frac{\overline{U}}{2} + \lambda S w_1 (1 - t^S) + S\left[(\sigma - 1)t^U + t^S\right][\lambda w_1 + (1 - \lambda) w_2] \qquad (4.10) \\ Y_2 &= \frac{\overline{U}}{2} + (1 - \lambda) S w_2 (1 - t^S). \qquad (4.11) \end{aligned}$$

The market outcome is a solution to the four equations (4.7), (4.8), (4.10) and (4.11) in the four unknowns w_1, w_2, Y_1 and Y_2. The unique solution in w_1 and w_2 to this linear system is given by:[18]

$$\begin{aligned} w_1^* &= \frac{(1-\lambda)(\mu+\sigma)\phi^2 + 2\lambda\sigma\phi + (1-\lambda)(\sigma-\mu) + (1-\lambda)\mu\left(1-\phi^2\right)\left(2t_S + (\sigma-1)t_U\right)}{D(t^S, t^U)} \mu\bar{U} \\ w_2^* &= \frac{2\sigma\phi + \lambda(1-\phi)(\sigma - \mu - (\mu+\sigma)\phi) - \lambda\mu(\sigma-1)\left(1-\phi^2\right)t_U}{D(t^S, t^U)} \mu\bar{U}, \end{aligned}$$

where

$$\begin{aligned} D(t^S, t^U) \equiv\; & 2S[(1-\lambda)\lambda(\mu+\sigma)\phi^2 + (2(\lambda-1)\lambda+1)\sigma\phi + (\lambda-1)\lambda(\mu-\sigma) \\ & +(\lambda-1)\lambda\mu\left(\phi^2-1\right)t_S]\left(\sigma - \mu - \mu(\sigma-1)t_U\right) > 0. \end{aligned}$$

It is readily verified that the expressions of w_1^* and w_2^* reduce to the ones in Forslid and Ottaviano (2003) in the no-tax case ($t^U = t^S = 0$).

Note that both w_1^* and w_2^* are *decreasing* in S, which will itself be endogenously determined later in our analysis. The reason for this is that more skilled production workers increase the mass of competing firms, which leads to global product market crowding and, therefore, lower equilibrium wages. Such an effect does not arise in standard NEG models where the mass of firms is usually proportional to the exogenously fixed mass of skilled workers (Krugman, 1991; Ottaviano *et al.*, 2002). Furthermore, straightforward but longer calculations show that w_2^* is decreasing in t^S, whereas w_1^* is increasing in t^S. Stated differently, increasing taxation of the skilled shifts nominal wages in favor of the capital region and away from the periphery. The reason is that as t^S increases, the elite spends proportionally more tax revenues on varieties produced in the capital region, thereby raising demand and wages there. As we show later, the widening interregional wage gap induced by taxation increases the tendency for agglomeration of the mobile sector.

[18]Because Y_1^* and Y_2^* are complicated terms that are not required for the subsequent analysis, we do not provide their analytical expressions in this chapter.

4.4 Spatial equilibrium

As usual in NEG models, we assume that the mobile skilled workers employed in the productive sector migrate to the region offering them the highest indirect utility. Let ΔV^* stand for the indirect utility differential between region 1 and 2, which is a function of tax rates and given as follows:

$$\Delta V^*(t^S, t^U) = \mu^\mu (1-\mu)^{1-\mu}(1-t^S)\left(\frac{w_1^*}{\mathbb{P}_1^\mu} - \frac{w_2^*}{\mathbb{P}_2^\mu}\right).$$

Some straightforward calculations, using the equilibrium expressions of w_1^*, w_2^*, and the two regional price indices (5.6), allow us to rewrite ΔV^* as follows:

$$\begin{aligned}\frac{\Delta V^*(t^S, t^U)}{K(t^S, t^U)} &= \frac{(1-\lambda)(\mu+\sigma)\phi^2 + 2\lambda\sigma\phi + (1-\lambda)(\sigma-\mu) + (1-\lambda)\mu\left(1-\phi^2\right)(2t_S + (\sigma-1)t_U)}{[\lambda + \phi(1-\lambda)]^{\frac{\mu}{1-\sigma}}} \\ &- \frac{2\sigma\phi + \lambda(\phi-1)(\mu-\sigma+(\mu+\sigma)\phi) + \lambda\mu(\sigma-1)\left(\phi^2-1\right)t_U}{[\phi\lambda + (1-\lambda)]^{\frac{\mu}{1-\sigma}}} \end{aligned} \tag{4.12}$$

where

$$K(t^S, t^U) \equiv \frac{(1-t^S)\bar{U}(1-\mu)^{1-\mu}\mu^{\mu+1}}{D(t^S, t^U)}\left[\frac{m\left(\frac{S}{F}\right)^{\frac{1}{1-\sigma}}\sigma}{(1-t^U)(\sigma-1)}\right]^{-\mu} \tag{4.13}$$

is a strictly positive bundle of parameters, parametrized by the tax rates t^S and t^U.

A *spatial equilibrium* is such that no skilled worker has an incentive to change location, conditional upon the fact that the product markets clear at the equilibrium prices, while the skilled labor market clears at the equilibrium wages. Formally, a spatial equilibrium arises at $\lambda^* \in (0,1)$ when $\Delta V^*(\lambda^*) = 0$, or at $\lambda^* = 0$ if $\Delta V^*(0) \leq 0$, or at $\lambda^* = 1$ if $\Delta V^*(1) \geq 0$. Following Fujita *et al.* (1999), an interior equilibrium is said to be stable if and only if the slope of the indirect utility differential ΔV^* is negative in a neighborhood of the equilibrium, whereas the two agglomerated equilibria are always stable whenever they exist.

4.4.1 The benchmark case

Let us start with the case where $t^S = t^U = 0$, which is the benchmark case of Forslid and Ottaviano (2003). In this case, ΔV^* reduces to

$$\begin{aligned}\frac{\Delta V^*(0,0)}{K(0,0)} &= \frac{2\lambda\sigma\phi + (1-\lambda)\left[\sigma\left(\phi^2+1\right) - \mu\left(1-\phi^2\right)\right]}{[(1-\lambda)\phi + \lambda]^{\frac{\mu}{1-\sigma}}} \\ &\quad - \frac{2(1-\lambda)\sigma\phi + \lambda\left[(\mu+\sigma)\phi^2 + (\sigma-\mu)\right]}{[\lambda\phi + (1-\lambda)]^{\frac{\mu}{1-\sigma}}}. \end{aligned} \tag{4.14}$$

As shown by Forslid and Ottaviano (2003), full agglomeration may be sustained as an equilibrium if and only if

$$\frac{\Delta V^*(0,0)}{K(0,0)}\Big|_{\lambda=1} = -\frac{\Delta V^*(0,0)}{K(0,0)}\Big|_{\lambda=0} = 2\sigma\phi - \frac{(\mu+\sigma)\phi^2 + (\sigma-\mu)}{\phi^{\frac{\mu}{1-\sigma}}} > 0, \tag{4.15}$$

which implicitly defines the sustain point ϕ^s as the value of ϕ that equates the above expression to zero. Full agglomeration can be sustained for all $\phi \geq \phi^s$. Additionally, there are at most three interior equilibria in the no-tax case (Robert-Nicoud, 2005), of which the symmetric one ($\lambda^* = 1/2$) always exists. The stability of the equilibrium $\lambda^* = 1/2$ depends on the sign of the derivative of the indirect utility differential, whereas the other two interior equilibria are always unstable. Computing $\partial(\Delta V^*)/\partial\lambda$ and evaluating it at $\lambda = 1/2$, the break-point is such that

$$\phi^b \equiv \frac{\sigma-\mu}{\sigma+\mu}\frac{\mu-\sigma+1}{1-\mu-\sigma}.$$

Hence, $\lambda^* = 1/2$ is a stable spatial equilibrium for all $\phi \leq \phi^b$. Note, finally, that both types of equilibria occur for values $\phi^s \leq \phi \leq \phi^b$, in which case both full agglomeration and full dispersion are stable spatial equilibria.

In what follows, we assume that agglomeration forces are sufficiently weak for dispersion to prevail as an equilibrium outcome for some parameter values. In other words, we assume that the 'no-black-hole' condition $\mu < \sigma - 1$ holds.

4.4.2 The tax case

Assume now that the elite levies taxes ($t^U > 0$ and/or $t^S > 0$). In this case, some longer calculations show that

$$\begin{aligned}\frac{\Delta V^*(t^S,t^U)}{K(t^S,t^U)} &= \frac{\Delta V^*(0,0)}{K(0,0)} \\ &+\mu\left(1-\phi^2\right)\left[\lambda\frac{t^U(\sigma-1)}{[\lambda\phi+(1-\lambda)]^{\frac{\mu}{1-\sigma}}} + (1-\lambda)\frac{2t^S+t^U(\sigma-1)}{[\lambda+\phi(1-\lambda)]^{\frac{\mu}{1-\sigma}}}\right],\end{aligned} \tag{4.16}$$

where the first term on the right-hand side corresponds to Ottaviano and Forslid's (2003) indirect utility differential, as given by (4.14). Since the second term on the right-hand side of (4.16) is unambiguously positive for all $t^U > 0$ and/or $t^S > 0$, and since what matters for the spatial equilibrium is the sign of ΔV^*, we have the following result.

Proposition 1 (taxation and agglomeration) *When compared with the no-tax case,*

1. *the capital hosts a larger share of the mobile skilled population under positive taxes, i.e.,* $\lambda^*(t^S, t^U) \geq \lambda^*(0,0)$ *when* $(t^S, t^U) > 0$*;*

2. *full agglomeration in the capital is more likely under positive taxes, i.e.,* $\phi^s(t^S, t^U) < \phi^s(0,0)$ *when* $(t^S, t^U) > 0$*.*

Proof. In both cases, let λ_0^* denote the value of λ such that $\Delta V^*(0,0) \mid_{\lambda=\lambda_0^*} = 0$. Evaluating (4.16) at this value yields

$$\frac{\Delta V^*(t^S, t^U)}{K(t^S)} \Big|_{\lambda=\lambda_0^*} = \mu \left(1-\phi^2\right) \left\{ \lambda_0^* \frac{t^U(\sigma-1)}{[\lambda_0^*\phi + (1-\lambda_0^*)]^{\frac{\mu}{1-\sigma}}} + (1-\lambda_0^*) \frac{2t^S + t^U(\sigma-1)}{[\lambda_0^* + \phi(1-\lambda_0^*)]^{\frac{\mu}{1-\sigma}}} \right\}$$

which is strictly positive. This shows that: (i) when there is a dispersed equilibrium without taxes, the equilibrium with taxes must involve more agglomeration since the utility differential at that no-tax equilibrium is strictly positive; and (ii) when there is full agglomeration without taxes, there is also full agglomeration with taxes since

$$\frac{\Delta V^*(t^S, t^U)}{K(t^S)} \Big|_{\lambda=1} = \frac{\Delta V^*(0,0)}{K(0,0)} \Big|_{\lambda=1} + \mu(1-\phi^2) t^U (\sigma-1) \phi^{\frac{\mu}{\sigma-1}} \tag{4.17}$$

is strictly positive because $\Delta V^*(0,0)/K(0) \mid_{\lambda=1} > 0$ (since full agglomeration is sustainable without taxes). Finally, since the second term in (4.17) is strictly positive we see that: (iii) there are parameter values for which full agglomeration is sustainable with taxation, whereas it is not in the absence of taxation.

When taken together, conditions (i), (ii), and (iii) show that taxation cannot make the economy less agglomerated in the capital, which establishes Proposition 1. ■

Proposition 1 shows that taxation increases agglomeration in the capital. This is because of the redistribution of purchasing power to the capital ('urban bias'), which entices firms and agents to locate there: "*The political power [induces] the government to transfer resources to the capital, and these transfers will attract migrants*" (Ades and Glaeser, 1995, p.199).[19]

Before proceeding, a few remarks are in order. First, although taxation by the elite increases agglomeration, we cannot compute the (partially agglomerated) stable interior equilibrium analytically. Indeed, the indirect utility differential is a transcendental function, which does not allow for explicit solutions in λ in the general case. By consequence, it is also impossible to explicitly characterize the break point for $\lambda \neq 1/2$, since this requires the evaluation of the sign of the derivative of ΔV^* at the interior equilibrium, which we do

[19]Note that the conditions for full agglomeration in the other region are no longer symmetric, due to the asymmetries in regional spending induced by the presence of the elite (see also Forslid and Ottaviano, 2003, pp. 237-239).

not know. In this respect, our analysis faces similar problems than the ones encountered by Forslid and Ottaviano (2003) and Baldwin and Krugman (2004). Yet, these authors have shown that exogenous size differences or differential taxation favor the degree of agglomeration in the region having the larger market (in terms of consumption expenditure), and this result continues to hold true in our model.

Second, contrary to Forslid and Ottaviano (2003) and Baldwin and Krugman (2004), the value of S will be endogenously determined by the process of elite formation in our model. Because what matters in the CES model are the expenditure share μ and trade costs τ, the size of the elite is not of fundamental importance for the spatial equilibrium, despite its endogeneity (see expression (4.12)). Yet, one should keep in mind that the size of the elite may influence tax-setting, which has itself an impact on the spatial equilibrium which is the focus of our analysis. Furthermore, the spatial equilibrium would depend on S if skilled workers and members of the elite had different expenditure shares for differentiated goods.[20]

Last, note that the degree of agglomeration increase monotonically with the taxation of the unskilled:

$$\frac{\partial\left(\Delta V^{*}/K\right)}{\partial t^{U}} = \mu(\sigma-1)(1-\phi^{2})\left\{\frac{1-\lambda}{[(1-\lambda)\phi+\lambda]^{\frac{\mu}{1-\sigma}}}+\frac{\lambda}{[\lambda\phi+(1-\lambda)]^{\frac{\mu}{1-\sigma}}}\right\}>0.$$

The reason is, as in the foregoing, the higher elite expenditure in the capital which attracts mobile firms and agents.

4.5 Elite formation and taxation

In this section we focus on the dual issues of *elite formation* and *tax setting* by this elite.

4.5.1 Elite formation

Only the skilled have the opportunity of entering the political elite of the country. Each skilled worker therefore faces the binary choice of: (i) staying in the productive sector, earning an after-tax wage of $w_i(1-t^S)$ when he is located in region i; or (ii) leaving the productive sector to become part of the political elite, thereby securing a claim to the revenues generated by taxation.

[20] Note also that the size of the elite is likely to influence the spatial equilibrium in more 'demand-side' oriented modeling frameworks that allow for variable demand elasticities and pro-competitive effects (e.g., Ottaviano *et al.*, 2002; Behrens and Murata, 2007).

Let E stand for the mass of elite agents. Since the total mass of skilled $\overline{S}$ is fixed, the mass of productive skilled workers is given by $S = \overline{S} - E$. As pointed out in Section 3.3, we assume that skilled workers do not incur any cost other than the forfeit of their productive wage when entering the elite. The benefit of doing so is to secure a claim to an equal share of tax rents levied by the elite. More formally, skilled workers will choose to enter the elite if doing so yields a higher indirect utility than staying in the productive sector.

