Analysing the spatial persistence of population and wealth during Apartheid

PD Niemand

22258817

Dissertation submitted in partial fulfillment of the requirements for the degree *Magister Commercii* in Economics at the Potchefstroom Campus of the North-West University

Supervisor: Prof WF Krugell

May 2015

Abstract

This dissertation undertakes an analysis of the spatial persistence of population in South Africa over the period 1911 to 2011. A comprehensive review is given of the history and development of geographical economics in order to understand the dynamics of the forces of agglomeration. In addition the history of the development of South Africa is discussed and special focus is directed to the geographical, economic and political factors that gave rise to the unequal distribution of population and wealth in the country. In the empirical analysis Zipf's law was applied and it was found that South Africa's population was more evenly spread in 1911. With the application of the law to the 2011 data the Pareto exponent of the OLS log-linear regression indicated that urban agglomeration was more persistent. Although this might indicate that apartheid did not influence agglomeration in South Africa it is argued that the nature of the agglomeration was in fact influenced by restrictive measures placed on the urbanisation of the population and industrial decentralisation policies. It is indicated that the apartheid policy altered the equilibrium spatial distribution of population and wealth which lead to a smaller than optimal primate and second largest magisterial districts, too many secondary cities of similar size, and also too many small and uneconomical rural settlements.

Key words: Geographical economics, Zipf's law, rank size distribution, spatial development, urbanisation, industrialisation, decentralisation, policy, South Africa

Opsomming

Hierdie verhandeling analiseer die ruimtelike konsentrasie van die bevolking in Suid-Afrika. ŉ Omvattende geskiedkundige oorsig word gegee van die ontwikkeling van geografiese ekonomie om die dinamiek van die kragte van agglomerasie beter te begryp. Daarbenewens word ŉ historiese oorsig van Suid-Afrika se ontwikkeling gegee, met spesiale klem op die geografiese, ekonomiese en politieke faktore wat aanleiding gegee het tot die oneweredige verspreiding van die bevolking en welvaart in die land. In die empiriese ontleding word Zipf se wet toegepas en daar word bevind dat Suid-Afrika se bevolking meer eweredig versprei was in 1911. Met die toepassing van die wet op die 2011 data het die Pareto-eksponent van die log-liniêre regressie aangedui dat stedelike agglomerasie meer gekonsentreerd was. Alhoewel hierdie ŉ aanduiding mag gee dat apartheid nie ŉ invloed gehad het op agglomerasie in Suid-Afrika nie word daar geargumenteer dat die aard van die agglomerasie wel beïnvloed was deur die beperkende maatstawwe op verstedeliking van die bevolking asook nywerheidsdesentralisasiebeleide. Daar word aangedui dat apartheidsbeleid die ruimtelike ewewigsverspreiding van die bevolking verander het wat daartoe gelei het dat daar ŉ minder as optimale primaat en tweede grootste munisipaliteite, te veel sekondêre stede van soortgelyke grootte, asook te veel klein en onekonomiese plattelandse nedersettings gevorm was.

Sleutelwoorde: Geografiese ekonomie, Zipf se wet, rang grootteverspreiding, ruimtelike ontwikkeling, verstedeliking, nywerheidsontwikkeling, desentralisasie, Suid-Afrika

Table of contents

Abstract ii
Opsomming iii
List of tables and figures vi
List of acronyms vii
Chapter 1: Introduction 1
1.1 Introduction 1
1.2 Literature review 3
1.3 Research objectives 8
1.3.1 General objective 8
1.3.2 Specific objectives 8
1.4 Research methodology 8
1.5 Chapter division 9
Chapter 2: Theory on economic geography 11
2.1 Introduction 11
2.2 The history of economic geography 12
2.2.1 The origins and early development of geographical economics 13
2.2.2 The quantitative revolution and the shift to a spatial science 20
2.2.3 1970s to 1990s: Urban system theories and industrial organisation revolution 23
2.2.4 Krugman's core model of geographical economics 27
2.2.5 Geographical economics after Krugman (1991a) 31
2.3 Agglomeration and economic growth 34
2.3.1 Geographical economics and endogenous growth theory 35
2.3.2 Economic development 37
2.3.3 Policy implications of a restriction on migration 44
2.4 Conclusion 48
Chapter 3: The history of the concentration and movement of people and economic activity in South Africa 54
3.1 Introduction 54

3.2 Development of South Africa 55
3.2.1 The impact of first nature aspects and the start of urbanisation and industrialisation 55
3.2.2 The establishment of the Union of South Africa 60
3.2.3 Apartheid and decentralisation 66
3.2.3.1 1950s and 1960s 66
3.2.3.2 The 1970s 72
3.2.3.3 The 1980s 75
3.2.4 Post-Apartheid South Africa 83
3.2.4.1 Urbanisation 83
3.2.4.2 Industrial policy 86
3.2.4.3 Effects of decentralisation and restrictive urbanisation on development 89
3.3 Conclusion 90
Chapter 4: South Africa's city size distribution 97
4.1 Introduction 97
4.2 An explanation of Zipf's law 98
4.3 South African data and the evolution of the urban system 102
4.4 An application of Zipf's law to 1911 and 2011 104
4.5 Conclusion 109
Chapter 5: Summary, conclusion and recommendations 112
5.1 Summary 112
5.2 Conclusion 114
5.3 Recommendations 114
Bibliography 116

List of tables and figures

Table 3.1: Urban share of population in South Africa, 1911-2011

Table 4.2: Description of smallest, largest and average population of magisterial districts in 1911 and 2011

Table 4.3: 21 largest agglomerations with 50 000 or more people in 1911

Table 4.4: Discrepancy between actual population and the predicted population if rank size rule is applied to 2011

Figure 1.1: Map indicating population density of the Republic of South Africa in 1911

Figure 4.2: Log of population versus log of rank of the 135 US metropolitan areas in 1991

Figure 4.3: Logarithm of size versus logarithm of 205 magisterial districts in 1911 in South Africa

Figure 4.4: Logarithm of size versus logarithm of rank of 205 magisterial districts in 2011 in South Africa

List of acronyms

ANC	African National Congress
ASGISA	Accelerated and Shared Growth Initiative
CES	Constant-Elasticity-of-Substitution
DBSA	Development Bank of Southern Africa
DPLG	Department of Provincial and Local Government
DTI	Department of Trade and Industry
FRG	Federal Republic of Germany
GDP	Gross Domestic Product
GDR	German Democratic Republic
IDZs	Industrial Development Zones
IPAPs	Industrial Policy Action Plans
IUDF	Integrated Urban Development Framework
LED	Local Economic Development
MAR	Marshall-Arrow-Romer
NDP	National Development Plan
NEG	New Economic Geography
NIPF	National and Industrial Policy Framework
NP	National Party
NPDP	National Physical Development Plan
NSDF	National Spatial Development Framework
NSDP	National Spatial Development Perspective
OFS	Orange Free State

OLS	Ordinary Least Squares
PWV	Pretoria-Witwatersrand-Vereeniging
RDP	Reconstruction and Development Programme
RIDS	Regional Industrial Development Strategy
RSA	Republic of South Africa
SACN	South African Cities Network
SBDC	Small Business Development Corporation
SDIs	Spatial Development Initiatives
TBVC	Transkei, Bophuthatswana, Venda and Ciskei
USA	United States of America
WW	World War

Chapter 1: Introduction

1.1 Introduction

In order for development of a country to be achieved it is essential to reshape its economic geography (World Bank, 2009:9). History has shown that most countries need to change the geographical distribution of economic activity and its people in order to successfully develop. As the economy of a nation develops, economic activities and people become more concentrated spatially, since economic development is associated with growing cities (Bos, 1992:36). Workers migrate from rural to urban areas as a result of greater economic opportunities and shorter distances to markets in densely populated areas. As more people migrate to live in or near large settlements the economic density of urban areas rapidly increases (World Bank, 2009:1-19).

The shift toward density is, usually, quick in rapid-growing economies, since density is strongly associated with economic activity. The expanding economic activity then spills over to areas connected to fast-growing agglomerations which, in their turn, lead to the growth of towns and cities.[1] When cities grow, more mobile people and specialisation of products are fundamental to this transformation and economic development. Baldwin and Martin (2003) state that agglomeration and economic growth are inextricably linked, since agglomeration can be compared as the "territorial counterpart of economic growth" (Brakman, Garretsen & Van Marrewijk, 2009:437). But, this process, however, does not happen without sacrifices. Economic growth does not spread smoothly across geographical areas. According to Dewar, Todes and Watson (1986:14-15) economic disparities occur between rural areas, where economic activity as well as development generally stagnates, and urban areas where populations generally prosper. In order to achieve "unbalanced" economic growth governments need to institute good policies. These policies, however, also need to address the rural-urban disparities in living standards that occur as a result of the spatial transformation (World Bank, 2009:33-49).

South Africa's spatial transformation, over the last century has, however, not been according to the general transformation other successful countries have undergone (Coetzee & De Coning, 1992:1). The spatial system of the country was dramatically affected by limitations

[1] "Fast-growing agglomerations" refer to rapidly growing towns, cities and metropolises as a result of increased economic activity through, for example, the growth of the manufacturing sector (Brakman *et al*., 2009:185).

on the movement of Africans even before the system of apartheid was introduced and after introduction of the regime tighter restrictive measures were implemented (Wessels & Wentzel, 1989:9). The state-enforced regulations, such as the influx control and Group Areas Act, meant that the spatial structure was racially based and that only white citizens had full industrial and urban citizenship in most places during the regime. In addition industrial decentralisation policies were implemented in order to strengthen the ideological view of segregation. The urban system was based on policies that aimed to decentralise and deconcentrate employment of people and accordingly economic activity of the country. As a consequence the balance between forces of agglomeration and dispersion in the country may have been significantly impaired (Davenport, 1991:9-16).

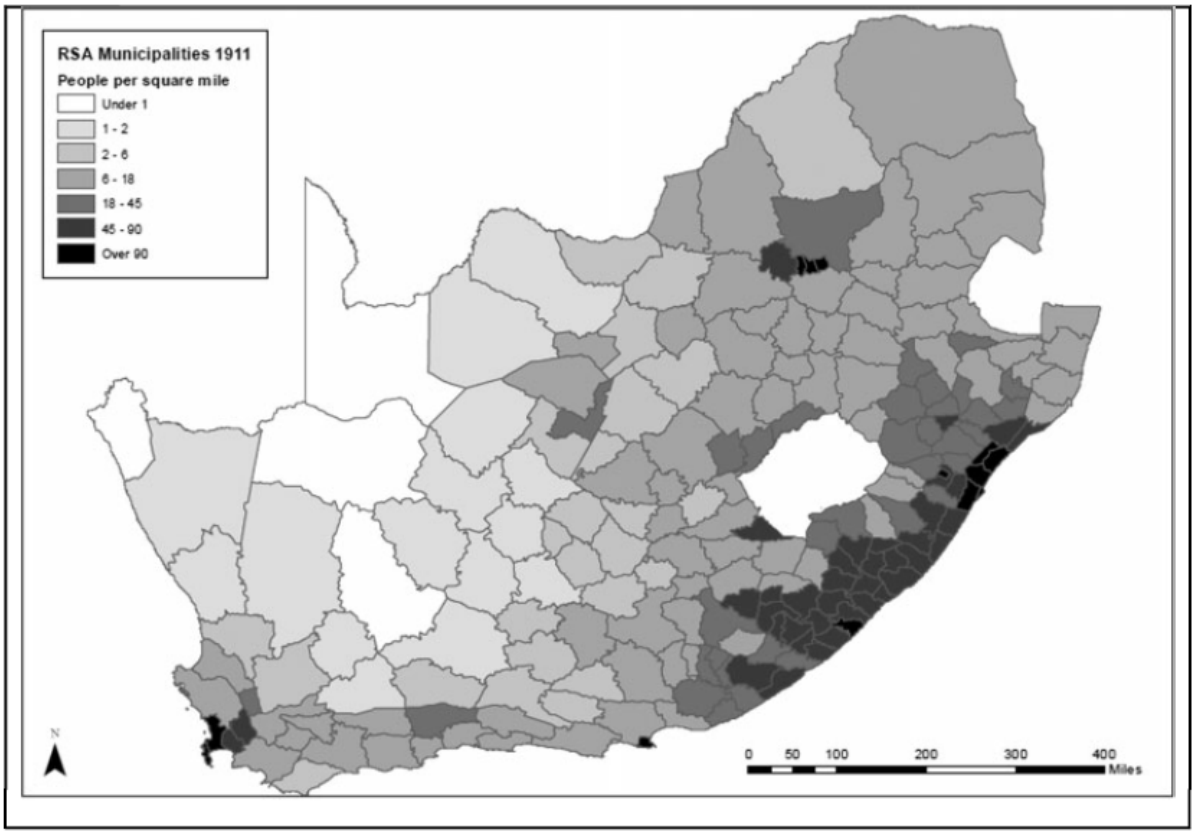

Figure 1.1: Map indicating population density of the Republic of South Africa in 1911

The growth and existence of cities are basically a result of the forces of agglomeration in the form of increasing returns to scale (Brakman *et al.*, 2009:299). Large and vibrant cities are essential since they foster both economic growth and development in a country (World Bank,

2009). However, what happens when 'irrational'[2] economic policies, such as apartheid, with a place-dependent character, are implemented? What effect may it have on urban agglomeration and the distribution of wealth and population within an economy?

1.2 Literature review

The importance of the field of geographical economics and the distribution of economic activity is highlighted by Brakman *et al.* (2009). The relevance of geographical economics and agglomeration is given by means of practical real world examples and also empirical evidence. Specific reference should also be made to Krugman's (1991a) core model of geographical economics and the important influence this model has had on the explanation of centripetal forces that pull population and economic activity together and centrifugal forces that drive them apart. However, Brakman *et al*. (2009:299) make an important point regarding literature on geographical economics. The literature gives a good indication of the sources and scope of urban agglomeration economies, however it does not suffice in explaining city size distributions. However, geographical economists and urban economists have turned to another becnhmark of analysis of the distribution of city sizes and urban systems. This benchmark is known as Zipf's law for cities. The law is a special case of the rank size distribution and it enables economists to come to probabilistic answers regarding agglomeration and the distribution of wealth and population within an economy (Brakman *et al*., 2009).

The World Bank (2009:51) defines Zipf's law as a special case of the rank size rule, where "the population of any city is equal to the population of the largest city, divided by the rank of the city in question within a country's urban hierarchy" (World Bank, 2009:51)[3]. Gabaix and Ioannides (2003:4-7) give an even more precise definition of Zipf's law and reviews empirical evidence, discusses econometric regression and looks at theories regarding the distribution. Gabaix (1999:739-765) gives an in-depth explanation of the power law and describes it as "the number of cities with populations greater than S is proportional to 1/S" (Gabaix, 1999:739)[4]. According to Brakman *et al.* (2009:299) the law can be useful in explaining city size distribution, where normal literature on the sources and scope of

[2] Irrational throughout this dissertation refers to the opposite of the conventional in terms of spatial policy. It is not used as a value judgement or to indicate a political position, it is merely used to economically indicate the disposition of the norm.

[3] The rank size rule states that the size, in population, of the n^{th} city or settlement is approximately one-n^{th} of the size of the largest city (Beckmann, 1958:243-248).

[4] Zipf's law, as a power law or regularity, was first noted by Felix Auerbach, but the first person to name it was George Kingsley Zipf (Gabaix, 1999:739).

agglomeration economies are redundant. The reason for this is that literature focuses on the size and growth of individual cities, without taking into account the spatial interconnections and interdependencies of cities. Many researchers, on the other hand, question Zipf's law and in general do not believe it to always hold true (Gabaix & Ioannides, 2003:14). A broad message can, however, be taken from the rank size distribution and Zipf's law – a wide range of settlement sizes coexist within a country (World Bank, 2009:52).

Zipf's law has been applied in a previous study conducted on South Africa by Naudé and Krugell (2003). They estimated that for the 123 largest cities and towns in the country in 2001 Zipf's law did not hold. However, they came to the conclusion that the city size distribution in South Africa at that time was more evenly spead and that these cities delivered urbanisation rather than localisation economies. On the international stage the law has been applied numerous times. Soo (2005) uses data for 73 countries in order to assess the empirical validity of Zipf's law. He found that when using Ordinary Least Squares (OLS), more often than not the law did not hold. On the other hand Brakman, Garretsen, Van Marrewijk and Van den Berg (1999) indicate that Zipf's law in fact holds true for the Netherlands in 1900. However, they also stressed the importance of the fact that the value of the Pareto coefficient in the application of Zipf's law is not always constant since the development of the country may influence the rank size distribution. For instance, in 1600 they found that Zipf's law did not hold and a more even spead of urban agglomeration was evident. In 1990 the development of the country had also changed from 1900 and consequently it was again found that the city size distribution in the country was more even. Gabaix (1999) in an example of Zipf's law in his article used 135 American metropolitan areas for 1991. He came to the conclusion in this example that Zipf's law held strongly for the United States of America (USA). Rosen and Resnick (1980) showed that Zipf's law in general applies for most countries in the modern period. Anderson and Ge (2005) found that the city size distribution of Chinese cities was in fact influenced by economic reforms made since 1979. The conclusion was drawn that place-dependent policies may have an impact on the city size growth of a country.