In what follows we focus, for analytical tractability, on the case with full agglomeration of all skilled into the capital region 1 only (i.e., $\lambda^* = 1$).[21] We know from Proposition 1 that the presence of an elite reinforces agglomeration, which implies that we can focus on the case in which there is dispersion in the absence of an elite, whereas there is full agglomeration in its presence (numerical examples are easy to construct). Although this scenario is a special case, our results are likely to extend to the cases with partial agglomeration.

We start with the question of how the unproductive political elite is formed, and how it sets taxes as to maximize its well-being. Let

$$\Delta W^* = \mu^\mu (1-\mu)^{1-\mu} \left(\frac{r^*}{\mathbb{P}_1^\mu} - \frac{w_1^*(1-t^S)}{\mathbb{P}_1^\mu} \right), \tag{4.18}$$

stand for the indirect utility differential between a member of the elite and a productive skilled in the capital region 1. Elite formation will take place until indirect utilities are equalized. Using the elite members income r^*, given by (4.9), the value of E will be determined such that

$$S \frac{(\sigma-1)t^U + t^S}{E} [\lambda w_1^* + (1-\lambda) w_2^*] - w_1^*(1-t^S) = 0.$$

Letting $\lambda^* = 1$ and using $S \equiv \bar{S} - E$, the mass E of the elite will be determined such that

$$(\bar{S} - E) \frac{(\sigma-1)t^U + t^S}{E} - (1-t^S) = 0. \tag{4.19}$$

The unique solution E^* to (4.19) is such that:

$$E^* = \bar{S} \left[1 - \frac{1-t^S}{1+t^U(\sigma-1)} \right], \quad \text{and hence} \quad S^* = \bar{S} \frac{1-t^S}{1+t^U(\sigma-1)}. \tag{4.20}$$

Expression (4.20) shows that there is no political elite when there is no taxation ($t^S = t^U = 0$), whereas when rent-extraction from the skilled is complete ($t^S = 1$) all skilled workers

[21] This assumption can be justified by the fact that the elite always wants to implement full agglomeration in its region of residence because this maximizes its access to product variety (see also Charlot *et al.*, 2005, Prop. 3). A convenient by-product of letting $\lambda^* = 1$ is that it is impossible to obtain analytical results for the case with partial agglomeration, because ΔV^* is transcendental in λ (see also Baldwin and Krugman, 2004, who face similar problems and focus on the case with full agglomeration only).

want to be part of the elite. Furthermore, taxation reduces the mass of productive skilled by inflating the unproductive elite, *thereby decreasing product variety and consumption benefits*. This latter aspect may be quite important in practice but has, to the best of our knowledge, not really been highlighted in models of agglomeration and rent-seeking until now.

Some straightforward calculation furthermore shows that $(\partial E^*)/(\partial\sigma) > 0$. These results may be summarized as follows:

Proposition 2 (size of the elite) *Under full agglomeration in the capital, the size of the unproductive elite decreases with the degree of product differentiation (smaller σ) and increases with both tax rates t^S and t^U.*

Note that a smaller degree of product differentiation (higher σ) tends to inflate the size of the unproductive elite. The intuition underlying this result is that when firms in the manufacturing sector have little market power they pay lower wages, which raises the payoff to political rent-seeking when compared to productive activity. Everything else equal, this fosters entry into the elite.

4.5.2 Tax setting

The final stage involves tax setting by the elite, taking into account how its decision influences subsequent 'entry into the elite' by skilled workers.[22] In what follows, we assume that the elite sets tax rates such as to maximize its indirect utility and not necessarily its tax revenue. Stated differently, the elite takes into account the fact that higher taxes and the subsequent 'entry into the elite' reduces product variety and welfare by diverting productive resources.[23]

Dropping the positive constant, the indirect utility of a member of the political elite is given by $V_E^* = r^*/P_1^{\mu}$, where the subscript E refers to the elite. Letting $S^* = \overline{S} - E^*$,

[22]It is worth emphasizing that our results are likely to depend on the timing of the game. Indeed, in our model the elite sets taxes by anticipating how its choice will affect further entry into the elite (and, therefore, the respective payoffs of being either elite or productive skilled). In a game where entry and elite formation occur simultaneously, results may be different.

[23]In public economic terms, the elite is not a 'leviathan' but a welfare maximizer (for itself). Note that the 'leviathan' case is likely to lead to an even larger elite as it disregards the negative variety effect. Note also that the 'variety constraint' faced by the elite, which constrains its ability to extract rents, may be less stringent nowadays because imports can supplement the needs of the elite. Stated differently, international trade may be a driver for agglomeration in a setting where unproductive elites maximize their welfare since it increases their extractive capacity. Though interesting, we disregard this possibility in what follows.

$\lambda^* = 1$, and using r^* as given by (4.9), we then obtain:

$$V_E^*\big|_{\lambda^*=1} = \left[\frac{(\sigma-1)(1-t^U)}{\sigma m}\right]^{\mu} \left(\frac{\overline{S}-E^*}{F}\right)^{\frac{\mu}{\sigma-1}} \frac{(\overline{S}-E^*)[(\sigma-1)t^U+t^S]}{E^*} w_1^*.$$

Substituting the expression of E^* as given by (4.20), and using w_1^*, we finally get

$$V_E^*\big|_{\lambda=1} = \kappa(1-t^S)^{\frac{\mu}{\sigma-1}}\xi(t^U), \tag{4.21}$$

where

$$\xi(t^U) \equiv \frac{\left(1-t^U\right)^{\mu}\left[(\sigma-1)t^U+1\right]^{1-\frac{\mu}{\sigma-1}}}{\sigma-\mu-(\sigma-1)t^U\mu} \tag{4.22}$$

and where $\kappa > 0$ is a bundle of parameters. The elite maximizes (4.21) with respect to the taxes t^U and t^S. Since $\mu/(\sigma-1) > 0$, we see that V_E^* is always decreasing in t^S. Stated differently, the elite sets $t^{S*} = 0$. This suffices to establish the following result:

Proposition 3 (optimal skilled taxation) *At any spatial equilibrium with full agglomeration, the elite will set a zero tax rate on the skilled ($t^{S*} = 0$).*

The intuition underlying Proposition 3 is that $V_E^* = V_S^*$ by arbitrage, so that the objective of maximizing the welfare of the elite is congruent with the objective of maximizing the welfare of the skilled. Consequently, the skilled will not be taxed.[24]

Results with respect to the unskilled tax rate t^{U*} are more complicated to establish. Yet, we can show the following:

Proposition 4 (optimal unskilled taxation) *At any spatial equilibrium with full agglomeration, there exists a threshold*

$$\sigma^t \equiv \frac{1}{2} + \mu + \frac{1}{2}\sqrt{1+4(1-\mu)\mu} < 2$$

such that

1. *$t^{U*} = 0$ and there is no entry into the elite ($E^* = 0$) when $\sigma \leq \sigma^t$;*

2. *there is a unique $0 < t^{U*} < 1$ with entry into the elite ($E^* > 0$) when $\sigma > \sigma^t$.*

Proof. See Appendix B. ■

[24]We thank Pierre M. Picard for pointing out this interpretation.

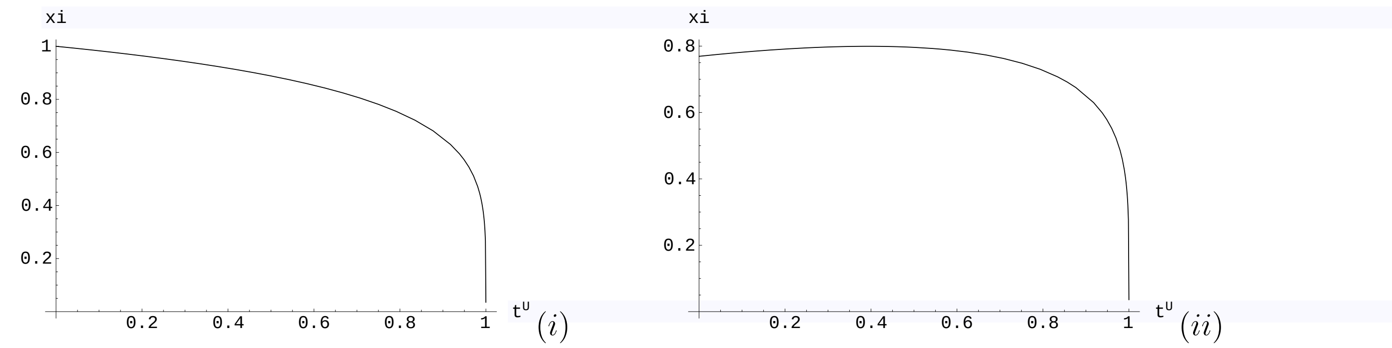

Figure 4.2: ξ as a function of the unskilled tax rate t^U

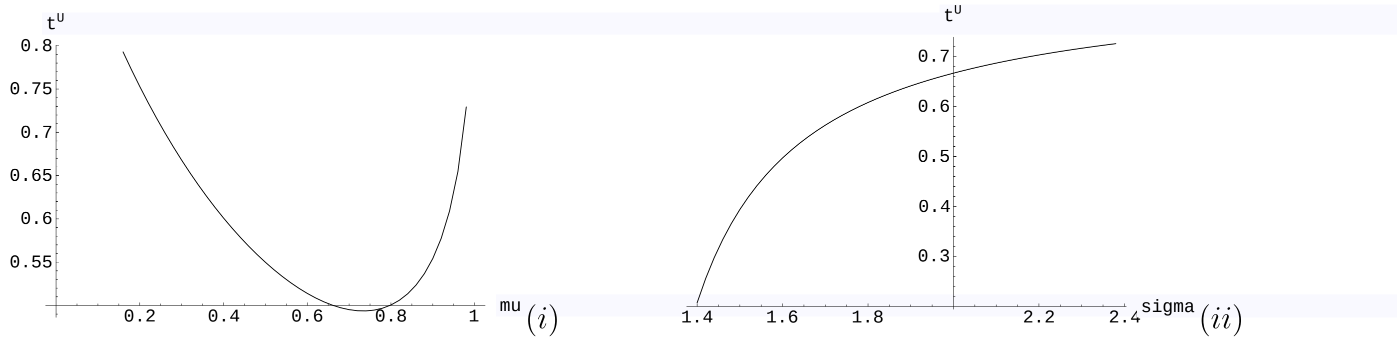

Figure 4.3: Equilibrium tax rate t^{U*} as a function of parameters

Proposition 4 shows that when products are sufficiently differentiated and firms have enough market power, there will be no elite formation in equilibrium, even if it is a priori possible. The intuition underlying this result is that since it is profitable enough to work in the productive sector, agents cannot win by pursuing unproductive rent-seeking. This finding suggests that a strong reliance on relatively homogeneous products, as often observed in developing countries, may be a strong driver for elite formation as the payoff of unproductive to productive activity rises.

Entry of skilled into the elite both erodes per capita rents in a standard way and, by taxing agents and thus raising production costs, increases wages in the modern sector. Both effects are penalizing and may be strong enough to prevent elite formation in equilibrium. Panel (i) of Figure 4.2 illustrates this case where no elite emerges in equilibrium, whereas there is a strictly positive tax rate and elite in equilibrium in the case depicted in panel (ii).[25] The former case occurs for a low elasticity of substitution, whereas the latter case occurs for a high elasticity of substitution.

Note that t^{U*} is strictly decreasing in μ when $\sigma \leq 2$, whereas it is first decreasing and

[25]The parameter values underlying Figure 4.2 are set as follows: panel (i) $\mu = 0.2$ and $\sigma = 1.2$; and panel (ii) $\mu = 0.2$ and $\sigma = 1.5$.

then increasing in μ when $\sigma > 2$ (see panel (i) in Figure 4.3).[26] The intuition underlying the second effect is that when expenditures on the modern sector account only for a small part of the budget, consumers are willing to buy even when the price is high, which allows for a higher taxation of the unskilled without reducing tax revenues and product variety. Finally, clear-cut comparative static results with respect to σ are not easily derived since the various thresholds are parametrized by σ. Yet, numerical examples suggest that, as expected, t^{U*} is increasing in σ (see panel (ii) in Figure 4.3). The less differentiated the goods are, the lower the unskilled tax rate. The reason is that higher unskilled taxes raise production costs in the differentiated sector, thereby eroding profits and triggering skilled entry into the elite.

To summarize our main findings, both a large expenditure share on the agricultural good and little product differentiation in the manufacturing sector are drivers for high tax rates on unskilled and a large elite, as they erode the return to productive activity. Both of these features seem to be largely prevalent in the developing world and may, therefore, serve to explain the presence of a sizable unproductive elite.

4.6 Conclusion

Our results suggest that rent-seeking behavior fuels the formation of large urban agglomerations in developing countries, via mechanisms of interregional income transfers. Stated differently, even if increasing returns and the freeness of trade are too low for agglomeration to occur in equilibrium, spatial concentration may arise because of the additional purchasing power generated by transfers from the country-side to the capital cities. Our findings may thus serve to explain a seemingly paradoxical aspect of urban development in SSA: *agglomeration despite high trade costs*. Indeed, the NEG literature roughly predicts a positive and monotone relationship between freeness of trade and the degree of agglomeration, whereas most SSA countries are characterized by high values of interregional trade costs.[27] Rent-seeking behavior may serve to explain why agglomeration nevertheless occurs in such

[26]The parameter values underlying Figure 4.3 are set as follows: panel (i) $\sigma = 2.7$ and μ varies from 0 to 1; and panel (ii) $\mu = 0.2$ and σ varies from 1.4 to 2.4.

[27]A particularly illuminating example is detailed in the article "The road to hell is unpaved" of the December 19, 2002, print edition of The Economist (available online at `http://www.economist.com/PrinterFriendly.cfm?Story_ID=1487583`). Due to administrative hassle, 47 road-blocks and poor infrastructure, a 500km trip by truck from Douala to Bertoua took the authors four days, with only two-thirds of the original load arriving at its destination. Bad infrastructure and high transaction costs therefore add substantially to total shipping costs.

a context.

We have shown that rent-seeking will always lead to the formation of an elite provided that goods are not too differentiated. Furthermore, we have shown that the elite sets zero tax rates for skilled workers when maximizing its own welfare. This result is consistent with the well-acknowledged fact that the economic and political power of rich taxpayers often allows them to prevent fiscal reforms which would otherwise increase their tax burdens, so that the main bulk of taxation falls heavily on the poor (Tanzi and Zee, 2001; Howard, 2001). The smaller the elite, the higher the incentives for skilled to engage in rent-seeking behavior because the payoff in doing so rises. If we proxy the share of skilled by that of people having benefited from higher education, we see that SSA (with only 2.43% of enrollment in higher education; Barro and Lee, 2001) has one of the worst performances in this respect. Hence, when combined with the low returns to private investment, the payoff to skilled workers of entering the political elite may be quite large in this region of the world. Other factors also contribute to increase the size of the elite in SSA. Indeed, we have shown that in equilibrium the size of the elite can be expected to be inversely related to the share of manufacturing expenses in national income (via the equilibrium tax rates) and to the degree of product differentiation. As a matter of fact, low product differentiation is likely to be a feature of SSA, because economic development is positively correlated with product diversity (Falkinger and Zweimüller, 1996; Frensch and Gaucaite-Wittich, 2004). Moreover, most of the countries in this region are characterized by small manufacturing sectors and small expenditure shares on manufactured goods. SSA was in 2002 the region exhibiting the lowest ratio of manufacturing value added to GDP (13.6% versus 15.8%, 18.2%, 20.1% and 33.1% for Latin America, Europe and Central Asia, the European Union, and East Asia and Pacific, respectively). Looking only at SSA countries is even more informative: of the 43 SSA countries, 23 have ratios lower than 10% and 9 have ratios of even less than 5%.[28] This point suggests that the payoff to productive work is quite low, enticing skilled workers to engage in rent-seeking activities, which might drive urban primacy.