As mentioned in the previous paragraph by Brakman *et al*. (1999) development of a nation can affect city size distributions. The spatial transformation – the growth of towns and cities as a result of urbanisation – is closely related to sectoral transformations of a country (World Bank, 2009:9). Sectoral transformation is a prerequisite for economic growth as has been established by a number of economic growth and development theorists. Lewis (1954)

describes the growth of a developing country in his dual-sector model or Lewis model. The model entails the movement of labour or people from the agricultural sector to the manufacturing sector, which is according to Lewis the transition from a traditional to an industrial economy. Furthermore, Rostow (1960) focused on the future of developing countries regarding economic development, based on the economic history of Europe, specifically Britain. He argued that nations go through five phases, which corresponds with the move from an agrarian society to a substantial manufacturing sector and escalates to a significant improvement of living standards of the people where production is aimed at consumer consumption, specialisation of products occur and technology is sophisticated (Cypher & Dietz, 2009:159-164). Kuznets (1971) summarised, in his Nobel Prize winning lecture, the structural transformations that occur with economic growth, which involve the shift away from agricultural to non-agricultural activities, as well as more recently, the movement away from industry to a service-oriented economy. Other economic development theories which mutually supported the sectoral transformations needed for development, although the emphasis and interpretation are different, include Hirschman's (1958) unbalanced growth theory, Nurkse's (1953) theory of balanced growth and Rosenstein-Rodan's (1976) "big push" theory.

Van Jaarsveld (1985) indicates that the historical sectoral transformation South Africa has undergone from as early as 1652. Industrialisation was for the first time relevant when gold and diamonds were discovered and along with that urbanisation occurred. Bos (1992:216-370) describes the historical pattern of industrialisation in the country and takes an in-depth look at decentralisation, as well as policies regarding decentralisation. Theory is also explained regarding the growth of the country. In addition, Badenhorst (1985:78-86) describes the development of cities and specifically focuses on the agglomeration patterns in South Africa.

Brakman *et al.* (2009:23) state that the clustering of people can be seen as the rule, not the exception. People and economic activities become more concentrated as a country develops (World Bank, 2009:8). According to the World Bank (2009:8) economic concentration can be measured by the rate of urbanisation, since the growth of population and economic density in towns and cities is a direct result of urbanisation. It is important to establish the historical background of urbanisation in South Africa and to determine the impact racially restrictive policies had on the migration of people (Wessels & Wentzel, 1989:5). The reason for this is that urbanisation and development of a country are positively related, since the concentration

of economic activity increases with development (World Bank, 2009:49; Van Jaarsveld, 1985:2). Brakman *et al.* (2009:10) also state that there is a strong correlation between the degree of urbanisation and the per capita income or living standards of a country.

Wessels and Wentzel (1989) give an early historical overview of urbanisation since the founding of the Union of South Africa in 1910. An indication is given of the impact diamond and gold mining had on urbanisation during the early parts of the 20th century in South Africa. In addition the report focuses on the development of the country in the form of industrialisation and discusses the restrictions placed on black migrants to access settlements and cities. Similarly, an analysis is given of the impact these restrictions or policies had on black urbanisation. Gelderblom and Kok (1994) also give a historical background of urbanisation in South Africa from the beginning of the 20th century to the early stages of democracy. Moreover, the dynamics of urbanisation is dealt with and a theoretical, as well as comparative background is provided. Theories concerning urbanisation, development, and migration are applied and challenges of the country in the future are emphasised.

Swilling, Humphries and Shubane (1991) depict the history of urbanisation and growth of cities, since the implementation of the apartheid regime in 1948 till the early 1990s. The migration of people, the limits placed on migration, apartheid legislation and policies, as well as the government's impact on agglomeration and growth of cities are discussed. Swilling *et al*. (1991) argue that the policies of the apartheid government were aimed at decentralisation and deconcentration of the population and that racial inequalities during this period impaired the growth of agglomerations in the country. Early literature and theory on urban locations, urbanisation, and the evolution of urban locations as well as the distribution of city sizes are discussed by Berry and Horton (1969). Turok (2012) provides a comprehensive review of urbanisation in South Africa from an early stage in its development and also provides explanations for the restriction or rapid movement of people to urban centres. Furthermore, Booysen (1985) and Choe and Chrite (2014) take a look at urbanisation strategies during apartheid. On the other hand, Bond (2003), discusses policies aimed at the integration of areas that were segregated by apartheid as well as urbanisation strategies implemented after the democratisation of the country.

Dewar *et al.* (1986) focus on the role of settlements and its impact on development in South Africa. A historical view is taken of urbanisation in the country and the relationship between development theory and settlement policy internationally is touched upon. The major themes

of the project report are the impact of policies such as influx control, decentralisation policy, rural development and it prescribes a framework for development. A specific focus should be given to the industrial decentralisation policies during apartheid and a number of authors including Wellings and Black (1984, 1986), Cobbett, Glaser, Hindson and Swilling (1985), Fine and Rustomjee (1996) and Hindson's (1987) work takes a look at these policies. On the other hand, Naudé and Krugell (2005) take a look at industrial policy after apartheid and discuss how the distribution of cities was influenced by geographical elements. Furthermore, Rogerson (1997; 2008), Phalatse (2001), Crush and Rogerson (2001) and Oranje and Merrifield (2010) all consider industrial policies after apartheid and what impact they have had spatially.

The World Bank (2009:20-47) argues that countries should institute policies that encourage unbalanced economic growth, yet ensure development throughout the country. The reason for this is that rural-urban disparities in living standards occur with the movement of people to densely populated areas, where living standards are generally higher than rural settlements. Consequently, the rural-urban and within urban disparities in development should be a critical consideration when governments institute policies.

In order for development of South Africa to be achieved it is essential that the economic geography should evolve. This evolution includes the growth of cities, more mobile people and increasingly specialised products (World Bank, 2009:9-10). The World Bank (2009:24) emphasises that no country has grown to middle income without industrialisation and urbanisation. Moreover, no country has grown to high income without vibrant cities. Agglomeration not only concentrates and increases economic production, but also leads to a convergence of living standards. In South Africa, agglomeration has been impaired by pre-democratic policies. On the other hand, the country has made good economic progress since 1994.

The objective of the research is to determine the spatial persistence of agglomeration in South Africa, which will give an indication of the impacts historical forces have had on the level of agglomeration in the country.

The following research questions can be formulated based on the above-mentioned description of the research problem:

- How persistent is agglomeration in South Africa?

- What effect did apartheid, especially restrictive measures on the movement of Africans and industrial decentralisation policies have on forces of agglomeration and consequently the city size distribution of South Africa?

In order to answer the above research questions, the following research objectives are set.

1.3 Research objectives

The research objectives are divided into a general and specific objectives.

1.3.1 General objective

The general objective of this research is to determine the persistence of agglomeration in South Africa. This will in turn give us an indication of whether the South African spatial economy had been affected by apartheid policies.

1.3.2 Specific objectives

The specific objectives of this research are:

- To research theory on economic geography and the history of the development of geographical economics in order to determine how the understanding of agglomeration has evolved towards contemporary practice?
- To research the historical development of South Africa and give an extensive review of policies that may have impaired the equilibrium spatial distribution of economic activity within the country.
- To assess Zipf's law of population and apply it to South Africa.

1.4 Research method

This research, pertaining to the specific objectives, consists of two phases, namely a literature review and an empirical study.

The first phase will look at literature regarding economic geography and its evolution towards the discipline of geographical economics. This will give one a good indication of how the understanding of forces of agglomeration has changed over the years to the modern-day understanding and application of agglomeration.

After this the history of the development of South Africa is examined in the context of the literature researched in the first phase. A comprehensive overview is given regarding the forces that may have impaired the modern day spatial distribution of economic activity in the country. Specific focus is put on the first nature aspects as well as apartheid's measures to

restrict the movement of Africans to urban areas in the country and post-apartheid policies. Accordingly a better picture is sketched regarding the geographical and political factors that may have influenced the distribution of population.