Finally, the presence of a rent-seeking elite also reduces the range of available goods because it diverts productive resources. In an endogenous growth setting à la Grossman and Helpman (1991), where varieties also serve as intermediate inputs, the reduction of variety due to the presence of the elite may damage long-run growth, which may shed light on the relatively bad growth performance of SSA. Indeed, this region has experienced an urban growth rate outpacing its economic growth rate. From 1980 to 1990 and from 1990

[28] Figures are from World Bank Development Indicators, 2005, available online at the following address: http://devdata.worldbank.org/data-query.

to 1999, the average GDP growth rate in that region was respectively 1.7% and 2.4%, but from 1980 to 1990 and from 1990 to 2000, its average urban growth rate was 4.71% and 4.49%, i.e., 277% and 187% of the corresponding GDP growth rate, respectively.[29]

So far we have explained the urban centralization prevailing in SSA by the political influence of their urban elites living in the political capitals. In the next chapter we rely on another crucial feature of their urbanization process, i.e. the locational advantages of SSA hugest cities, to grasp the urban primacy characterizing countries of that region.

4.7 Appendix A: Sample and data

The dataset consists of 149 countries and is available from the authors upon request. It draws on publicly available data only.

Country list: Afghanistan, Albania, Algeria, Angola, Argentina, Armenia, Australia, Austria, Azerbaijan, Bahrain, Bangladesh, Belarus, Belgium, Belize, Benin, Bolivia, Bosnia and Herzegovina, Botswana, Brazil, Bulgaria, Burkina Faso, Burundi, Cambodia, Cameroon, Canada, Chad, Chile, China, Colombia, Congo (Democratic Republic), Congo (Republic), Costa Rica, Côte d'Ivoire, Croatia, Cuba, Cyprus, Czech Republic, Denmark, Dominican Republic, Ecuador, Egypt, El Salvador, Equatorial Guinea, Eritrea, Estonia, Ethiopia, Fiji, Finland, France, Gabon, Georgia, Germany, Ghana, Greece, Guatemala, Guyana, Haiti, Honduras, Hungary, Iceland, India, Indonesia, Iran, Iraq, Ireland, Israel, Italy, Jamaica, Japan, Jordan, Kazakhstan, Kenya, Kuwait, Kyrgyzstan, Laos, Latvia, Lebanon, Lesotho, Liberia, Lithuania, Luxembourg, Macedonia, Madagascar, Malawi, Malaysia, Mali, Malta, Mauritius, Mexico, Moldova, Mongolia, Morocco, Mozambique, Myanmar, Namibia, Nepal, Netherlands, New Zealand, Nicaragua, Niger, Nigeria, Norway, Oman, Pakistan, Palestine (Occupied territories), Panama, Papua New Guinea, Paraguay, Peru, Philippines, Poland, Portugal, Qatar, Romania, Russia, Rwanda, Saudi Arabia, Senegal, Serbia and Montenegro, Slovakia, Slovenia, Somalia, South Africa, South Korea, Spain, Sri Lanka, Sudan, Suriname, Swaziland, Sweden, Switzerland, Syria, Tajikistan, Thailand, Tunisia, Turkey, Turkmenistan, Uganda, Ukraine, United Arab Emirates, United Kingdom, Uruguay, USA,

[29]The figures for urban growth rates were computed from the Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat (2004); from the World Urbanization Prospects: The 2003 Revision; and from UNWUP2003 (POP/DB/WUP/Rev.2003/Data set 1/File 6), data set in digital form. The figures for the GDP growth rates were obtained from the World Development Report 2000–2001.

Uzbekistan, Venezuela, Vietnam, Yemen, Zambia, Zimbabwe.

Variables:

1. *Urban primacy* is defined as the percentage of urban population living in the country's largest city. This data is taken from the UN Populations Division "World Urbanization Prospects Population Database: The 2003 Revision" for the year 2000, available online at `http://esa.un.org/unup/`.

2. *Ratio of largest to second-largest city* is computed from the UN Populations Division "World Urbanization Prospects Population Database: The 2003 Revision" for the year 2000, available online at `http://esa.un.org/unup/`.

3. *Food expenditure share* in consumption is taken from ERS, United States Department of Agriculture, available online at `http://www.ers.usda.gov/Data/International FoodDemand/`. It covers the year 2004.

4. *Non-corruption index* is the "2005 Transparency International Corruption Perceptions Index". It ranges from 1 (most corrupt) to 10 (least corrupt). It "[...] focuses on corruption in the public sector and defines corruption as the abuse of public office for private gain. The surveys used in compiling the CPI ask questions that relate to the misuse of public power for private benefit, with a focus, for example, on bribe-taking by public officials in public procurement." It is available online at `http://www.transparency.org/policy_and_research /surveys_indices/cpi/2005`.

5. *Top income tax rate* and *top corporate tax rate* are taken from the "2004 Index of Economic Freedom", available online at `http://cf.heritage.org/ index2004test/`.

6. *Size of the shadow economy* in 2002/2003 as percentage of GDP is taken from Schneider (2004, Table 5.4).

7. *GDP per capita* in 2003 $US is taken from the "2005 World Development Indicators", available at `http://www.devdata.worldbank.org/wdi2005/cover.htm`.

8. *Central government expenditures on goods, services, and compensation of employees* in % of central government expenditures in 2003 are from the "2005 World Development Indicators", available online at `http://www.devdata.worldbank.org/ wdi2005/cover.htm`.

4.8 Appendix B: Proof of Proposition 4

To prove our claim, we need to analyze the function ξ, as given by (4.22). First, note that ξ is continuous in t^U and non-negative on the interval $[0,1]$. Furthermore, when $\mu < \sigma - 1$ holds (the 'no black hole condition'), we have

$$\lim_{t^U \to 1} \xi(t^U) = 0 \qquad \text{and} \qquad \lim_{t^U \to 0} \xi(t^U) = \frac{1}{\sigma - \mu} > 0. \tag{4.23}$$

Stated differently, prohibitive taxation leads to zero tax revenues and zero utility for the elite, whereas sufficiently low taxation yields positive utility. Second, one can check that $\partial\xi/\partial t^U = 0$ if and only if

$$t^U = \frac{(4-3\sigma)\mu^2 + (\sigma-2)\sigma\mu + \sigma\left[\sigma \pm \sqrt{(1-\mu)(\sigma-\mu-1)(\sigma+\mu(\mu+3\sigma-4)-1)} - 1\right]}{2\mu^2(\sigma-2)(\sigma-1)}$$

It is readily verified that the larger of the two roots exceeds 1 and, therefore, does not belong to the range of admissible values. Consequently, ξ is either monotonously decreasing on $[0,1]$, or admits a unique maximum (the existence of an interior minimum is ruled out by (4.23) and because $\xi \geq 0$; see Figure 2 for the two cases).

Some straightforward computations show that

$$\lim_{t^U \to 0} \frac{\partial \xi(t^U)}{\partial t^U} = \frac{2\mu^2 - 2\sigma\mu + (\sigma-1)\sigma}{(\mu-\sigma)^2}.$$

Equating this expression to zero, the two roots in σ are such that the smaller one is less than 1 and must be ruled out, whereas the larger one is given by:

$$\sigma^t = \frac{1}{2} + \mu + \frac{1}{2}\sqrt{1 + 4(1-\mu)\mu} < 2.$$

Hence, when $\sigma < \sigma^t$ the function ξ si strictly decreasing for all tax rates and the elite's best choice is to not tax the unskilled. Obviously, there will be no elite in this case. When $\sigma > \sigma^t$, the elite chooses $0 < t^{U*} < 1$ and there will be elite formation. The optimal tax rate is therefore given as follows:

$$t^{U*} = \begin{cases} 0 & \text{if} \quad 1 < \sigma < \sigma^t \\ \frac{(4-3\sigma)\mu^2+(\sigma-2)\sigma\mu+\sigma\left(\sigma-\sqrt{(1-\mu)(\sigma-\mu-1)(\sigma+\mu(\mu+3\sigma-4)-1)}-1\right)}{2\mu^2(\sigma-2)(\sigma-1)} & \text{if} \quad \sigma^t < \sigma,\ \sigma \neq 2 \\ \frac{1-\mu}{1+\mu} & \text{if} \quad \sigma = 2 \end{cases}$$

which establishes Proposition 4.

Bibliography

[1] Ades, A. (1995) Economic development with endogenous political participation, *Economics and Politics* 7, 93-117.

[2] Ades, A. and E.L. Glaeser (1995) Trade and circuses: Explaining urban giants, *Quarterly Journal of Economics.*

[3] Bairoch, P. (1985) *De Jericho à Mexico: Villes et économie dans l'histoire.* Paris: Gallimard. English translation *Cities and Economic Development. From the Dawn of History to the Present* (1988), London: Mansell Publishing Ltd.

[4] Baldwin, R.E., R. Forslid, Ph. Martin, G.I.P. Ottaviano and F. Robert-Nicoud (2003) *Economic Geography and Public Policy.* Princeton: Princeton Univ. Press.

[5] Baldwin, R.E. and P.R. Krugman (2004) Agglomeration, integration and tax harmonization, *European Economic Review* 48, 1-23.

[6] Barro, R. J. and J.-W. Lee (2001) International data on educational attainment: updates and implications, *Oxford Economic Papers* 3, 541-563.

[7] Bassey, M.O. (1999) *Western Education and Political Domination in Africa. A Study in Critical and Dialogical Pedagogy.* Westport: Bergin & Garvey.

[8] Baumol, W.J. (1990) Entrepreneurship: Productive, unproductive, and destructive, *Journal of Political Economy* 98, 893-921.

[9] Behrens, K. and Y. Murata (2007) General equilibrium models of monopolistic competition: a new approach, *Journal of Economic Theory*, forthcoming.

[10] Cantillon, R. (1730) *Essai sur la Nature du Commerce en Général.* Full english version available online at `http://www.econlib.org/library/NPDBooks/Cantillon/cntNTContents.html`

[11] Charlot, S., C. Gaigné, F. Robert-Nicoud and J.-F. Thisse (2006) Agglomeration and welfare: the core-periphery model in the light of Bentham, Kaldor, and Rawls, *Journal of Public Economics* 90, 325-347.

[12] Coquery-Vidrovitch, C. (1993) *Histoire des Villes d'Afrique Noire: Des Origines à la Colonisation*. Paris: Albin-Michel.

[13] Dascher, K. (2000) Are politics and geography related? Evidence from a cross-section of capital cities, *Public Choice* 105, 373-392.

[14] De Long, J.B. and A. Shleifer (1993) Princes and merchants: European city growth before the industrial revolution?, *Journal of Law and Economics* 36, 671-702.

[15] Falkinger, J. and J. Zweimüller (1996) The cross-country Engel-curve for product diversification, *Structural Change and Economic Dynamics* 7, 79-97.

[16] Forslid, R. and G.I.P. Ottaviano (2003) An analytically solvable core-periphery model, *Journal of Economic Geography* 3, 229-240.

[17] Frensch, R. and V. Gaucaite-Wittich (2004) The benefits from product differentiation in modern economies. Forost Arbeitspapier #19.

[18] Fujita, M., P.R. Krugman and A.J. Venables (1999) *The Spatial Economy: Cities, Regions, and International Trade*. Cambridge, MA: MIT Press.

[19] Fujita, M. and J.-F. Thisse (2002) *Economics of Agglomeration: Cities, Industrial Location, and Regional Growth*. Cambridge: Cambridge Univ. Press.

[20] Grossman, G.M., and E. Helpman (1991) *Innovation and Growth in the Global Economy*. Cambridge, MA: MIT Press.

[21] Howard, M. (2001) *Public Sector Economics for Developing Countries*. Barbados: Univ. of The West Indies Press. Available online at `http://www.questia.com/`

[22] Krueger, A.O. (1974) The political economy of the rent seeking society, *American Economic Review* 3, 291-303.

[23] Krugman, P. (1991) Increasing Returns and Economic Geography, *Journal of Political Economy* 99, 483-499.

[24] Mbaku, J.M. (2003) Corruption and economic development. In: Nnadozie, E. (ed.) *African Economic Development*, Academic Press, San Diego, pp. 313-349.

[25] Mbeki, M. (2005) Underdevelopment in sub-Saharan Africa: the role of the private sector and political elites, CATO Institute Foreign Policy Briefing #85.

[26] Moomaw, R.L. and A.M. Shatter (1996) Urbanization and economic development: a bias toward large cities?, *Journal of Urban Economics* 40, 13-37.

[27] Ottaviano, G.I.P., T. Tabuchi and J.-F. Thisse (2002) Agglomeration and trade revisited, *International Economic Review* 43, 409-436.

[28] Puga, D. (1998) Urbanization patters: European vs less developed countries, *Journal of Regional Science* 38, 231-252.

[29] Robert-Nicoud, F. and F. Sbergami (2004) Home-market vs. vote-market effect: Location equilibrium in a probabilistic voting model, *European Economic Review* 48, 155-179.

[30] Robert-Nicoud, F. (2005) The structure of simple 'New Economic Geography' models (or, On identical twins), *Journal of Economic Geography* 5, 201-234.

[31] Schneider, F. (2004) The size of the shadow economies of 145 countries all over the world: First results over the period 1999-2003, IZA Discussion Paper #1431.

[32] Stahl, K. (1983) A note on the microeconomics of migration, *Journal of Urban Economics* 14, 318-326.

[33] Tanzi, V. and H., Zee (2001) *Tax Policy for Developing Countries.* International Monetary Fund: available online at `http://www.imf.org/external/pubs/ft/issues/issues27/`.

[34] Williamson, J. G. (1988) Migration and Urbanization. In: Chenery, H. and T.N. Srinivasan (eds.) *Handbook of Development Economics, vol. 1*, North-Holland: Elsevier Science Publishers, pp. 424-465.

Chapter 5

Hubs and Urban Primacy in Sub-Saharan Africa

5.1 Introduction

Sub-Saharan Africa (henceforth, SSA) faces an historically unprecedented absolute rate of urban growth. With an urban population growth rate averaging almost 5 percent per year, its urban population is expected to double every 15 years. Yet, while rates of urban population growth of cities of SSA remain the highest of the World, urbanization in that region is taking place in a context of severe constraints that did not face other country groups in other periods: full exposure to pressures of global competition, depredation of the productive workforce, weak industrial sector etc. (Kessides, 2005).

The ability of African cities to cope economically, environmentally, and politically with such acute concentrations of people is subject to serious concerns. Many conjecture that disadvantages of African agglomerations such as social costs of a progressive overloading of housing and social services, of increased crime, pollution, and congestion outweigh their expected urban advantages in terms of agglomeration economies (Todaro, 1997).

Economics don't appear as the only drive of urban development in that region. In the fourth chapter we have indeed shown how the synergy between scale economies and political factors may explain the apparition of urban agglomerations that would not have emerged otherwise. According to this theoretical setting cumulative causation arises thanks to the ability of political leaders to extract rents from hinterland to the capital city. In this framework the primate city appears merely as 'parasitic' in the Hoselitz (1955) sense since such transfers divert resources from productive use and therefore may impede long run economic growth.

However, relying on this rent seeking story may not be the only way to explain emergence of agglomerations in Black Africa. From the few papers attempting to explain spatial distribution in developing countries, Krugman and Elizondo (1996)' gives some clues by outlining the linkage between trade policies and urban development. They explain urban concentration in developing countries and especially in Mexico by the import-substitution policies which by closing local markets strengthen backward and forward linkages and thus favor agglomerations. So they predicted that once an economy is opened up urban concentration may shrink.

Actually import-substitution policies applied after political independence in many countries of Africa and Latin America were characterized by a strong urban bias flavor. With their focus on industrialization, technological sophistication, modern education and metropolitan growth, such strategies induced a significant spatial imbalance in economic and noneconomic opportunities between rural and urban areas and therefore contributed significantly to rural-urban migration (Todaro, 1997; Mabogunje, 1994). However, the failure of industrialization strategies in SSA has not been followed by a shrinking of the size of african agglomerations. Moreover, despite of this industrial collapse Africa urban growth rates go unabated. Therefore, Krugman and Elizondo argument does not provide a convincing explanation of Africa 's urban development.

One way to tackle satisfactorily that issue may be to analyze the impact of international trade in the spatial distribution of economic activities in a setting featuring an asymmetric location of regions. Most of NEG contributions addressing the impact of international trade in the distribution of firms in a domestic economy (Krugman and Elizondo, 1996; Monfort and Nicolini, 2000; Behrens *et al.*, 2007) imply a symmetric location of regions. In such frameworks there is no room for locational advantages or disadvantages. They would rather imply that all locations of SSA have the same accessibility to foreign countries. This is actually hardly plausible, a basic stylized fact of SSA urban geography being that most of the largest cities of Africa are located along the coast. The logic behind such a locational preference for coastal sites is that, because of the heavy import dependence of SSA economic strategies, port cities are the preferred location for industrial development (Mabogunje, 1994). This feature is not new, it holds since the beginning of colonization: at that time ports were vital for the outward shipment of raw materials back to the colonizing countries and the inward shipment of manufactured goods. By 1900 when the partition of the continent was effectively achieved, 25 (i.e. about 69%) of the 36 capitals of countries having a sea access were located in the coast (See Figure 5.7 in Appendix A). From that time on population redistribution toward these coastal cities did not cease. It rather increased as those cities retained and extended their dominance as the primary centers of economic activities (Kempe, 1996). Indeed, 15 of the 25 coastal colonial capitals in existence in 1900 still retain their status in 1991 (Figure 5.8 in Appendix B). Independence therefore has not induced dramatic changes in spatial distribution of economic agents. Even in the few cases where the capital were moved from the hub to an interior location, the hub remained the primate city. This is the case in Côte d'Ivoire, Nigeria, Cameroun, Tanzania where the capitals were respectively moved from Abidjan to Yamoussoukro, from Lagos to Abuja, from Douala to Yaounde and from Dar es Salaam to Dodoma. Therefore, as in colonial times most of the primate cities, 23 out of 33 (i.e. about 70%), have a coastal location.

The fact that the spatial structures of most African economies are strongly focused on a small number of port cities clearly points out to a hub effect which discards the symmetry assumption. So one of the explanations of the localization of firms and consumers in SSA may rely on gate effects. There is some literature on hub effects (Fujita and Mori, 1996; Krugman, 1993). But only few recent NEG contributions have addressed convincingly hub effects issues, reducing further the gap between a reality where geography and locational advantages (the so-called 'First Nature') are part of the story of economic agents localization and most of NEG papers which abstracted from geographical features to focus on purely economic mechanisms (the so-called 'Second Nature').

Ago *et al.* (2006) for instance analyze the impacts of falling transport costs on the spatial distribution of economic activities and welfare for a network economy consisting of three regions located on a line. They showed that, conversely to a Krugman setting which implies concentration in the central region (the so-called hub), this may not be the case in a Ottaviano *et al.* (2002) (henceforth, OTT) model because price competition is so intense in the central region that it may reduce welfare. Our framework adopts a similar structure with three regions located on a line, but departs from the symmetry assumption implied by Ago *et al.* (2006)' setting with two similar 'peripheral regions' endowed with the same mass of skilled and unskilled workers, and transport costs that are equivalent between the hub and each one of the peripheral regions. To better capture the impact of international trade on the distribution of firms in the domestic economy, we make the sensible assumption that the two 'peripheral' regions of our setting namely the rest of the World and the Hinterland are heterogeneous. Moreover, we assume that transport costs between the hub and each other region differ: interregional trade frictions between the Hub and the hinterland are merely constituted by transport costs and are called as such. Conversely, trade between the hub and the rest of the World also includes in addition to transport costs other international trade impediments related to institutional factors like trade policy, customs duties and formalities, or adaptation to foreign legislation etc. We denote all those international trade barriers by trade costs.

Behrens *et al.* (2006) on the other hand study the impacts of international trade and domestic transport costs on the internal geography of a country by using a two-country four regions model in which one country has a region that exhibits a 'geographical advantage' in terms of better access to the other country's market. Their main results are that the space-economies of the trading partners are interdependent and that agglomeration in one country reduces the occurrence of agglomeration in the other. They further find that the landlocked region may be the location that attracts the larger share of firms especially when transport costs in the gated country are high. Like us they make a distinction between interregional trade barriers, 'transport costs', and international trade impediments, the 'trade costs'. Nevertheless, the major difference between our model and those two contributions is that we consider firms distribution in the rest of the world as exogenous while they consider repartition of firms in any region as endogenous. Based on this assumption, Behrens *et al.* (2006) find that distribution of firms in countries involved in trade are interdependent. We believe however that such result may not be sensible for SSA. Indeed, with a share of international trade converging to that region not exceeding 3%, it can not be expected to have a substantial influence on the location of foreign activities.[1]

Therefore, conversely to the latter contributions, our framework has the specificity to assume asymmetry between the Rest of the World and the Hinterland and the independence of foreign firms localization with respects to localization choices in the Domestic Economy and is likely to offer a more realistic explanation of spatial location of firms in SSA. Moreover, our setting may deal with the paradox of the increasing integration of SSA with foreign countries through international trade, and its low interregional integration because of its poor communication infrastructure, and subsequent high transport costs.[2]

The remainder of this chapter is organized as follows. Section 2 presents the common structure of the two models presented in this chapter. In Section 3, we develop the Footloose

[1]According to OECD data (`http://www.oecd.org/dataoecd/53/47/39759637.pdf`), from 2000 to 2006, African exports has increased to reach the value of 290 billions dollars and African share of International trade has increased from 2.0% to 2.3%.

[2]A remarkably vivid description of communication infrastructure in SSA may be found in the article "The road to hell is unpaved" of the December 19, 2002, print edition of The Economist (available online at `http://www.economist.com/ PrinterFriendly.cfm?Story_ID=1487583`). Due to administrative hassle, 47 road-blocks and poor infrastructure, a 500 km trip by truck from Douala to Bertoua took the authors four days, with only two-thirds of the original load arriving at its destination. Therefore, high transaction costs add substantially to total shipping costs.

Entrepreneur Model and present simulation results. Section 4 then investigates the spatial equilibrium in the OTT model and presents the most significant analytical findings. We finally conclude in Section 5.

5.2 Structure of the economy

We consider an economy consisting in three locations, Regions $i \in \{0, 1, 2\}$, located equidistantly on a line. Region 0 depicts the rest of the world, while Regions 1 and 2 are domestic locations. Without loss of generality we assume that Region 1 is the capital and the hub of our domestic economy, while Region 2 represents its hinterland.

There are two factors of production: skilled and unskilled labor. Total labor endowments for skilled and unskilled labor are respectively H and L. Masses of skilled and unskilled labor are respectively $H_F = H_0$ and $L_F = L_0$ in the rest of the World and H_D and L_D in the domestic economy. Unskilled workers are immobile. To avoid giving to any domestic region an advantage in terms of its unskilled demand, we assume that all of them have the same share of domestic unskilled workers $L_i = \frac{1}{2}L_D$ for $i \in \{1, 2\}$. Skilled workers may move within the domestic economy while foreign skilled workers are immobile. So while the mass of skilled workers in foreign locations H_F is exogenous because of factor immobility between the Domestic Economy and the rest of the World, the distribution of skilled between domestic locations is endogenous with H_1 and H_2 respectively the masses of skilled in the hub and the hinterland regions. Those assumptions are in line with empirical evidence: there is no evidence of any significant brain drain from Rest of the World to SSA, it is rather in the other direction than skilled migration is substantial.[3] Moreover, in SSA educated persons have a higher propensity to migrate than less qualified people (Byerlee, 1974).[4]

In each region there are two production sectors, manufacturing and agriculture. By using exclusively unskilled labor, the agricultural sector produces a homogeneous good. We assume this good as costlessly tradable across regions. Moreover, we normalize, without any loss of generality, the unit input coefficient in this sector to one. Then by perfect

[3]Considering this brain drain, we could actually allow 'Domestic' skilled workers to move from the Domestic Economy to the Rest of the World, but since empirically the number of SSA skilled workers represents only a small share of Foreign skilled workers, this would complicate the calculations without providing in exchange any additional insight.

[4]Byerlee even stated that the dominance of school-leavers in the migration stream in SSA is stronger relative to Latin America and Asia where illiterate landless laborers and tenants make up a significant proportion of migrants.

competition and costless trade unskilled wages w^L are equalized across regions: $w_i^L = 1$ for $i \in \{0, 1, 2\}$. On the other hand the manufacturing sector requires both skilled and unskilled labor to produce horizontally differentiated varieties of manufactured good.

As previously mentioned trade is inhibited by frictional trade barriers that are different according to the origin and destination involved. More exactly transactions between the rest of the world and the hub and those between the hub and the hinterland imply different 'trade' costs.

We assume that there is a continuum of potential firms so that the impact of each firm on the market outcome is negligible. Since we assume that there is no economy of scope, each variety is produced by a single firm in only one region. Because varieties are symmetric, each firm's output is equalized in equilibrium. We further make the standard assumption that mobile workers are short-sighted and choose their locations as to maximize their well-being captured by their indirect utility. Supposing that market clearing conditions hold, the equilibrium distribution of firms in the Domestic Economy is given by the scalar λ and mobile labor migration is regulated by the following Marshallian adjustment process:

$$\dot{\lambda} \equiv \frac{d\lambda}{dt} = \begin{cases} \Delta V(\lambda) & \text{if } 0 < \lambda < 1 \\ \min\{0, \Delta V(\lambda)\} & \text{if } \lambda = 1 \\ \max\{0, \Delta V(\lambda)\} & \text{if } \lambda = 0 \end{cases} \tag{5.1}$$

where ΔV depicts the indirect utility differential.

To check up the robustness of our results, we will analyze the impact of international trade on the domestic space-economy by using two different models, namely the so-called 'Footloose Entrepreneur Model' of Forslid and Ottaviano (2003) and the linear model of Ottaviano *et al.* (2002). Because of its relative analytical intractability, we will use the 'Footloose Entrepreneur Model' as a benchmark pointing to key results through simulations. For crucial analytical results we will rely mostly on the OTT version.

5.3 Footloose Entrepreneur Model

This model is based on Forslid and Ottaviano (2003) with CES utility function, iceberg transport costs.

5.3.1 Preferences

A representative consumer in Region $i \in \{0, 1, 2\}$ has Cobb-Douglas upper-tier preferences over agricultural and manufactured goods, with a CES sub-utility over a continuum of horizontally differentiated varieties. Therefore, he maximizes the following utility function:

$$A_i^{1-\mu}\left(\int_{\Omega_0} q_{ii}(\omega)^{\frac{\sigma-1}{\sigma}} d(\omega) + \int_{\Omega_1} q_{ji}(\omega)^{\frac{\sigma-1}{\sigma}} d(\omega) + \int_{\Omega_2} q_{ki}(\omega)^{\frac{\sigma-1}{\sigma}} d(\omega)\right)^{\frac{\mu\sigma}{\sigma-1}}$$

given the constraint:

$$p^A A_i + \left(\int_{\Omega_0} p_{ii} q_{ii}(\omega) d(\omega) + \int_{\Omega_1} p_{ji} q_{ji}(\omega) d(\omega) + \int_{\Omega_2} p_{ki} q_{ki}(\omega) d(\omega)\right) = y_i$$

where A_i is the consumption of agricultural good; $q_{ji}(\omega)$ and $p_{ji}(\omega)$ represent the quantity and the price of variety ω consumed in country i and produced in country j; Ω_i stands for the set of varieties produced in country i, with measure n_i; y_i is the income of the representative consumer in region i.

By the homotheticity of preferences, we obtain the following aggregate demand for firm ω in region i when it is located in region j:

$$D_{ji}(\omega) = \frac{p_{ji}(\omega)^{-\sigma}}{\mathbb{P}_i^{1-\sigma}} \mu Y_i \tag{5.2}$$

where Y_i is the total income of agents in region $i \in \{0, 1, 2\}$ including skilled (w_i) and unskilled wages (w_i^L):

$$Y_i = w_i H_i + w_i^L L_i \tag{5.3}$$

and $\mathbb{P}_i$ is the CES price aggregate. Assuming that all varieties produced in each region are symmetric allows us to alleviate notation by dropping the variety index ω. The price aggregate $\mathbb{P}_i$ then reduces to

$$\mathbb{P}_i = \left(n_i p_{ii}^{1-\sigma} + n_j p_{ji}^{1-\sigma} + n_k p_{ki}^{1-\sigma}\right)^{\frac{1}{1-\sigma}}. \tag{5.4}$$

5.3.2 Technology and transportation

Frictional trade barriers are modeled as iceberg costs. As stated previously they are asymmetric: trade of one unit of differentiated good between the rest of the world to the hub region entails shipping of ρ units of differentiated good, while it entails τ units of differentiated goods from the hub to the hinterland region.

In this model factor wages equalization only holds when the homogeneous good is produced in all the regions (Forslid and Ottaviano, 2003; Baldwin *et al.*, 2003). Such non-full-specialization condition is verified only if the agricultural good has an important weight in the utility (μ small) and if product variety is highly valued by consumers.[5]

[5] Formally, in this three regions framework, if each region has an equal share of unskilled workers $L_i = \frac{1}{3}L$ for $i \in \{0, 1, 2\}$, the exact condition is $\mu < \sigma/(3\sigma - 2)$.

Each firm of the manufacturing sector requires F units of skilled labor as a fixed input requirement and m units of unskilled labor per unit of output as a variable input requirement. Total production costs of producing a quantity Q in region $i \in \{0, 1, 2\}$ are then given by

$$\mathrm{TC}_i(Q) = mQ + Fw_i.$$

With such a fixed cost requirement, skilled labor market clearing then requires that the masses of firms in the regions are as follows:

$$n_0 = \frac{H_F}{F} \qquad n_1 = \frac{\lambda H_D}{F} \qquad n_2 = \frac{(1-\lambda)H_D}{F} \qquad N = n_0 + n_1 + n_2 \tag{5.5}$$

where λ represents the share of the domestic unskilled labor endowment located in the hub region.