The empirical phase then commences. Place-level census data from 1911 and 2011 for the population of 205 magisterial districts are used and applied to determine the rank size distribution of South Africa's magisterial districts, to see whether Zipf's law holds true for South Africa. The period from 1911 to 2011 captures many of the forces that have influenced the city size distribution of the country and a comparison of 1911 and 2011 will give a good indication of the impact apartheid had on agglomeration.

The specific design to be used: a comparison is drawn between the results of 1911 and 2011's rank size distribution. A simple OLS log-linear regression has been run with the log of population on the log of rank of the place in order to determine whether Zipf's law and the rank size rule holds. The results have been interpreted and probabilistic answers derived from the results. Furthermore, measures such as the primacy ratio was used in order to support some of the arguments made with the analysis.

1.5 Chapter division

The chapter division of the dissertation is as follows: Chapter 1 gives an introduction. Chapter 2 revolves around theory on economic geography and how the discipline has historically developed and given rise to different sub-disciplines such as geographical economics. The Chapter then discusses contemporary geographical economics and how it is applied to agglomeration and economic growth. Chapter 3 discusses the historical development of South Africa in terms of the concentration of population and economic activity throughout the country. Firstly, the Chapter looks at the impact of first nature geography on the distribution of economic activity. Thereafter chronological phases of development of the country will be discussed. Urbanisation and its restriction as well as industrial decentralisation policies during apartheid are comprehensively discussed in order to indicate the political forces that may have influenced the distribution of wealth and population during apartheid. The chapter also looks at post-apartheid industrial policies and urbanisation of the population. Chapter 4 forms the foundation for analysis in this dissertation. The city size distribution of South Africa has been determined by applying the rank size rule and Zipf's law to population data from 1911 and 2011. Chapter 5 provides a

summary of the dissertation, deriving certain conclusions from the findings and lastly providing recommendations for further research.

Chapter 2: Theory on economic geography

2.1 Introduction

It is an empirical fact that economic activity is not distributed randomly across space (Brakman *et al*., 2009:23). This fact can be intelligibly shown by the clustering of people and firms in major cities around the world. Take for a start the largest city in the world – Tokyo. The city's surface area is more or less four per cent of Japan's total land area; nonetheless it holds roughly a quarter of the country's population with more than 35 million people residing in the country's capital. Another similar example would be the uneven spatial distributions that occur in the United States of America (USA). The three states of Texas, California and New York had 19.8 per cent of the US population, yet they took up only 12.8 per cent of the country's land in 2005 (World Bank, 2009:19-39). Millions of people are living close together in densely populated cities around the world and in the process subject themselves to the unpleasantness of living in these packed areas. In contrast there are vast empty landscapes and areas with only a few small towns with minimal human interaction. The unequal distribution of population and economic activity across the world is one of the most conspicuous phenomena of the global economic system (Van Marrewijk, 2012:294). But, why is it the case? According to Brakman *et al*. (2009:3-4) sociological, psychological, historical, cultural and geographical motives may influence the clustering of people in cities and economic centres. However, a motive exists that may even be a prerequisite for the other motives – the economic rationale behind clustering, or as it is technically known agglomeration.[5] The agglomeration or clustering of economic activity occurs at various geographical levels and has many different compositions. Examples of types of agglomeration include the clustering of a number of restaurants and shops within a region; the growth of a number of industrial districts; or the formation of cities, which all have different sizes, which range from large metropolis to small towns. Another relevant type of agglomeration is the existence of regional disparities within the same country (Fujita & Krugman, 2004:140). Research on agglomeration continues to grip economic geographers, although the social, cultural and institutional dimensions of clusters are getting more and more attention. Consequently contemporary economic geography's primary concern is to ascertain and explain the cause and effect of uneven development of population and economic activity within and between regions (Aoyama, Murphy & Hanson, 2011:1-3). This

[5] According to Brakman *et al*. (2009:4) psychological, cultural, sociological and historical motives may have developed mainly in response to an economic motive, which gave an incentive for people to settle in towns and cities.

chapter focuses on the history of economic geography, from the earliest form of the Von Thünen (1826) model to the core-periphery model of Krugman (1991a) to give the reader an idea of how economic theory regarding the study of spatial agglomerations has evolved and what geographical economics contemporarily entails.[6] Furthermore, the focus will shift towards the link between agglomeration and economic growth, and also its impact on development, industrialisation and migration of a country.

2.2 The history of economic geography

Economic geographers have dedicated their studies to geographically specific factors that form economic processes and determine important agents and drivers which trigger unequal territorial development and transformation.[7] The discipline of economic geography's explanations has, however, differed somewhat over its history (Aoyama *et al.*, 2011:1-2). The reason for this is that economic geographers have considered a number of geographically specific endowments as drivers of territorial development and also the fact that economic geography, as a sub-discipline of geography, has sometimes been misunderstood. This misunderstanding can be due to the various roots[8] of the discipline, its diverse links with other social science disciplines and the use of heterodox methodologies within the field (Aoyama *et al.*, 2011:1-2). Some authors even prefer the use of a certain term when describing the sub-discipline, like "geographical economics", "new regional science" or "new economic geography" (Brakman *et al.*, 2009:22-23). Contemporarily the linkages between fields such as geographical economics, which is mostly populated by economists, and economic geography, which is populated by geographers, have strengthened (Aoyama *et al.*, 2011:7).[9] Although authors may prefer the use of a different term for geographical economics, the topics covered remain the essence of the sub-discipline (Brakman *et al.*, 2009:23).

---

6 Core-periphery traditionally refers to the arrangement of a country into a manufacturing "core" and an agricultural "periphery" (Krugman, 1998:13). However, contemporarily it may refer to large urban areas as the "core" and abandoned rural areas as the "periphery".

7 Agents include the state, labour and firms. Drivers include innovation, accessibility, institutions and entrepreneurship.

8 Economic geography has various influences from fields like geographical economics, economic sociology, regional science and regional and urban studies (Aoyama *et al.*, 2011:2).

9 "Geographical economics" as a term will henceforth be adopted by this dissertation, since I tend to agree with Brakman *et al.* (2009) regarding the use of the term. Brakman *et al.* (2009) prefer the use of "geographical economics" rather than "new regional science" or "new economic geography" due to the fact that the "new" will eventually wear of as time passes and furthermore the latter term implies that economic geographers developed the theory, which is not the case. The roots of geographical economics can be found in international economics, modern international trade and economic growth theory (Brakman *et al*, 2009:21-23).

According to Fujita and Krugman (2004:140-141) the essence of the issue of geographical economics is how to explain the structure of economic agglomeration or concentration geographically across space. Agglomeration and concentration deal with the question of whether a certain part of economic activity can be found at a few locations, be it a country, a region or a city (Brakman *et al*., 2009:185). Aoyama *et al.* (2011:1) state that the main goal has been to provide multi-faceted explanations for economic processes, such as prosperity and growth and also crises and decline, proven across different territories at the local, regional, national and international scale (Aoyama *et al*., 2011:1). However, the field of geographical economics has been subject to change since the discipline emerged as a result of contextual factors and these factors are subject to change over time (Barnes, 2000:12; Scott, 2000:484). Latest trends in the sub-discipline have been the movement towards research that focuses on territorial development contributions, that are regularly unquantifiable and intangible, and social endowments like institutions, culture, knowledge and networks. History and change play a major role in the formation of the geographical economics and therefore it is important to look back on classical theories, since they may be useful in explaining contemporary field specific questions such as: "What are the linkages and interdependencies between towns and cities?" and "why does uneven development and competition between regions occur?" Reconnecting traditional theories with new phenomena and linking historical concepts with contemporary debates may be an invaluable tool in showing that new challenges and questions have long intellectual backgrounds (Aoyama *et al*., 2011:1-8). As Barnes (2012:4) states, the history of the ideas of the discipline matters, since "ideas become true only in history and are not born true outside history" (Barnes, 2012:4).