The profit of a representative firm in region i is given by

$$\Pi_i = (p_{ii} - m)\, D_{ii} + (p_{ij} - m)\, D_{ij} + (p_{ik} - m)\, D_{ik} - Fw_i,$$

where the demands are evaluated at (5.2). Taking into account the frictional trade barriers previously mentioned, profit-maximizing prices exhibit a constant mark-up over marginal cost:

$$\begin{aligned}
p^*_{00} &= p^*_{11} = p^*_{22} = \frac{\sigma m}{(\sigma - 1)} \\
p^*_{01} &= p^*_{10} = \frac{\sigma m \rho}{(\sigma - 1)} \\
p^*_{12} &= p^*_{21} = \frac{\sigma m \tau}{(\sigma - 1)} \\
p^*_{02} &= p^*_{20} = \frac{\sigma m \tau \rho}{(\sigma - 1)}
\end{aligned}$$

Replacing prices by those expressions into (5.4), and using the skilled labor market clearing conditions (5.5), we get the following price indices:

$$\mathbb{P}_0 = \frac{\sigma m}{(\sigma - 1)} \left(\frac{H_D}{F}\right)^{\frac{1}{1-\sigma}} \left[\frac{H_F}{H_D} + \lambda \rho^{1-\sigma} + (1-\lambda)\, (\tau\rho)^{1-\sigma}\right]^{\frac{1}{1-\sigma}} \tag{5.6}$$

$$\mathbb{P}_1 = \frac{\sigma m}{(\sigma - 1)} \left(\frac{H_D}{F}\right)^{\frac{1}{1-\sigma}} \left[\frac{H_F}{H_D}\rho^{1-\sigma} + \lambda + (1-\lambda)\, \tau^{1-\sigma}\right]^{\frac{1}{1-\sigma}} \tag{5.7}$$

$$\mathbb{P}_2 = \frac{\sigma m}{(\sigma - 1)} \left(\frac{H_D}{F}\right)^{\frac{1}{1-\sigma}} \left[\frac{H_F}{H_D}\, (\tau\rho)^{1-\sigma} + \lambda\tau^{1-\sigma} + (1-\lambda)\right]^{\frac{1}{1-\sigma}} \tag{5.8}$$

In this chapter we want to study the distribution of skilled workers (i.e. of firms) between the hub and the hinterland regions. To do so we need to derive equilibrium skilled wage expressions for the aforementioned regions. Because the three locations have different accessibility, product market clearing conditions yield asymmetric expressions for total quantities produced in Regions 0, 1 and 2. So we get:

$$\begin{aligned}
X_0 &= D_{00} + \rho D_{01} + \tau\rho D_{02} \\
&= \frac{\mu(\sigma-1)}{m\sigma}\left(\frac{Y_0}{n_0 + n_1\phi_\rho + n_2\phi_\tau\phi_\rho} + \frac{\phi_\rho Y_1}{n_0\phi_\rho + n_1 + n_2\phi_\tau} + \frac{\phi_\tau\phi_\rho Y_2}{n_0\phi_\tau\phi_\rho + n_1\phi_\tau + n_2}\right) \quad (5.9) \\
X_1 &= D_{11} + \rho D_{10} + \tau D_{12} \\
&= \frac{\mu(\sigma-1)}{m\sigma}\left(\frac{\phi_\rho Y_0}{n_0 + n_1\phi_\rho + n_2\phi_\tau\phi_\rho} + \frac{Y_1}{n_0\phi_\rho + n_1 + n_2\phi_\tau} + \frac{\phi_\tau Y_2}{n_0\phi_\tau\phi_\rho + n_1\phi_\tau + n_2}\right) \quad (5.10) \\
X_2 &= D_{22} + \tau D_{21} + \tau\rho D_{20} \\
&= \frac{\mu(\sigma-1)}{m\sigma}\left(\frac{\phi_\tau\phi_\rho Y_0}{n_0 + n_1\phi_\rho + n_2\phi_\tau\phi_\rho} + \frac{\phi_\tau Y_1}{n_0\phi_\rho + n_1 + n_2\phi_\tau} + \frac{Y_2}{n_0\phi_\tau\phi_\rho + n_1\phi_\tau + n_2}\right) \quad (5.11)
\end{aligned}$$

where $\phi_\tau = \tau^{1-\sigma}$ and $\phi_\rho = \rho^{1-\sigma}$ stand for freeness of respectively interregional and international trade.

As firms price above marginal cost, there exist pure operating profits which are competed away by firms' bidding for skilled labor. Therefore, in equilibrium the skilled wages absorb all operating profits:

$$\begin{aligned}
w_0 &= \frac{\mu}{\sigma F}\left(\frac{Y_0}{n_0 + n_1\phi_\rho + n_2\phi_\tau\phi_\rho} + \frac{\phi_\rho Y_1}{n_0\phi_\rho + n_1 + n_2\phi_\tau} + \frac{\phi_\tau\phi_\rho Y_2}{n_0\phi_\tau\phi_\rho + n_1\phi_\tau + n_2}\right) \quad (5.12) \\
w_1 &= \frac{\mu}{\sigma F}\left(\frac{\phi_\rho Y_0}{n_0 + n_1\phi_\rho + n_2\phi_\tau\phi_\rho} + \frac{Y_1}{n_0\phi_\rho + n_1 + n_2\phi_\tau} + \frac{\phi_\tau Y_2}{n_0\phi_\tau\phi_\rho + n_1\phi_\tau + n_2}\right) \quad (5.13) \\
w_2 &= \frac{\mu}{\sigma F}\left(\frac{\phi_\tau\phi_\rho Y_0}{n_0 + n_1\phi_\rho + n_2\phi_\tau\phi_\rho} + \frac{\phi_\tau Y_1}{n_0\phi_\rho + n_1 + n_2\phi_\tau} + \frac{Y_2}{n_0\phi_\tau\phi_\rho + n_1\phi_\tau + n_2}\right) \quad (5.14)
\end{aligned}$$

Using (5.5), (5.12), (5.13) and (5.14) can take the final expressions:

$$\begin{aligned}
w_0 &= \frac{\mu}{\sigma H_D}\left(\frac{Y_0}{\lambda_{FD} + \lambda\phi_\rho + (1-\lambda)\,\phi_\tau\phi_\rho} + \frac{\phi_\rho Y_1}{\lambda_{FD}\phi_\rho + \lambda + (1-\lambda)\,\phi_\tau} + \frac{\phi_\tau\phi_\rho Y_2}{\lambda_{FD}\phi_\tau\phi_\rho + \lambda\phi_\tau + (1-\lambda)}\right) \quad (5.15) \\
w_1 &= \frac{\mu}{\sigma H_D}\left(\frac{\phi_\rho Y_0}{\lambda_{FD} + \lambda\phi_\rho + (1-\lambda)\,\phi_\tau\phi_\rho} + \frac{Y_1}{\lambda_{FD}\phi_\rho + \lambda + (1-\lambda)\,\phi_\tau} + \frac{\phi_\tau Y_2}{\lambda_{FD}\phi_\tau\phi_\rho + \lambda\phi_\tau + (1-\lambda)}\right) \quad (5.16) \\
w_2 &= \frac{\mu}{\sigma H_D}\left(\frac{\phi_\tau\phi_\rho Y_0}{\lambda_{FD} + \lambda\phi_\rho + (1-\lambda)\,\phi_\tau\phi_\rho} + \frac{\phi_\tau Y_1}{\lambda_{FD}\phi_\rho + \lambda + (1-\lambda)\,\phi_\tau} + \frac{Y_2}{\lambda_{FD}\phi_\tau\phi_\rho + \lambda\phi_\tau + (1-\lambda)}\right) \quad (5.17)
\end{aligned}$$

with $\lambda_{FD} = \frac{H_F}{H_D}$.

We can now analyze the market outcome for any *given* spatial distribution of domestic skilled workers across domestic locations. Considering that net aggregate incomes are given by expression (5.3), the market outcome yields unique solution in w_0, w_1 and w_2 .[6] Having expressions of this solution we can now discuss the spatial equilibrium. As indicated earlier mobile labor migration in the Domestic Economy is regulated by the aforementioned

[6] Expressions of this solution are quite long, so we put them in the Appendix C

Marshallian adjustment process described by (5.1). It is a myopic adjustment process whose driving force is skilled workers' current utility differential between the hub and the hinterland. This utility differential has the following expression:

$$\Delta V(\lambda) = \mu^{\mu}(1-\mu)^{1-\mu}\left(\frac{w_1}{\mathbb{P}_1^{\mu}} - \frac{w_2}{\mathbb{P}_2^{\mu}}\right).$$

5.3.3 Benchmark case: a closed domestic economy

When the Domestic economy is closed, that is when $\rho \to \infty$ (or $\phi_\rho \to 0$) , we are in the benchmark case of Forslid and Ottaviano (2003). Indeed, ΔV reduces to

$$\begin{aligned} \frac{\Delta V(\lambda)}{\Phi} &= \frac{2\lambda\sigma\phi_\tau + (1-\lambda)\left[\sigma\left(\phi_\tau^2+1\right) - \mu\left(1-\phi_\tau^2\right)\right]}{\left[(1-\lambda)\phi_\tau + \lambda\right]^{\frac{\mu}{1-\sigma}}} \\ &\quad - \frac{2(1-\lambda)\sigma\phi_\tau + \lambda\left[(\mu+\sigma)\phi_\tau^2 + (\sigma-\mu)\right]}{\left[\lambda\phi_\tau + (1-\lambda)\right]^{\frac{\mu}{1-\sigma}}}. \end{aligned} \tag{5.18}$$

where Φ is a strictly positive bundle of parameters given by the following expression

$$\Phi \equiv \frac{L(1-\mu)^{1-\mu}\mu^{\mu+1}}{\eta}\left[\frac{m\left(\frac{H}{F}\right)^{\frac{1}{1-\sigma}}\sigma}{(\sigma-1)}\right]^{-\mu}$$

with

$$\eta \equiv \left((1-\lambda)^2 + \lambda^2\right)\sigma\,\phi_\tau + \left(\sigma - \mu + (\mu+\sigma)\,{\phi_\tau}^2\right)(1-\lambda)\,\lambda$$

It is shown in Forslid and Ottaviano (2003) that full agglomeration may be sustained as an equilibrium if and only if

$$\frac{\Delta V^*(\lambda)}{\Phi}\bigg|_{\lambda=1} = -\frac{\Delta V^*(\lambda)}{\Phi}\bigg|_{\lambda=0} = 2\sigma\phi_\tau - \frac{(\mu+\sigma)\phi_\tau^2 + (\sigma-\mu)}{\phi_\tau^{\frac{\mu}{1-\sigma}}} > 0,$$

Therefore, the sustain point ϕ_τ^s may be defined as the value of ϕ_τ that equates the above expression to zero and full agglomeration can be sustained for all $\phi_\tau \geq \phi_\tau^s$. Moreover, there are at most three interior equilibria in the closed Domestic economy case (Robert-Nicoud, 2005), of which the symmetric one ($\lambda^* = 1/2$) always exists. The stability of the equilibrium $\lambda^* = 1/2$ depends on the sign of the derivative of the indirect utility differential, whereas the other two interior equilibria are always unstable. Computing $\partial(\Delta V^*)/\partial\lambda$ and evaluating it at $\lambda = 1/2$, the break-point is such that

$$\phi_\tau^b \equiv \frac{\sigma-\mu}{\sigma+\mu}\frac{\mu-\sigma+1}{1-\mu-\sigma}.$$

Hence, $\lambda^* = 1/2$ is a stable spatial equilibrium for all $\phi_\tau \leq \phi_\tau^b$. Note, finally, that both types of equilibria occur for values $\phi_\tau^s \leq \phi_\tau \leq \phi_\tau^b$, in which case both full agglomeration and full dispersion are stable spatial equilibria.

5.3.4 Case of an open domestic economy

The general case of an open economy yields a much longer and more complicated expression of the indirect utility differential. As a consequence we can not characterize spatial equilibrium analytically.[7] We therefore rely on simulations to provide sensible results.[8] Figure 5.1 shows the real wage differential in case of a close economy and for a high value of interregional transport cost. It indicates that the symmetric equilibrium is stable since the real wage differential is zero for $\lambda = 0.5$ and it has a negative slope at that point.

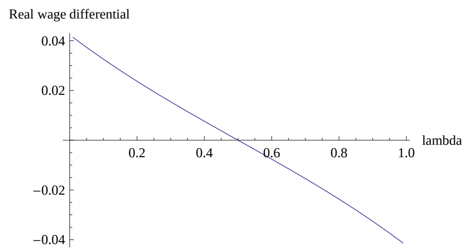

Figure 5.1: Real wage differential in case of high transport costs ($\tau = 2.1$) and a closed economy ($\phi_\rho = 0$)

Allowing for international trade implies a different story. Figure 5.2 displays a graph of the real wage differential as a function of the spatial distribution of firms in the Domestic economy in case of the absence of any international trade friction. It shows that there is not any interior equilibrium; the only equilibrium being agglomeration in the Hub. International trade induces an asymmetry between the two regions of the market, since the latter have different accessibility to the foreign market and domestic firms prefer to locate in the region having the higher market potential.

In the case of low ($\rho = 1.2$) or even relatively high international trade costs ($\rho = 3$), the same story holds: no interior equilibrium and agglomeration in the Hub (Figure 5.3).

[7] Ago *et al.* (2006) were able to obtain meaningful analytical results in a three regions model à la Krugman featuring symmetry. However, in a asymmetric framework things are much more involved

[8] To perform those simulations we gave the following values to model parameters: $\mu = 0.4$, $\sigma = 2.7$, $H = 10$, $H_D = 1$, $L_D = 0.852$, $H_F = 9$, $L_F = 7.668$, $F = 1$, $m = 0.4$. Values of H_D and L_D have been set conforming to standard normalization in the FE model (Baldwin *et al.*, 2003). We assume that the Domestic Economy and the Rest of the World have the same ratio of unskilled to skilled workers. The values of $H_F = 9$ and $H = 10$ come from the fact that conforming to World Bank data on labor force, the mass of workers in the Rest of the World is about tenfold that of SSA (the total labor force respectively in World and in SSA were in 2006 $3,077.9$ and 322.8 billions. Data are available online on the World Bank Website http://www.worldbank.org)

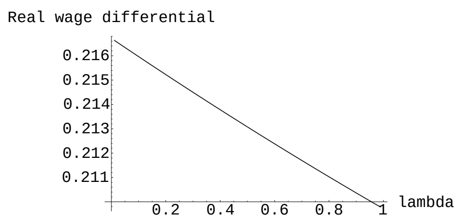

Figure 5.2: Real wage differential in case of high transport costs ($\tau = 2.1$) and an open economy ($\phi_\rho = 1$)

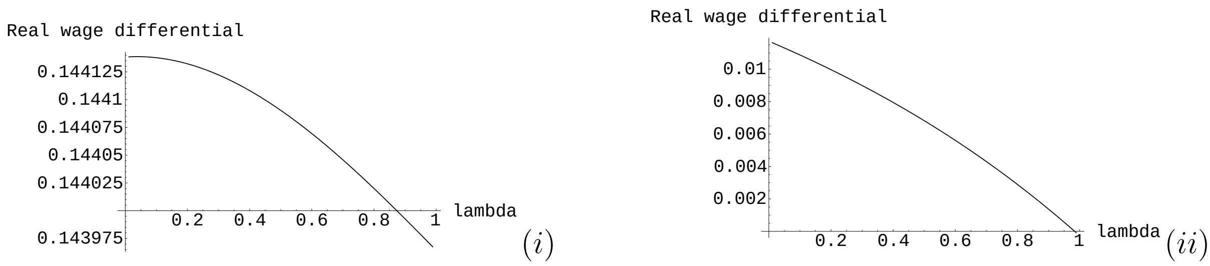

Figure 5.3: Real wage differential in case of high transport costs ($\tau = 2.1$) and with respectively low trade costs ($\rho = 1.2$, panel (i)) and high trade costs ($\rho = 3$, panel (ii))

It is only for much higher international trade costs that a stable interior spatial equilibrium appears. But this spatial equilibrium may not be the symmetric one. Indeed, with $\rho = 3.5$ the spatial equilibrium implies a share of skilled workers slightly greater than one half in the hub. But for higher trade costs, for instance $\rho = 4$ (respectively $\rho = 5$) the spatial equilibrium implies a value of λ equal to 1/3 (respectively 26%) of skilled workers in the gated region (Figure 5.4). Therefore, in case of an open economy, very high trade costs induce spatial equilibrium with partial agglomeration in the hinterland. The rationale of this localisation of most firms in the hinterland is that, because of higher trade costs, the market potentiel of firms in the hub is reduced. Therefore, the gated region is less profitable and interregional transport costs provide hinterland firms a good protection against foreign firms competition.