2.2.1 The origins and early development of geographical economics

It is difficult to confirm an exact date for the creation of geographical economics (Barnes, 2000:12). The origin and historical descent of geographical economics have a number of interpretations and discourses due to the contextual nature of the discipline (Aoyama *et al*., 2011:2).[10] Some argue that the origin or roots of geographical economics are connected to West-European imperialism. Colonialism led to growth of global commercial trade which required further geographical examination in order understand and enhance the performance

[10] Discourse in the sense that a network of concepts, practices and statements gives rise to a prominent body of knowledge (Barnes, 2000:13).

of trade routes and means of transportation (Barnes, 2000:12-13).[11] Other possible roots include the distinguishing of commercial and economic geography in 1882 by the German geographer Götz; the first teaching of geographical economics courses at Cornell University and the University of Pennsylvania in 1893; or when the journal *Economic geography* was initially published in 1925 (Barnes, 2000:12-13; Aoyama *et al*., 2011:2). Another source may be Alfred Marshall, a British economist, who transformed economics through the marginal revolution during the early 1900s (Aoyama *et al*., 2011:2).[12] What is evident out of these possible roots is that the origins of geographical economics can be narrowed down to the nineteenth and early twentieth century. This period was significant for the development of geographical economics since it was the first time that the discipline had been institutionalised within Western European and North American Universities (Barnes, 2000: 13).

The Germanic location theories of Heinrich Von Thünen and Alfred Weber are also considered by many as the origin of the field (Aoyama *et al*., 2011:2). Von Thünen's book, *Der Isolierte Staat* (The Isolated State) published in 1826, is for some the first classical monograph of geographical economics (Barnes, 2000:13). Although Von Thünen's (1826) model, the monocentric city model, remains a point of reference for regional and urban economics, it shows that the ideas and concepts of geographical economics have been examined before in a basic form without increasing returns to scale[13] – a term that is strongly linked with the economics of agglomeration (Brakman *et al*., 2009:34-35; Anas, Arnott & Small, 1998:1427).[14] The model assumes the existence of cities and central business districts and tries to determine the consequences for land use and land rents (Fujita *et al*., 1999b:3). Weber (1909) focused on the location of modern manufacturing industries at that time. Although they focused on different aspects of the location of economic activity, Von Thünen (1826) and Weber's (1909) goals were similar (Brakman *et al*., 2009:44). According to Aoyama *et al.* (2011:2) they wanted to create optimal location patterns for the most effective operating farms, factories and cities, considering geographical endowments, accessibility

[11] Colonialism and imperialism refer to the political and economic domination of another country, but they differ in the sense that colonialism refers to application and imperialism to the idea (Oxford Dictionary, 1964).

[12] Marshall was one of the first to introduce the incidence of industrial agglomerations and also emphasised the influence of economies of scale (Aoyama *et al*., 2011:3; Fujita, Krugman & Venables, 1999:18).

[13] In-depth discussion of increasing returns to scale will follow later on in the dissertation.

[14] Increasing returns to scale refer to the decline in average costs as output or production takes place within a specific area (Anas *et al*., 1998:1427).

(like transport costs) and trade-offs that needed to be made by the locater. According to Anas *et al*. (1998:1435) a number of facts about urban spatial structure are in line with the monocentric model and theory of Weber. First, the density of population has an inverse relationship with the distance to the central business centre, in other words, population decreases as distance increases. Second, most major cities in the Western world decentralised in the 1900s as a result of people starting to locate further away from city centres due to the fall in transport costs. There are, however, some significant limitations to the monocentric model. According to Brakman *et al*. (2009:36-37), firstly, the interaction between cities is excluded from the model and secondly it doesn't answer the question as to why there is a city in the first place. According to urban economists the introduction of increasing returns to scale is necessary in order to deal with the abovementioned limitations of the model (Glaeser, 2007:15-16).

During Weber's (1909) era there was also another important book published by George G. Chisholm. The Scottish geographer was the author of perhaps the first English-language geographical economics textbook called the *Handbook of Commercial Geography* in 1889 (Barnes, 2001:165). The handbook, which had a great impact in educating Britain in economic geography, contained an abundance of information regarding world commodity production and geographical conditions for trade in a time where Britain was the production centre of the world and dependent on growing international trade. Chisholm (1889) emphasised the importance of local geographical factors and their interrelationship with the geography of economic activities. Chisholm's role as educator had a great impact in the promotion and study of the relatively new discipline of geographical economics.[15] While Chisholm focused more on the empirical and statistical side of geography, an American transportation economist, J. Russel Smith, followed a different approach on the other side of the Atlantic (Barnes, 2000:14-16). At that time it was evident that not only economic geography but also the economic history required further consideration after the historical school had introduced the concept of relativity as to time and space (Robinson, 1909:249). Interest grew in geographical economics as America moved from an agrarian to an urbanised and industrialised society. Consequently, the number of institutions with courses concerning geography oriented towards economics increased and by the early twentieth century geographical economics had become a familiar university-level subject (Fellmann, 1986:316-

[15] By the time he died in 1930 economic geographical discourse was a well-established field (Barnes, 2000:16).

317). One such an institution was Wharton School at the University of Pennsylvania where Smith (1913) was founded within the Department of Geography and Industry (Barnes, 2000:16). In 1913 he published the handbook *Industrial and commercial geography* which was a new and upgraded American version of Chisholm's book (Starkey, 1967:200). The book discusses the production of certain resources and manufactured goods as well as international trade. Smith's (1913) work differs from Chisholm's (1889) in the sense that he focuses on technological changes regarding transportation and communication, which was a more dynamic and analytical approach (Barnes, 2000:16-17; Fellmann, 1986:319). Smith, like his counterpart, was trying to invent his own distinct discipline (Barnes, 2001:165). He was one of the new generation of American geographical economists who were influenced and intellectually challenged by German historicism and wanted to steer the discipline in a new direction (Fellmann, 1986:316).[16]

By taking Chisholm (1889) and Smith (1913) as point of reference, geographical economics was rooted within empirical detail, a world-wide geographical classification based on commodity specialisation, and the conditions and spatial patterns of commercial trade. Both authors ensured that the discipline was not based solely upon abstract theory, which was mostly the path of economics. The fact that it did not follow the same route emphasised the importance of historical and geographical context. However, the discipline changed again some fifteen years after Smith's book was published. By the 1930s the focus shifted towards the geography of regions, unlike Smith's focus on transportation, commodities and trade. The movement to region was in part due to a persistent debate in the literature of geographical economics (Barnes, 2000:18). Some argued that the focus should be geographical (Whitbeck, 1915-16:197). A proponent of this school of thought was Ray Whitbeck, an American economic geographer, who wrote a critical review of *Industrial and commercial geography,* written by Smith. Whitbeck (1914:540) stated that Smith's textbook concentrated on commerce and industry and did not suffice as a textbook for commercial and industrial geography. He said that the measure should not be the commodity, but needed to be the country or region (Whitbeck, 1915-16:197). Whitbeck's influence was large and consequently a regional perspective became the focus of inquiry of geographical economics. From the mid-1920s onwards, that perspective dominated various textbooks (Barnes, 2000: 18). A book like *Economic geography* by Whitbeck and Finch (1924) justifies the regional perspective of the discipline based upon the fact that it provides insight on regional

[16] German Historicism refers to the likes of Von Thünen (1826) and Weber (1909).

differences that "educated people need and use" (Whitbeck & Finch, 1935). Another example was C.F. Jones's (1935) textbook, also titled *Economic Geography*, which compares facts from various regions by following a grid of categorisation first used by Whitbeck and Finch (1924). The facts of each region were ordered under categories like agriculture, minerals, manufacture, commercial trade, communication and transportation. The differences in regions were evident out of the categories and facts (Barnes, 2000:19).