The role of high interregional transports as a shield against foreign competition is further emphasized in simulations with very high transport costs ($\tau = 5$). In this case we find as before that, for a closed Domestic economy, the symmetric distribution of firms is as previously the only stable equilibrium and that in case of the lack of any international trade friction there is no interior equilibrium. But with higher trade costs partial agglomeration in the hinterland occurs for lower values of international trade frictions than it was the case before. Figure 5.5 shows that for $\rho = 2.5$, a stable partial equilibrium occurring with a

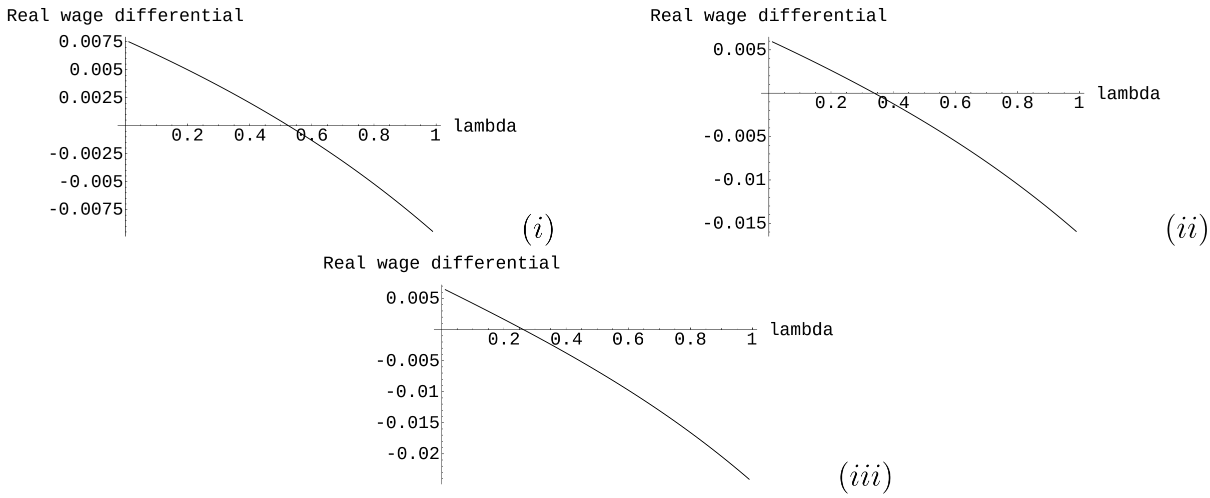

Figure 5.4: Real wage differential in case of high transport costs ($\tau = 2.1$) and with very high trade costs ($\rho = 3.5$ for panel (i), $\rho = 4$ for panel (ii), and $\rho = 5$ for panel (iii))

value of λ slightly below 20%. So the higher are transport costs, the more they provide protection against foreign firms competition.

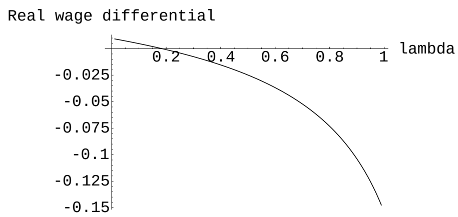

Figure 5.5: Real wage differential for very high transport costs ($\tau = 5$) and high trade costs ($\rho = 2.5$)

Those numerical examples provide interesting results. They show up that openness with low international trade costs provide good impulse for hub agglomeration. However, once those trade costs exceed some threshold, the hinterland became the favorite location for most of firms and the higher are transport costs, the lower are those thresholds. However, simulations can not provide a complete gallery of results.[9] So now we consider results

[9]One may for instance want to check what would be the spatial structure of a much smaller domestic economy. Simulations with mass of foreign skilled workers ninety nine times greater than mass of domestic skilled workers imply systematically full agglomeration in the hub. The intuition behind such a result is that more firms in the rest of the World intensifies competition in the same way as lower trade costs. When the rest of the World is large relatively to the domestic economy, locating in the landlocked region does not protect firms anymore, so agglomeration takes place

provided by the OTT Model which allows for more analytical tractability.

5.4 OTT Model

Some underlying assumptions of the Footloose Entrepreneur Model implies, as it is the case for the classical Core Periphery Model, several shortcomings. Firstly it entails equilibrium prices that are independent of the spatial distribution of firms and consumers, a result that conflicts with research in spatial pricing theory that shows that demand elasticity varies with distance while prices change with the level of demand and the intensity of competition. Secondly the iceberg assumption implies the unrealistic result that any increase in the price of the transported good is accompanied by a proportional increase in its trade cost.

Referring to Ottaviano *et al.* (2002), we use another modeling strategy that is short of these drawbacks. It is based on quadratic utility and on additive transport costs that are not incurred in the good itself. This allows us to derive analytically the results previously obtained.

5.4.1 Preferences

We assume that each worker is endowed with one unit of labor and $\bar{q}_0 > 0$ units of the numéraire and the initial endowment $\bar{q}_0$ is large enough for her consumption of the numéraire to be strictly positive at the market outcome.

Consumers have identical preferences described by a quasi-linear utility with a quadratic subutility. Therefore, a typical resident of region i faces the following consumption problem:

$$\max_{q_i(\omega),\omega\in[0,N]} \quad \alpha \int_0^N q_i(\omega)\, d(\omega) - \frac{\beta-\gamma}{2}\int_0^N q_i(\omega)^2 d(\omega) - \frac{\gamma}{2}\left[\int_0^N q_i(\omega) d(\omega)\right]^2 + q_0$$

$$\text{s.t.} \quad \int_0^N p_i(\omega)\, q_i(\omega)\, d(\omega) + q_0 = y_i + \bar{q}_0$$

where $\alpha > 0$ and $\beta > \gamma > 0$ are parameters, $p_i(\omega)$, $q_i(\omega)$ are respectively consumer price and quantity of variety ω in region i and y_i is the individual's labor income in region i.

in the gate. This gives some ground to the empirical result that smaller countries have larger primate cities. We thank Kristian Behrens for this intuition. Adjusting for labor productivity may also be relevant as labor productivity in Sub-Saharan Africa is the twelfth of that of developed nations (`http://www.ilo.org/public/english/employment/strat/kilm/download/kilm18.pdf`). Such an adjustement would increase the relative size of the rest of the World and thus would magnify hub locational advantages.

Assuming that all varieties produced in each region are symmetric, we may alleviate notation by dropping the variety index ω. Considering that q_{ij} denotes the output of a firm located in region i demanded by a consumer in region j, it is readily verified that the individual demand functions are given by:

$$q_{ij} = a - (b + cN)\, p_{ij} + cP_j \tag{5.19}$$

where

$$a = \frac{\alpha}{\beta + (N-1)\,\gamma}, \quad b = \frac{1}{\beta + (N-1)\,\gamma}, \quad c = \frac{\gamma}{(\beta - \gamma)\,(\beta + (N-1)\,\gamma)}$$

p_{ij} is the price a firm located in region i charges to consumers in region j and with

$$P_j = \sum_{i\,\in\,\{0,1,2\}} n_{ij} p_{ij}. \tag{5.20}$$

the price index of varieties in region j.

5.4.2 Technology

As in the Footloose Entrepreneur Model each firm of the manufacturing sector requires a constant amount of skilled labor, denoted hereafter by ϕ, as a fixed input requirement and m units of unskilled labor per unit of output as a variable input requirement. Without loss of generality, we set $m = 0$ in what follows. As demand functions are linear, this amounts to rescaling firms' demand intercepts (Ottaviano *et al.*, 2002). Given the technology in the modern sector, skilled labor market clearing requires the following masses in each region $i = 0, 1, 2$:

$$n_0 = n_F = \frac{H_F}{\phi}, \quad n_1 = \lambda \frac{H_D}{\phi}, \quad n_2 = (1-\lambda)\,\frac{H_D}{\phi}, \quad n_D = n_1 + n_2, \quad N = n_D + n_F \tag{5.21}$$

Making the standard assumptions that product markets are segmented, that labor markets are local and that firms bear all trade and transportation costs, firms in regions 0, 1 and 2 maximize profit given respectively by:

$$\begin{aligned}
\pi_0 &= p_{00}q_{00}\left(L_F+H_F\right)+\left(p_{01}-\tau\right)q_{01}\left(\frac{L_D}{2}+\lambda H_D\right) && (5.22)\\
&+ \left(p_{02}-\left(t+\tau\right)\right)q_{02}\left(\frac{L_D}{2}+\left(1-\lambda\right)H_D\right)-\phi w_0 \\
\pi_1 &= \left(p_{10}-\tau\right)q_{10}\left(L_F+H_F\right)+p_{11}q_{11}\left(\frac{L_D}{2}+\lambda H_D\right) && (5.23)\\
&+ \left(p_{12}-t\right)q_{12}\left(\frac{L_D}{2}+\left(1-\lambda\right)H_D\right)-\phi w_1 \\
\pi_2 &= \left(p_{20}-\left(\tau+t\right)\right)q_{20}\left(L_F+H_F\right)+\left(p_{21}-t\right)q_{21}\left(\frac{L_D}{2}+\lambda H_D\right) && (5.24)\\
&+ p_{22}q_{22}\left(\frac{L_D}{2}+\left(1-\lambda\right)H_D\right)-\phi w_2
\end{aligned}$$

Given those profits functions, profit-maximizing prices are as follows:

1. Intraregional prices

$$p_{ii}=\frac{a+cP_i}{2\left(b+cN\right)} \tag{5.25}$$

2. Interregional prices

$$p_{ij}=p_{jj}+\frac{t}{2} \tag{5.26}$$

   with $i,j\neq 0$

3. International prices

$$p_{ij}=p_{jj}+\frac{\tau_{ij}}{2} \tag{5.27}$$

   with i or $j=0$, $i\neq j$ and

$$\begin{aligned}
\tau_{01} &= \tau_{10}=\frac{\tau}{2}\\
\tau_{02} &= \tau_{20}=\frac{t+\tau}{2}
\end{aligned}$$

We may notice that the price a firm sets in a region depends on the price index P_i of this region, which depends itself on the prices set by all other firms. Since there is a continuum of firms, each firm is negligible and considers aggregate market conditions as given when setting its optimal price. But these aggregate market conditions must be consistent with firms' optimal pricing decisions. Hence, the (Nash) equilibrium price indices must satisfy the following equilibrium conditions:

$$P_0 = n_0 p_{00} + n_1 p_{10} + n_2 p_{20} \tag{5.28}$$
$$P_1 = n_0 p_{01} + n_1 p_{11} + n_2 p_{21} \tag{5.29}$$
$$P_2 = n_0 p_{02} + n_1 p_{12} + n_2 p_{22} \tag{5.30}$$

The equilibrium price indices can be found by solving (5.28) – (5.30) using expressions (5.25) – (5.27). This yields:

$$P_0 = \frac{aN + (b + cN)\left((n_1 + n_2)\tau + n_2 t\right)}{2b + cN} \tag{5.31}$$
$$P_1 = \frac{aN + (b + cN)(n_0\tau + n_2 t)}{2b + cN} \tag{5.32}$$
$$P_2 = \frac{aN + (b + cN)(n_0\tau + (n_0 + n_1)t)}{2b + cN} \tag{5.33}$$

Substituting (5.31) – (5.33) into (5.25) gives the intraregional prices:

$$p_{00} = \frac{2a + c(n_D\tau + (1-\lambda)n_D t)}{2(2b + cN)} \tag{5.34}$$
$$p_{11} = \frac{2a + c(n_F\tau + (1-\lambda)n_D t)}{2(2b + cN)} \tag{5.35}$$
$$p_{22} = \frac{2a + c(n_F\tau + (\lambda n_D + n_F)t)}{2(2b + cN)} \tag{5.36}$$

Up to now we have implicitly assumed that trade and transport costs are sufficiently low for interregional and international trade to be bilateral, regardless of firm distributions. We precise now the conditions on t and τ for trade to occur between any two regions at these equilibrium prices.

For interregional transport costs between Region 1 and Region 2, the following conditions

$$t \leq t_{12}^{trade} = \frac{2a\phi + c\tau H_F}{2b\phi + c(1-\lambda)H_D} \tag{5.37}$$
$$t \leq t_{21}^{trade} = \frac{2a\phi + c\tau H_F}{2b\phi + c(H_F + \lambda H_D)} \tag{5.38}$$

must hold for trade to occur respectively between Region 1 and Region 2 and between Region 2 and Region 1.

Evaluating (5.38) for $\lambda = 1$ yields the most stringent conditions for trade threshold concerning flows from region 1 to region 2 that hold for any spatial repartition of firms in the Domestic economy:

$$t_{21}^{trade} \to \frac{2a\phi + c\tau H_F}{2b\phi + c(H_F + H_D)}. \tag{5.39}$$

Those expressions outline clearly that trade between the two regions of the domestic economy is asymmetric. Indeed, requirements for trade between Region 2 and Region 1 are more stringent than between Region 1 and Region 2 reflecting the locational advantage of the hub and its status as the favored domestic region for trade. However, for both directions trade thresholds decrease with the value of trade costs. This implies that lower international trade costs may lead to a break down of internal trade when the regional markets of a country are poorly integrated, especially when one of the two regions has a good access to the international marketplace, an intuitive result already put forward by Behrens *et al.* (2006). This captures the fact that consumers tend to prefer cheaper imports to more expensive nationally produced ones.

Considering international trade costs between Region 0 and Region 1, the following conditions

$$\tau \leq \tau_{01}^{trade} = \frac{2a\phi + c(1-\lambda)H_D t}{2b\phi + cH_F} \tag{5.40}$$

$$\tau \leq \tau_{10}^{trade} = \frac{2a\phi + c(1-\lambda)H_D t}{2b\phi + c(H_F + \lambda H_D)} \tag{5.41}$$

apply. Once more those conditions point out trade asymmetry. For values of $\lambda > 0$, fulfilling conditions for existence of international trade is more involved from region 1 to region 0 than the other way round. It is therefore easier for the Rest of the World to export.