Regional categorisation furthermore had a meaningful impact on geographical economics through the work of German geographer, Walter Christaller, and German professor of spatial economics, August Lösch (Barnes, 2012:8-11).[17] All the German contributions, including Von Thünen (1826) and Weber (1909), analyse where location of economic activity is situated by taking the national or economy-wide space into consideration. This is a question of import since transportation associated with people and goods is not costless and production is related to some type of increasing returns to scale (Brakman *et al*., 2009:44). A good example of the impact of transportation and production costs is shown by historian Albion (1939) who explained that New York City gained initial advantages out of the Erie Canal and the innovations introduced by merchants in the early 1800s. The development of the city was consequently spurred on by the low transportation costs and increasing returns to scale in production. One can, however, come to the conclusion that the function of cities changes over time, since the advantages of the city are now due to existing agglomeration to certain industries, especially communication and finance (Fujita & Krugman, 2004:141). The concepts first invented and tested by Christaller (1933) and Lösch (1940) not only try to explain the differentiation of various functions of cities that they perform, but also try to explain the location of cities and determine the relationship between cities and "non-cities" (Brakman *et al*., 2009:44). Christaller's book, *Die Zentralen Orte in Süddeutschland* (The central places of southern Germany), and Lösch's book, *Die Räumliche Ordnung der Wirtschaft* (The economics of location), were the first to depict the abovementioned approach, which is known as central place theory. The approach shows that different locations of cities on the economic landscape have various degrees of centrality and that products and services are optimally provided on hierarchical grounds (Mulligan, 1984:4). The provision of the products and services is determined by increasing returns to scale, while the location is important since it affects transportation costs of consumers, who want to minimise

[17] Christaller and Lösch followed the German location theories of Von Thünen and Weber (Aoyama *et al*., 2011:2).

these costs by accessing nearby suppliers. Accordingly many small locations (e.g. towns) deliver only a few basic functions, while a few large locations (e.g. cities and central areas) deliver multiple functions within the economic landscape (Brakman *et al*., 2009:44-45).[18] Christaller and Lösch's approach in central place theory was in contrast to the Germanic tradition of regional geography. The theoretical inclinations and use of laws, mathematics and regularities were irregular. Central place theory was rationalist in the sense that it was based upon logical deduction, exchanged in the vocabulary of laws, intended intervention and was concerned with prediction and explanation (Barnes, 2012:9-15). Christaller's (1933) search for laws was clear in his introduction to his book. His assumption was that because economic laws existed to determine the existence of the economy, there should also be specific economic geographical laws which explain the sizes, distribution and number of towns. Consequently, he came to the conclusion that it would not be senseless to search for such laws.[19]

A distinct advantage of the central place theory is that it is concerned with the location of economic activity. A major problem with the theory is that the economic rationale behind the individual agents' (e.g. consumers and firms) decisions is absent, and consequently the central place outcome is not explained. Imperfect competition needed to be introduced in conjunction with increasing returns to scale at the firm level in order to explain the outcome of locations (Brakman *et al*., 2009:46). As a result central place theory, as a regional approach to geographical economics, was prescientific and is contemporarily used to illustrate a descriptive story, since it does not constitute a causal model (Barnes, 2012:3; (Fujita *et al*., 1999b:27). However, the regionalist perspective constituted a shift in the discourse of geographical economics as the object of inquiry changed. Richard Hartshorne, an American geographer, played a significant role in the evolution of the regionalist perspective.[20] He may not have been the first to develop regionalism, but he gave structure and gave an accurate justification to it. He was the first to give economic geography a considerable purpose (Barnes, 2000:19). Hartshorne is best known for his work on the history and philosophy of economic geography, and, especially, his book, *The Nature of Geography* (1939). The book was rooted in economic geographical concerns and followed an approach which disregarded the methodological approach of social sciences like economics (Barnes,

[18] The economy can only sustain a certain number of locations. A hierarchy of locations is provided where large centres perform more functions, while small centres provide only some functions (Brakman *et al*., 2009:44-45).

[19] The significance of such a law, Zipf's law, will later be the focus of the dissertation.

[20] Hartshorne is maybe the geographer most associated with regionalism (Barnes, 2000:19).

2001:165). The main aim of the book was to determine a geographical unit which could codify and incorporate different unorganised pieces of information gathered by geographers (Barnes, 2000:19). Hartshorne (1939) determined that the region as geographical unit enabled geographers to integrate dissimilar geographical facts which made the complexity of the world understandable. He defined region in such a manner that economic speciality was fundamental for delimitation of the specific region. His regional perspective is justified on the grounds that there is no boundary between economic and regional geography. The world according to him consists of economically defined regions, which interconnect with one another, and each region is made up of a complex of interconnected elements (Hartshorne, 1939:334-408).[21] The complex interaction among the various elements within a region is important, since it creates a distinctive regional entity. Hartshorne's (1939) work had a couple of fundamental differences to the work of Chisholm (1889) and Smith (1913), and it was important for the reinvention of geographical economics. According to Barnes (2000:20-21) firstly, regionalism is depicted by the geographically unique. Because of the uniqueness of the region, the object of inquiry of geography could not be generalised in the form of "exact" or "natural" sciences (Barnes, 2012:11). Unlike Christaller, Hartshorne (1939) implies that geographical economics cannot be a predictive science, and must be a descriptive science, since regions' characteristics can differ substantially. On the other hand Chisholm and Smith did rely on description and geographical difference, but emphasise the economic and geographical system that connects regions together.[22] Secondly, Hartshorne's principle of regional uniqueness pointed to the importance of determining the elements out of which a region was comprised. He then used different typologies to demonstrate that these elements came in different mixtures and subsequently formed different regions with unique characteristics. Chisholm and Smith are not as concerned with typologies. They only focus on production of commodities and trade of a certain place, with the aim of emphasising geographical connectivity. Lastly, for Hartshorne, it was important to be in contact with the immediate geographical region in order to understand its complexity. For Chisholm and Smith physical geography was not an essential part of their work. The reasons for geographical economics' regionalist transformation may be due to the changing historical and geographical context. The slowing down of colonialism, the impact of the 1930s depression

[21] The elements are, for example, lands and plants, tools and methods of production, livestock and buildings, and also intangible components, such as, markets and prices, and knowledge.

[22] The system relies on trade and world production.

on trade and commerce, and later on the World War (WW) II were significant factors that influenced the field (Barnes, 2000:20-21; Aoyama *et al*., 2011:3).

2.2.2 The quantitative revolution and the shift to a spatial science

The idea of external economies was first articulated by Alfred Marshall. He highlighted the concept by considering the advantages of producing within an industrial district (Aoyama *et al*., 2011:3). Marshall recognised three reasons why producers found it advantageous to locate in the same area as other producers in the same industry. Firstly, industries that are geographically clustered could assist specialised local suppliers of inputs. Secondly, a large pool of labour existed since workers of the same type were employed by the concentration of firms. Thirdly, the sharing of information would be promoted as a result of geographic proximity (Brakman *et al*., 2009:41; Fujita *et al*., 1999b:18).[23] Consequently, spatial concentration has been associated with external economies since its inception (Aoyama *et al*., 2011:2-3; Fujita *et al*., 1999b:18). However, external economies were only given a fundamental role in urban theory at a later stage through the works of Hoover (1948), Isard (1949, 1953) and Berry and Garrison (1958b).[24] During that same period descriptive regional economic geography started to transform when economic geographers took up rationalist theories and methods (Barnes, 2012:3). Fred K. Schaefer (1953) was the first to reject Hartshorne's stance that economic geography could not be a predictive science, and called for a scientific approach based upon the search for geographical laws. Schaefer's article, *Exceptionalism in geography*, was the beginning of a movement by the new generation of economic geographers to reinvent the discipline as a science (Barnes, 2000:20-21).[25] These occurrences marked the beginning of the quantitative revolution and economic geography became an actual science later known as spatial science (Barnes, 2000:21, 2011a, 2011b). The discipline started to shift from specific descriptive analysis to predictive scientific analysis as economic geographers started to explore and apply economic theories, methods

[23] However it has been difficult to formalise Marshall's trio of external economies due to technical reasons (Fujita & Krugman, 2004:142; Fujita *et al*., 1999b:18). This is due to the fact that agglomeration in terms of micro-decisions relies on external economies at the level of the individual producer, which in turn translates into the fact that perfect competition cannot be assumed like in the Von Thünen model (1826). Imperfect competition needs to be assumed in a general-equilibrium model along with an integrated depiction of transportation costs, which is a difficult thing to do (Fujita & Krugman, 2004:142). But, Marshall's argument showed urban economists that they did understand why central business districts and cities were in existence, and furthermore by including external economies into their models they could give important analysis in view of the whole economy as a system of cities (Fujita *et al*., 1999b:18-19).

[24] Berry and Garrison (1958a) also published an article, *Alternate explanations of urban rank-size relationships*, which discusses the correspondence between Zipf's law and the theories of Christaller, Rashevsky and Simon (Berry & Garrison, 1958b:83).

[25] Laws can be considered to be the epitome of a scientific generalisation (Barnes, 2000:21).

and techniques for the first time (Barnes, 2001:165). Consequently, the economical geographical discourse was also redesigned and the economic geographical world was explained by abstract space, Greek symbols, geometrical axioms and regression lines (Barnes, 2000:22).The object of the analysis was to develop generalised theories in order to explain practical geographical phenomena like (a) industrial location and regional growth, (b) patterns of urbanisation, and (c) spatial flows and interactions (Aoyama *et al*., 2011:4; Barnes, 2001:165; Scott, 2000:485).

Another movement during the 1950s that secured economic geography's direction towards spatial science was regional science, led by the American economist Walter Isard (Barnes, 2001:166; 2000:23). According to Isard (1956:25) the aim of regional science was to incorporate spatial relationships into, as he described, the previous "wonderland of no dimensions" (Barnes, 2000:23). The lengthy tradition of analysis that came from Von Thünen (1826) described the pattern of land use around the central business district and cities, but the existence of such a central focus was taken as a given (Fujita & Krugman, 2004:141; Fujita *et al*., 1999b:3-18). This is why the shift towards spatial science was to a large extent inspired by the works of the Germanic location theorists like Weber (1909), Christaller (1933) and Lösch (1940), who failed to explain the interaction of decisions by individual firms or families (Fujita *et al*., 1999b:27). The regional scientists were aware of the limitations of the traditional location theories and accordingly they responded (Martin, 1999:66). The reactions of quantitative theoretical economic and regional scientists were to provide abstract and universal explanations for regional economic evolution and spatial theories of industrial location (Aoyama *et al*., 2011:4). Isard (1956, 1960) and his regional scientist followers' work attempted to build on the fundamental concepts of the central place theory, since they wanted to give a formalised theoretical economic basis for this theory (Mulligan, 1984). They were creating a new hybrid discipline, which combined economic elements with elements of geography. The purpose of this hybrid regional science was to revise neoclassical competitive equilibrium theory in accordance with spatial arrangements in order to derive a function of location that included all supplies, demands and price variables. They also focused on distinguishing the regularities of the neoclassical spatial economy (Scott, 2000:486).[26] However, limitations were also evident out of most of these new models since they were

[26] Zipf's law is an example of such regularity.

often of partial equilibrium nature.[27] Essentially the geometric pattern of the central place system was formalised, but the individual agents' underlying behaviour and also their decisions and market interactions were still not entirely explained. Accordingly the central place outcome was simply rationalised (Brakman *et al*., 2009:46). The reason why the reinvention happened at that time may be found in the fact that tertiary education expanded after the WW II. This expansion might have been due to the fact that structural changes in economies and societies were occurring and increased funding was given to educational attainment. Furthermore, the 1950s was characterised by optimism in science and technology (Barnes, 2000:23-24). It was a time when geographical economics as a discipline was dominated by economics (Barnes, 2001:166).

In the 1960s the Von Thünen (1826) model had an important revival through the work of William Alonso (1964), which most likely marked the beginning of urban economics as a separate discipline. He reinterpreted the model by replacing farmers by commuters and the isolated town or city by a central business centre (Fujita *et al*., 1999b:17; Brakman *et al*., 2009:36). Alonso's (1964) spatial equilibrium model, which was expanded by Mills (1967) and Muth (1969) to form the Alonso-Muth-Mills model, continues to be the essence of contemporary urban economics. The model is based upon the assumption that, within a metropolitan area, amenities and income are unchanged. The implication of these assumptions is that housing costs plus transport costs are spatially constant, which means the cost of housing will decrease as distance to the central business centre increases, and consequently the transport cost increases. Subsequently, commuters derive a utility from space for living but face a trade-off in the form of land rents, which are higher near the central centre and decline with distance. This implies that the bid-rent method can be used, and that the efficient allocation of land occurs as a result of competition between commuters for land (Glaeser, 2007:6; Brakman *et al*., 2009:36).[28] As mentioned before in this chapter, the monocentric model has limitations, but the relevance of the model for understanding urban spatial structures and the choice of location was significant, since much of the research's goal was to establish policy benchmarks for the recommendation of governmental decision makers and planner (Scott, 2000:486-487). However, the introduction of increasing returns to scale in these models were still necessitated (Glaeser, 2007:15-16).

---

[27] A partial equilibrium model explains some aspects of a phenomenon but ignores other relevant aspects, or assumes them (Brakman *et al*., 2009:46).

[28] Glaeser (2007) confirms the relevance of the Alonso-Muth-Mills model in contemporary cases. Although other factors also influence the relationship, the prediction the model makes cannot be rejected.

2.2.3 1970s to 1990s: Urban system theories and industrial organisation revolution

Regional science and spatial analysis began to be rejected as a whole by economic geographers in the early 1970s (Barnes, 2001:166; Isserman, 1993; Scott, 2000:487). According to Thisse (1997) the reason for this was that their models reached their height because they were based upon perfect competition and restricted ideas of convexity.[29] On the other hand research on urban economics with theoretical models of structure of metropolitan areas, like Alonso's (1964) work, along with transportation and accessibility research dominated the discipline, and the inclination remained throughout the 1970s (Aoyama *et al*., 2011:4). During that period an important strand of work was published by Henderson (1974, 1977, 1988), following Mills (1967) and Muth's (1969) writings, who extended the single city framework to a framework of the system of cities (Tabuchi, 1998:335). The work still contemporarily remains the basis of research regarding the actual distribution of sizes and types of urban areas and, unlike the monocentric model, external economies to scale are specifically the focus (Fujita *et al*., 1999b:19; Brakman *et al*., 2009:40). Henderson's model assumes that there are no interurban transport costs, intraurban commuting costs are relevant and space outside the city has no impact on the analysis (Tabuchi, 1998:335). The reason for this absence of non-city space is that modern developed countries' populations and economic activities are mostly clustered in urban areas as a result of industrialisation and urbanisation. Accordingly it is assumed that the transactions between urban and non-urban or rural areas are insignificant (Brakman *et al*., 2009:40-41).

The focus of analysis is the agglomerating forces that determine the size and interactions between cities. These agglomerating forces are the industry-specific positive external economies of scale (Brakman *et al*., 2009:40-41). On the other hand, the positive external economies, which cause centripetal motion of industries within a city, are met by diseconomies, like congestion, caused by large cities. The diseconomies are negative external economies of scale and lead to spreading of population and economic activity. Henderson's assertion is that the net effect of the trade-off between the positive and negative external economies of scale determines the relationship between the utility of a resident and city size, which in turn creates an equilibrium point that determines the optimum size of a city (Fujita *et al*., 1999b:19-20). But, if he argues that each city is of optimal size, what determines the different sizes of cities? The negative external economies of scale are a function of overall

[29] According to Scott (2000:487) another reason was the fact that regional science and spatial analysis' persuasiveness was lost due to wider political reasons.

city size. This means that a large city will have high costs, such as commuting costs and land rents. However, the diseconomies do not rely on the type of production in the city. Accordingly, overall size of the city in conjunction with industry specific positive external economies has two significant implications. Firstly, cities have different sizes since they need to accommodate different industries and perform different functions. Consequently, the systems of cities are rationalised based on the assumption that each industry has its optimum size since positive spatial spill-overs are industry specific.[30] Secondly, a rationale for specialisation is provided by giving a micro-economic base for increasing returns, and subsequently an urban network is established in which cities of differing sizes interact with each other. It does not make sense not locating similar industries together in a single complex, since they benefit from each other and to some extent negate the effect of diseconomies of scale.[31] Furthermore, the extent of these external economies varies considerably over different industries (Fujita *et al*., 1999b:19-21; Brakman *et al*., 2009:40-41). In comparison with Alonso and (1964) and Von Thünen's (1826) models, modern urban economics, through Henderson's framework, is more multi-purposed since it gives a theoretical basis for external economies that determine the existence of cities and gives meaningful insight as to urban systems (Brakman *et al*., 2009:41-42).

Another revolution began at the end of the 1970s in the field of industrial organisation. Theorists for the first time were trying to create manageable models of competition that could be combined with increasing returns to scale (Fujita *et al*., 1999b:3). Avinash Dixit and Joseph Stiglitz's (1977) work was especially important in this aspect as they developed a Chamberlinian monopolistic competition model, parameterised using their constant-elasticity-of-substitution individual utility function (Neary, 2001:537). The model would revolutionise theoretical modelling in a number of fields, including industrial organisation, trade theory, growth theory and geographical economics (Brakman *et al*., 2009:93). To date the Dixit-Stiglitz model forms a significant foundation of economic theory in economic

[30] Henderson argues that in equilibrium each city is of optimal size and the utility of the residents is maximised (Brakman *et al*., 2009:41). If a city is not of optimum size (e.g. too few typically large cities) a profit opportunity would be created (Fujita *et al*., 1999b:20). Consequently, if the cities are too large the profit opportunity would be for people to move out of the city and seek welfare in another smaller city. The effect on the large city would also then be welfare improvement (Brakman *et al*., 2009:41). Henderson (1988) introduced a city entrepreneur who gives people the incentive to move as a result of the profit opportunity he gives in a smaller or larger city. According to Becker and Henderson (2000) these entrepreneurs are necessary, since individuals in a city of non-optimal size do not have an incentive to relocate on his or her own.