Nevertheless, we can see from (5.40) and (5.41) that for both directions the feasibility of international trade improves when interregional transport costs are high, and when concentration of firms in the hub is weak. As Behrens *et al.* (2007) point out this is because lower transport costs and firms' agglomerations exacerbate price competition in local markets, thus making penetration by outside firms more difficult.

Finally for trade to occur between regions 0 and 2, the following conditions have to be fulfilled

$$\tau \leq \tau_{02}^{trade} = \frac{2a\phi - (2b\phi + cH_F + c(1-\lambda)H_D)t}{2b\phi + cH_D} \tag{5.42}$$

$$\tau \leq \tau_{20}^{trade} = \frac{2a\phi - (2b\phi + cH_F + c\lambda H_D)t}{2b\phi + cH_F} \tag{5.43}$$

We can deduce from (5.42) and (5.43) that the higher are transport costs, the less feasible is trade between regions 0 and 2. Thus, costly interregional trade induces the closeness of the Hinterland with respect to International Trade.

It is easy to verify that the equilibrium gross profits earned by a firm established in region 1 on each separated market are as follows:

$$\pi_{11} = (b+cN)\left(\frac{L_D}{2}+\lambda H_D\right)p_{11}^2 \tag{5.44}$$

$$\pi_{10} = (b+cN)(L_F+H_F)(p_{10}-\tau)^2 \tag{5.45}$$

$$\pi_{12} = (b+cN)\left(\frac{L_D}{2}+(1-\lambda)H_D\right)(p_{12}-t)^2 \tag{5.46}$$

Because of the specifity of our set up, profit earned by hinterland firms in the rest of World is asymmetric with respect to hub firms. This is readily shown in the following expressions:

$$\pi_{22} = (b+cN)\left(\frac{L_D}{2}+(1-\lambda)H_D\right)p_{22}^2 \tag{5.47}$$

$$\pi_{20} = (b+cN)(L_F+H_F)(p_{20}-(t+\tau))^2 \tag{5.48}$$

$$\pi_{21} = (b+cN)\left(\frac{L_D}{2}+\lambda H_D\right)(p_{21}-t)^2 \tag{5.49}$$

Concerning local labor markets, the equilibrium wages of the skilled are determined by the standard bidding process in which firms compete for workers by proposing higher wages until no firm can profitably enter or exit the market. Consequently all operating profits are absorbed by the wage bill. Therefore, in equilibrium the skilled wage rate in region i of the Domestic Economy satisfies the condition $\pi_i(w_i)=0$ which yields the following expressions for skilled wages in the Domestic economy:

$$\begin{aligned} w_1(\lambda) = & \left(\frac{(b\phi+cL)\phi}{4(2b\phi+cL)^2}\right)\left[\left(\frac{L_D}{2}+\lambda H_D\right)(2a\phi+c((1-\lambda)L_Dt+L_F\tau))^2\right. \\ & +\left(\frac{L_D}{2}+(1-\lambda)H_D\right)(2a\phi+c((\lambda L_D+L_F)t+L_F\tau)-(2b\phi+c)t)^2 \\ & \left.+(L_F+H_F)(2a\phi+c((1-\lambda)L_Dt+L_D\tau)-(2b\phi+c)\tau)^2\right], \end{aligned} \tag{5.50}$$

$$\begin{aligned} w_2(\lambda) = & \left(\frac{(b\phi+cL)\phi}{4(2b\phi+cL)^2}\right)\left[\left(\frac{L_D}{2}+\lambda H_D\right)(2a\phi+c((1-\lambda)L_Dt+L_F\tau)-(2b\phi+c)t)^2\right. \\ & +\left(\frac{L_D}{2}+(1-\lambda)H_D\right)(2a\phi+c((\lambda L_D+L_F)t+L_F\tau))^2 \\ & \left.+(L_F+H_F)(2a\phi+c((1-\lambda)L_Dt+L_D\tau)-(2b\phi+c)(t+\tau))^2\right] \end{aligned} \tag{5.51}$$

The individual consumer surplus in region 1 associated with the equilibrium prices p_{11}, p_{21}, and p_{F1} is given by:

$$\begin{aligned} S_1(\lambda) = & \frac{a^2H}{2b\phi}-\frac{aH_D}{\phi}\left[\lambda p_{11}+(1-\lambda)p_{21}+\frac{H_F}{H_D}p_{F1}\right] \\ & +\frac{(b+cN)H_D}{2\phi^2}\left[\lambda p_{11}^2+(1-\lambda)p_{21}^2+\frac{H_F}{H_D}p_{F1}^2\right] \\ & -\frac{cH_D^2}{2\phi^2}\left[\lambda p_{11}^2+(1-\lambda)p_{21}^2+\frac{H_F}{H_D}p_{F1}^2\right]^2, \end{aligned} \tag{5.52}$$

a symmetric expression holds for region 2. Mobile skilled workers living in the Domestic Economy move to the region offering the highest indirect utility. The indirect utility of a skilled worker living in region $i \in \{1, 2\}$ is given by:

$$V_i(\lambda) = S_i(\lambda) + w_i(\lambda) + \bar{q}_0$$

5.4.3 Benchmark case: a domestic economy without the Rest of the World

In case of the insignificance of the Rest of the World, $L_F \to 0$ and $H_F \to 0$, our set-up reduces to the one of Ottaviano *et al.* (2002). Indeed, after straightforward calculations the following indirect utility differential is obtained:

$$\begin{aligned} \Delta V(\lambda) &\equiv V_1(\lambda) - V_2(\lambda) = S_1(\lambda) - S_2(\lambda) + w_1(\lambda) - w_2(\lambda) \\ &= Ct(t^* - t)\left(\lambda - \frac{1}{2}\right) \end{aligned} \tag{5.53}$$

where

$$C = \left[2b\phi(3b\phi + 3cH + cL) + c^2H(L + H)\right]\frac{H(b\phi + cH)}{2\phi^2(2b\phi + cH)^2} > 0$$

and

$$t^* = \frac{4a\phi(3b\phi + 2cH)}{2b\phi(3b\phi + 3cH + cL) + c^2H(L + H)}$$

It is clear for (5.53) that $\lambda = 1/2$ is always an equilibrium. As $C > 0$, for $\lambda \neq 1/2$, the indirect utility differential has always the same sign as $(\lambda - 1/2)$ whenever $t < t^*$; otherwise it has the opposite sign. Thus, when $t < t^*$, the symmetric equilibrium is unstable and workers agglomerate in region 1 (2) provided that the initial fraction of skilled workers residing in this region is greater than 1/2. This yields the standard result obtained by Krugman (1991) that agglomeration arises when transport costs are low enough.

5.4.4 Spatial equilibrium analysis in the general case

Assuming that $L_F \neq 0$ and / or $H_F \neq 0$, after cumbersome but straightforward calculations the utility differential can be written as follows

$$\begin{aligned} \Delta V(\lambda) &\equiv V_1(\lambda) - V_2(\lambda) = S_1(\lambda) - S_2(\lambda) + w_1(\lambda) - w_2(\lambda) \\ &= \frac{t}{8\phi^2(cH + 2b\phi)^2}\left[2\left(\lambda - \frac{1}{2}\right)(\eta_1 + \eta_2\tau - \eta_3 t) + (\eta_4 - \eta_5\tau - \eta_6 t)\right] \end{aligned} \tag{5.54}$$

where

$$\eta_1 = 16a\phi H_D(b\phi + cH)(3b\phi + 2cH) > 0 \tag{5.55}$$

$$\eta_2 = 4cH_D H_F(b\phi + cH)(4b\phi + 3cH) > 0 \tag{5.56}$$

$$\eta_3 = 4H_D(b\phi + cH)\left(2bc\phi(L + 3H) + c^2H(L + H) + 6b^2\phi^2\right) > 0 \tag{5.57}$$

$$\eta_4 = 8a\phi(b\phi + cH)\left(L_F(2b\phi + cH) + H_F(3b\phi + 2cH)\right) > 0 \tag{5.58}$$

$$\begin{aligned}\eta_5 = {} & 2(b\phi + cH)\left(2L_F(2b\phi + cH)\left(2b\phi + cH_F\right) + H_F\left(cH_F(4b\phi + cH)\right.\right.\\ & \left.\left. +(2b\phi + cH)(6b\phi + cH)\right)\right) > 0\end{aligned} \tag{5.59}$$

$$\eta_6 = 2(b\phi + cH)\left(H_F\left(2bc\phi(L + 3H) + c^2H(L + H) + 6b^2\phi^2\right)\right. \tag{5.60}$$

$$\left. +2b\phi L_F(2b\phi + cH)\right) > 0 \tag{5.61}$$

where η_1, η_2, η_3, η_4, η_5 and η_6 are positive bundles of parameters independent of the distribution of domestic firms.

Since the indirect utility differential includes an additional term independent of $(\lambda - 1/2)$, the conditions for the prevalence of the symmetric equilibrium in the domestic economy are much more restrictive in this framework than in the benchmark OTT model.[10] Thus, in this set-up dispersion is almost never an equilibrium.

The stringency of the conditions required to allow dispersion to be an equilibrium outlines the peculiarity of our set-up. With one region benefiting of a locational advantage, the framework is asymmetric and dispersion is no longer the 'natural' equilibrium it was in the Ottaviano *et al.* (2002) framework. Therefore, our set-up is characterized by a 'bias' towards (at least partial) agglomeration in either Domestic region.

The analysis of the impact of transport and trade costs on the spatial distribution of the domestic economy may be developed through four subcases as shown in figure (5.6).[11] We will focus mostly on the analysis of the two subcases corresponding to Regions I and II in figure (5.6) since they provide the most interesting and clear-cut analytical results. The first one implies that the following conditions hold simultaneously.

$$(\eta_1 + \eta_2\tau - \eta_3 t) \geq 0 \tag{5.62}$$

$$(\eta_4 - \eta_5\tau - \eta_6 t) \geq 0 \tag{5.63}$$

[10] The explicit conditions for the existence and the stability of a symmetric equilibrium are the following: $(\eta_1 + \eta_2\tau - \eta_3 t) < 0$ and $(\eta_4 - \eta_5\tau - \eta_6 t) = 0$. It is clear that they are hardly fulfilled simultaneously.

[11] As $\frac{\eta_1}{\eta_3} < \frac{\eta_4}{\eta_6}$ the two lines in figure (5.6) cross. Therefore, they divide the first quadrant in four regions corresponding to our four subcases.

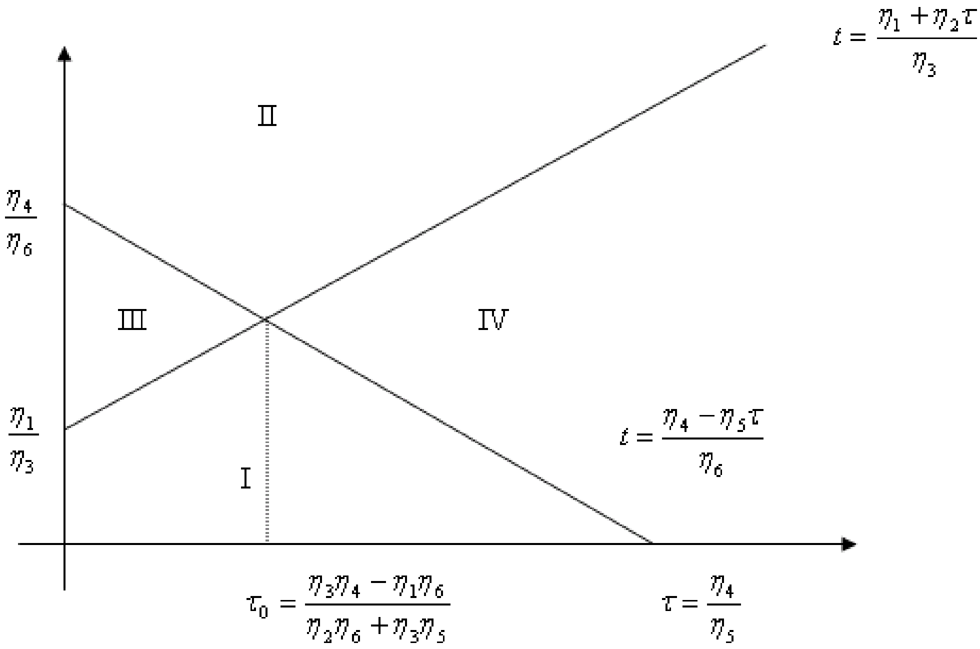

Figure 5.6: Trade and transport costs

Depending on the level of trade costs, we may express those conditions in terms on different transport thresholds. When

$$\tau < \tau_0 = \frac{\eta_3\eta_4 - \eta_1\eta_6}{\eta_2\eta_6 + \eta_3\eta_5}, \tag{5.64}$$

(5.62) is the more stringent condition.[12] It implies that :

$$t < \frac{\eta_1 + \eta_2\tau}{\eta_3}. \tag{5.65}$$

Conversely, when $\tau > \tau_0$, (5.63) is more stringent and implies that:

$$t < \frac{\eta_4 - \eta_5\tau}{\eta_6}. \tag{5.66}$$

Conditions (5.65) and (5.66) entail relatively low values of interregional transport costs. In this first case, agglomeration in the hub is an equilibrium. Indeed, for any value of $\lambda > \frac{1}{2}$ the indirect utility differential is positive.

This prompts to the first proposition of this chapter:

[12] τ_0 is the value of trade costs where the two lines of figure (5.6) cross.

Proposition 5 (Transport Costs, Trade Costs and Hub Agglomeration) *When $\tau < \tau_0$ (respectively $\tau > \tau_0$), for any transport costs values satisfying (5.65) (respectively (5.66)), Hub agglomeration is an equilibrium.*

The case where (5.62) and (5.63) hold doesn't discard partial agglomeration in the hinterland. Assuming that the values of η_1 to η_6 bundles of parameters are such that (5.54) may be equal to 0, the model admits the following equilibrium

$$\lambda^* = \frac{1}{2} - \frac{(\eta_4 - \eta_5\tau - \eta_6 t)}{2(\eta_1 + \eta_2\tau - \eta_3 t)} \tag{5.67}$$

with $\lambda^* < \frac{1}{2}$, which implies partial agglomeration in the hinterland.[13] However, as (5.62) holds this equilibrium is unstable, short deviations from it may drive the economy towards full agglomeration in the hub.

The second subcase corresponds to the situation where conditions (5.62) and (5.63) are both violated (Region II in figure (5.6)), i.e.:

$$(\eta_1 + \eta_2\tau - \eta_3 t) < 0$$
$$(\eta_4 - \eta_5\tau - \eta_6 t) < 0$$

It implies, when $\tau < \tau_0$, transport costs exceeding the upper treshold of condition (5.66). When $\tau > \tau_0$, it rather implies friction costs above the upper bound of condition (5.65).

In this case, Hinterland agglomeration may be an equilibrium provided that the indirect utility differential evaluated at $\lambda = 0$ is negative that is:

$$-(\eta_1 + \eta_2\tau - \eta_3 t) + (\eta_4 - \eta_5\tau - \eta_6 t) < 0$$

which implies that

$$\tau > \frac{\eta_4 - \eta_1 + (\eta_3 - \eta_6)t}{\eta_2 + \eta_5} \tag{5.68}$$

Therefore, high transportation and trade costs induce full agglomeration in the hinterland. Protected by the relative closeness induced by expensive domestic trade and without a good access to international markets, domestic firms prefer locating to the hinterland.

Proposition 6 (Transport costs, trade costs and agglomeration in the hinterland) *Any values of transport and trade costs violating both conditions* (5.62) *and* (5.63) *and fullfilling condition* (5.68) *induce Hinterland agglomeration.*

[13]This solution is acceptable (i.e. $\lambda^* > 0$) if trade costs exceed the following treshold: $\tau > \frac{\eta_4 - \eta_1 + (\eta_3 - \eta_6)t}{\eta_2 + \eta_5}$.

If condition (5.68) doesn't hold, the model admits a spatial distribution, $\lambda^* < 1/2$, implying partial agglomeration in the hinterland. This spatial equilibrium is stable since (5.62) doesn't hold. Therefore, when trade integration prevails in a context of market fragmentation, a minority of firms remain in hub to benefit from its higher market potential.

Proposition 7 (Transaction costs and partial agglomeration in the hinterland) *For any values of transport and trade costs violating both conditions* (5.62)*,* (5.63) *and* (5.68)*, partial agglomeration in the Hinterland is a stable equilibrium.*

We still have to determine when η_1/η_3 is lower than t_{21}^{trade}, so that for any transport costs below that treshold, bilateral trade between the two domestic regions is possible. It is the case when the mass of unskilled workers is sufficiently large so that:

$$L > \frac{c\tau H_F \left(6b^2\phi^2 + 6bcH\phi + c^2H^2\right) - 2a\phi \left(6b^2\phi^2 + 8bcH\phi + 3c^2H^2\right)}{c(2b\phi + cH) \left(2a\phi + c\tau H_F\right)}$$

Such a condition is likely to be fullfilled since the fraction of highly educated people represented only 14.3 % of the world total population in 2000 (cfr Barro and Lee (2001, Table 3)).

Propositions (5) and (6) yield opposite results. When trade and transport costs are low, the hub is the preferential location for firms because of its better access to either market. Conversely, when they are high hub firms have only a poor access to the foreign market. Therefore, firms prefer to locate in the hinterland in order to take advantage of the protection granted by high transportation costs.

Proposition (7) describes the spatial equilibrium prevailing in case of trade integration coupled with substantial interregional friction barriers. In this configuration, the hub remains attractive to a minority of firms because of its higher market potentiel. However most of firms will relocate in the hinterland to avoid a fierce price competition.

5.5 Conclusion

In this chapter we rely on international trade and gate effects to explain the formation of agglomerations in SSA hubs. According to the two models developed, openness is likely to trigger agglomeration in the hub especially when transport costs are low. This result is consistent with Weber's theory of location (Beckmann and Thisse, 1986) which states that in a star-shaped network without any dominant location, entry points are the optimal locations.[14]

Those results shed light on agglomeration processes in SSA. Indeed, as described in the introduction SSA is characterized by the stability of spatial concentration of economic activities along coastal locations. This persistence of the location of several of its biggest cities in hubs is quite appealing, especially in the context of increasing trade integration facilitated by the increasing efficiency of transport technologies and by the general decrease of tariffs. It discards Krugman and Elizondo (1996)'s model which explains urban concentration in developing countries by backward and forward linkages triggered by import-substitution policies and predicts that urban concentration may shrink with openness. Modeling explicitly locational disadvantages may therefore be one way to provide a convincing explanation of spatial distribution in SSA.

One relevant issue has still to be discussed: the impact of the magnitude of interregional transport costs on agglomeration location. According to the three regions OTT model we developed, increasing trade integration may induce agglomeration on the hub provided that transport costs are not too high. Otherwise, according to Proposition (7), partial agglomeration will take place in the hinterland. The last prediction while intuitive doesn't seem to have been backed by any empirical evidence: in countries of SSA having access to the sea primate cities are located at the hub despite the high transport costs characterizing most of countries of that region. Kenya is the only country where the primate and capital city moved from the port of Mombasa to Nairobi, an interior location. But as Nairobi is the hub of the Kenyan transportation network, this Kenyan exception is not really a genuine one (Obudho, 1997). The model à la OTT would rather imply that firms would have moved to the hinterland to escape from price competition. The persistence of coastal locations of primate city challenges this prediction.

One way to go through that problem is to recall that in the real world all the factors favoring the emergence of urban agglomerations are mixed: political factors, scale economies and locational (dis)advantages as well. It is therefore difficult to find a real situation where hub effects are the only at play. The primate city in a typical african country is often at the same time the capital city, the hub and the nascent industrial center. Therefore, the political and administrative role of primate port cities induces a lock in effect that may explain their persistence in spite of very high transport costs. This argument yields an appeal for a model encapsulating both political and hub effects. We wish to develop it in a future research.

[14]The network we consider here : a segment line with three locations is a degenerate form of a star network.

5.6 Appendix A: Colonial Capitals in Africa

Figure 5.7: Colonial Capitals in Africa. Source: Christopher (1994)

5.7 Appendix B: Change in National Capitals

Figure 5.8: Change in National Capitals, 1900-1991. Source: Christopher (1994)

5.8 Appendix C: FE model market outcome solutions

Replacing net aggregate incomes are given by expression (5.3), into wage expressions given by (5.15), (5.16) and (5.17) yields the following solutions in w_0, w_1 and w_2:

$$
\begin{aligned}
w_0 &= \frac{1}{D(\phi_\tau,\phi_\rho)}\mu\left(-2\sigma^2 H_F^2 L\phi_\rho^2\phi_\tau + \sigma H_D H_F \phi_\rho\left((1-\lambda)(\mu-\sigma)(L+L_F)\right.\right.\\
&\quad -2\lambda\left(\sigma L_D\left(1+\phi_\rho^2\right)+L_F\left(-\mu+2\sigma+\mu\phi_\rho^2\right)\right)\phi_\tau + (-1+\lambda)\left(2L_F\left(\sigma+\mu\phi_\rho^2\right)\right.\\
&\quad \left.\left.+L_D\left(\mu+\sigma+2\sigma\phi_\rho^2\right)\right)\phi_\tau^2\right) + H_D^2\left(-\left(\sigma L_D\phi_\rho^2(-\lambda+(-1+\lambda)\phi_\tau)\left((1-\lambda)(\mu-\sigma)\right.\right.\right.\\
&\quad \left.\left.-2\lambda\sigma\phi_\tau + (-1+\lambda)(\mu+\sigma)\phi_\tau^2\right)\right) + 2L_F\left((1-\lambda)\lambda(\mu-\sigma)\left(-\mu+\sigma+\mu\phi_\rho^2\right)\right.\\
&\quad -\sigma\left((-1-2(-1+\lambda)\lambda)(\mu-\sigma)+\lambda^2\mu\phi_\rho^2\right)\phi_\tau\\
&\quad \left.\left.\left.+(-1+\lambda)\lambda(\mu+\sigma)\left(-\mu+\sigma+\mu\phi_\rho^2\right)\phi_\tau^2-(-1+\lambda)^2\mu\sigma\phi_\rho^2\phi_\tau^3\right)\right)\right)\\
w_1 &= \frac{1}{D(\phi_\tau,\phi_\rho)}\mu\left(-2\sigma H_F^2\phi_\rho\left(\sigma L_F\phi_\rho^2+L_D\left(-\mu+\sigma+\mu\phi_\rho^2\right)\right)\phi_\tau\right.\\
&\quad -\sigma H_D^2\phi_\rho(-\lambda+(-1+\lambda)\phi_\tau)\left((1-\lambda)(\mu-\sigma)(L+L_F)-2\lambda\sigma L\phi_\tau\right.\\
&\quad \left.+(-1+\lambda)((\mu+\sigma)L_D+2\mu L_F)\phi_\tau^2\right) + H_D H_F\left((-1+\lambda)(-\mu+\sigma)\left(2\sigma L_F\phi_\rho^2\right.\right.\\
&\quad \left.+L_D\left(-\mu+\sigma+\mu\phi_\rho^2\right)\right) - 2\lambda\sigma\left(2\sigma L_F\phi_\rho^2+L_D\left(-\mu+\sigma+(\mu+\sigma)\phi_\rho^2\right)\right)\phi_\tau\\
&\quad \left.\left.+(-1+\lambda)\left(2\sigma(\mu+\sigma)L_F\phi_\rho^2+L_D\left(-\mu^2+\sigma^2+\left(\mu^2+\mu\sigma+2\sigma^2\right)\phi_\rho^2\right)\right)\phi_\tau^2\right)\right)\\
w_2 &= \frac{1}{(\lambda-1)D(\phi_\tau,\phi_\rho)}\left(-L_D-\sigma\left(H_F\phi_\rho\phi_\tau+H_D(1-\lambda+\lambda\phi_\tau)\right)\right)\left(\sigma(-\mu+\sigma)H_F^2 L_D\phi_\rho\right.\\
&\quad +\sigma H_D^2\phi_\rho(-\lambda+(-1+\lambda)\phi_\tau)\left(\lambda(\mu-\sigma)L_D+(-1+\lambda)((\mu+\sigma)L_D+2\mu L_F)\phi_\tau\right)\\
&\quad +H_D H_F\left(\lambda(\mu-\sigma)L_D\left(\mu-\sigma-(\mu+\sigma)\phi_\rho^2\right)-(-1+\lambda)\left(2\mu\sigma L_F\phi_\rho^2\right.\right.\\
&\quad \left.\left.\left.+L_D\left(-\mu^2+\sigma^2+\left(\mu^2+\sigma^2\right)\phi_\rho^2\right)\right)\phi_\tau\right)\right)\right)
\end{aligned}
$$

where

$$
\begin{aligned}
D(\phi_\tau,\phi_\rho) &= 2(\mu-\sigma)\left(\sigma^2 H_F^3\phi_\rho^2\phi_\tau+\sigma H_D^3\phi_\rho(-\lambda+(-1+\lambda)\phi_\tau)\left((1-\lambda)\lambda(\mu-\sigma)\right.\right.\\
&\quad \left.-(1+2(-1+\lambda)\lambda)\sigma\phi_\tau+(-1+\lambda)\lambda(\mu+\sigma)\phi_\tau^2\right)\\
&\quad +\sigma H_D H_F^2\phi_\rho\left((-1+\lambda)(\mu-\sigma)+\lambda\left(-\mu+2\sigma+(\mu+\sigma)\phi_\rho^2\right)\phi_\tau\right.\\
&\quad \left.-(-1+\lambda)\left(\sigma+(\mu+\sigma)\phi_\rho^2\right)\phi_\tau^2\right)+H_D^2 H_F\left((\mu-\sigma)\left((1-\lambda)\lambda(\mu-\sigma)\right.\right.\\
&\quad \left.-(1+2(-1+\lambda)\lambda)\sigma\phi_\tau+(-1+\lambda)\lambda(\mu+\sigma)\phi_\tau^2\right)\\
&\quad +\phi_\rho^2\left((-1+\lambda)\lambda(\mu-\sigma)(\mu+\sigma)+\phi_\tau\left(\sigma\left(\sigma-2\lambda\sigma+\lambda^2(\mu+3\sigma)\right)\right.\right.\\
&\quad \left.\left.\left.\left.+(-1+\lambda)\phi_\tau\left(-\left(\lambda\left(\mu^2+2\mu\sigma+3\sigma^2\right)\right)+(-1+\lambda)\sigma(\mu+\sigma)\phi_\tau\right)\right)\right)\right)\right)
\end{aligned}
$$

Bibliography

[1] Ago, T., I., Isono and T., Tabuchi (2006) Locational disadvantage of the hub, *Annales of Regional Science* 40, 819-848.

[2] Baldwin, R., R. Forslid, P. Martin, G. Ottaviano and F. Robert-Nicoud (2003) *Economic Geography and Public Policy.* Princeton: Princeton University Press.

[3] Barro, R. J. and J.-W. Lee (2001) International data on educational attainment: updates and implications, *Oxford Economic Papers* 3, 541-563.

[4] Beckmann, M. and J.-F. Thisse (1986) The location of production activites In: Nijkamp, P. (ed.) *Handbook of Regional and Urban Economics, vol. 1*, North-Holland: Elsevier Science Publishers, pp. 21-95.

[5] Behrens, K., C. Gaigné, G.I.P. Ottaviano and J.-F. Thisse (2006) Is remoteness a locational disadvantage, *Journal of Economic Geography* 6, 347-368.

[6] Behrens, K., C. Gaigné, G.I.P. Ottaviano and J.-F. Thisse (2007) Countries regions and trade: On the welfare impacts of economic integration, *European Economic Review* 51, 1277-1301.

[7] Byerlee, D. (1974) Rural-Urban Migration in Africa: Theory, Policy and Research Implications, *International Migration Review* 8, 543-566.

[8] Christopher, A.J. (1994) Urbanization and National Capitals in Africa. In: Tarver, (ed.) *Urbanization in Africa: a handbook*. Westport: Greenwood Press , pp. 408 - 421.

[9] Forslid, R., and G.I.P., Ottaviano (2003) An analytically solvable core-periphery model, *Journal of Economic Geography* 3, 229-240.

[10] Fujita, M. and T., Mori (1996) The role of ports in the making of major cities: Self-agglomeration and hub-effect, *Journal of Development Economics* 49, 93-120.

[11] Hoselitz, B. F. (1955) Generative and Parasitic Cities, *Economic Development and Cultural Change* 3, 278-294.

[12] Kempe, R. H. (1996) *African Political Economy.* Cambridge, MA: M.E. Sharpe.

[13] Kessides, C. (2005) The Urban Transition in Sub-Saharan Africa: Implications for Economic Growth and Poverty Reduction, Africa Region Working Paper Series #97.

[14] Krugman, P. (1991) Increasing Returns and Economic Geography, *Journal of Political Economy* 99, 483-499.

[15] Krugman, P. (1993) The hub effect: or, threeness in International trade. In: Ethier W.J., E. Helpman, and J.P. Neary (eds.) *Theory, policy and dynamics in international trade*. Cambridge: Cambridge University Press, pp. xxi - xxxii.

[16] Krugman, P. and R.L., Elizondo (1996) Trade policy and the Third World metropolis, *Journal of Development Economics* 49, 137-150.

[17] Mabogunje, A.L. (1994) Introduction: Cities and Africa's Economic Recovery. In: Tarver, (ed.) *Urbanization in Africa: a handbook*. Westport: Greenwood Press , pp. xxi - xxxii.

[18] Monfort, P. and R., Nicolini (2000) Regional Convergence and International Integration, *Journal of Urban Economics* 48, 286-306.

[19] Obudho, R.A. (1997) Nairobi: National capital and regional hub. In: Rakodi, C. (ed.) *The urban challenge in Africa: Growth and management of its large cities.* Tokyo - New York - Paris: United Nations University Press available online at `http://www.unu.edu/unupress/unupbooks/uu26ue/uu26ue00.htm`.

[20] Ottaviano, G., T. Tabuchi and J.-F. Thisse (2002) Agglomeration and Trade Revisited, *International Economic Review* 43, 409-435.

[21] Robert-Nicoud, F. (2005) The structure of simple 'New Economic Geography' models (or, On identical twins), *Journal of Economic Geography* 5, 201-234.

[22] Todaro, M. P. (1997) Urbanization, Unemployment and Migration in Africa: Theory and Policy, Policy Research Division Working Paper #104.