[31] These mutual spill-overs are positive external economies which occur due to the same reasons as previously noted by Alfred Marshall.

growth, international trade and contemporary economic geography. The model assumes that a large amount of goods (though the consumers regard these as distinctive products) enter symmetrically into demand and also that the utility function takes a specific yet unorthodox form (Fujita *et al*., 1999b:6). Dixit and Stiglitz (1977) firstly assume that the consumers love a variety of goods and firms require a fixed amount of limited productive resources, accordingly the "love-of-variety" effect is captured by the Constant Elasticity of Substitution (CES) utility function that is symmetrical in a package of differentiated products.[32] Furthermore, each firm is assumed to have no impact on overall market conditions (Ottaviano, Tabuchi & Thisse, 2002:410).[33] Secondly, transportation is assumed to be a costly activity since a fraction of the goods does not arrive at the destination when goods are moved between regions (Brakman *et al*., 2009:106). These transportation costs are known as iceberg transport costs, since a constant fraction of output is lost or "melts away" (Neary, 2001:539). The cost of transportation is represented by the fraction of goods that does not arrive (Brakman *et al*., 2009:106-107).[34] Thirdly, in order to select spatial equilibria, stability analysis is used which is based upon myopic adjustment processes wherein the location of factors of mobility is determined by differences in current returns. Lastly, numerical analysis needs to be used since analytical resolution of the models is not feasible (Ottaviano *et al*., 2002:410).

According to Fujita *et al*. (1999b:6) the advantages of the Dixit-Stiglitz model are that it makes the analysis of geographic issues manageable without taking the relevance of that analysis away. However, some limitations are evident out of the model: the model may deliver unrealistic results, since the elasticity of substitution ends up representing an index of returns to scale, which in turn raises questions in interpreting the results and stresses the indistinct role played by individual firms in the model; equilibrium prices are independent of the spatial location of consumers and firms, which conflicts with spatial pricing theory that emphasises that demand elasticity differs with distance while the level of demand and intensity of competition influence price changes; the iceberg transport cost assumption makes the implication that an increase in the price of the transported product is associated with proportional increase in the cost of trade and also that the technology used to transport the good is indistinguishable to that used to produce it, which is also unrealistic; and the role of

[32] When the extent of the market increases the number of varieties in the market from which the consumer can choose increases, which translates into a proportional increase in utility that the individual can derive (Brakman *et al*., 2009:92-93).

[33] Firms ignore the impact of their actions on income and industry price index (Neary, 2001:538).

[34] Iceberg transport costs were first introduced by Paul Samuelson (1952).

expectations is absent in the analysis which is an essential component of individual agent location decisions (Fujita *et al*., 1999b:7-8; Krugman, 1991; Matsuyama, 1991; Neary, 2001:548-552; Ottaviano *et al*., 2002:410). Although the Dixit-Stiglitz approach ignores some questions and limitations may exist, the importance of the model is emphasised by Krugman (1995) who states that formalism in geographical economics slowed down progress in spatial economics and development and furthermore that before the Dixit-Stiglitz model there was no efficient way in modelling imperfect competition with increasing returns (Neary, 2001:553; Fujita *et al*., 1999b:3).

At the end of 1970s, theorists started applying new industrial organisation theory, through its analytical tools, to international trade; a couple of years later to technological change as well as economic growth. New ideas needed to be formulated in order to apply the Dixit-Stiglitz framework and at first authors were creating seemingly inconsistent models in which direction and motivation were lacking. However, eventually a core of useful insights emerged which created new tightly integrated fields such as "new trade" and "new growth" theory (Fujita *et al*., 1999b:3).[35] From the 1980s economic geography shifted away from the traditional economic analysis and changed into a more interdisciplinary approach which uses insights from political, social and cultural sciences. In economic geography this movement has been known as the "institutional turn" or the "cultural turn" which found its origin in French and German philosophy (Aoyama *et al*., 2011:5; Boschma & Frenken, 2006:273). At the same time, the recognition that innovation and technological change hold significance for job creation and new industries, many economic geographers' focus shifted towards the various connections between innovation and economic growth. Furthermore, the 1980s marked growing interest in organisational aspects of manufacturing or production through the so-called post-Fordist industries.[36] These industries were known for a relatively high degree of intra-local business networking, spatial agglomeration, innovation and growth. The emphasis was mainly on high-technology manufacturing industries and their geographic consequences, as well as an important shift towards determining policies for urban and regional growth in the national and global contexts (Aoyama *et al*., 2011:5; Scott, 2000:492). Thus, since the 1970s economic geography has become even more heterodox by incorporating ideas from French regulation theory; Schumpeterian models of technological evolution; ideological orientations such as Ricardian, Marxian, Keynesian, Polanyian, and

[35] The second half of the 1980s saw the emergence of "new trade" and "new growth" theory (Fujita & Krugman, 2004:142).

[36] Fordism is named after Henry Ford and refers to mass production in a modern economic system.

Gramscian; and also neoclassical, institutional, cultural and evolutionary approaches to combine in the discipline (Aoyama *et al*., 2011:6; Martin, 1999:66).

2.2.4 Krugman's core model of geographical economics

Since the 1990s economic geography has reflected the multiple changes that had occurred since the 1970s (Aoyama *et al*., 2011:6). The 1990s marked the beginning of the "new economic geography" which is considered the fourth wave of the increasing returns revolution in economics (Fujita *et al*., 1999b:3). The reinvention of economic geography, the New Economic Geography (NEG), was claimed by various groups, including the group that provided the epistemological constituent of the "cultural turn" and the group who focused on econometric analysis, importantly impacted on by the seminal contribution of Paul Krugman (1991a). Krugman, along with writers like Michael Porter, Robert Barro, Xavier Sala-i-Martin and Anthony Venables, signified a movement of economists who intended to reconcile economics with geography, after the two disciplines had been relatively separated for a while (Martin, 1999:66; Boschma & Frenken, 2006:273; Fujita *et al*., 1999b; Duranton & Storper, 2006:1).[37] Krugman's (1991) and other associated applications of neoclassical economics in economic geography is the latest revolution that has transformed economic geography (Boschma & Frenken, 2006:275).[38] The main contribution of the new economic geography is a framework in which fundamental aspects of mainstream economics, particularly general equilibrium models and rational decision making, are used to model the trade-off between centripetal forces, that pull economic activity together and cause agglomeration, and centrifugal forces that disperse economic activity (Neary, 2001:536; Fujita & Krugman, 2004:141). The framework aids in explaining why a geographical structure of an economy is formed in a certain way by the tension between the opposing forces, and it provides an explanation for these forces in terms of more fundamental micro-decisions (Fujita & Krugman, 2004:141).

2.2.4.1 Why is the model not the core model of New Economic Geography

In 1991, Krugman (1991a, 1991b) developed a general equilibrium core-periphery or agglomeration model which to date is described as the core model of geographical economics (Brakman *et al*., 2009:81-86). The question may arise why the model is not described here as the core model of the NEG? The term NEG was after all proposed by Krugman who called

[37] As Boschma and Frenken (2006:273) state, neoclassical economists have re-entered the discipline of economic geography.

[38] NEG was launched by Krugman in 1991 and extended by articles written by himself, Masahisa Fujita, Tony Venables and other associates (Neary, 2001:536).

NEG his "current research project" (Krugman, 1998).[39] The answer may come from an earlier time when Ohlin (1933) stated that international trade on the one hand and urban and regional economics on the other hand had, essentially, identical research objectives. Both research fields want to explain *who* produces what, *why* and *where*. Although, international trade and regional and urban economics evolved differently over time, they still provide valuable literature regarding their respective fields. Geographical economics productively combines both with mainstream economic theory to determine the *who*, *why* and *where* of the location of economic activity (Brakman *et al*., 2009:23-81).

2.2.4.2 The influence of new trade theory – Krugman's 1979 and 1980 models

The work of Dixit and Stiglitz (1977), on monopolistic competition, and Krugman's (1979, 1980), on international trade theory, constitutes a large part of the core model of geographical economics (Neary, 2004:159). The latter two articles were later expanded by Krugman (1991a) – the core geographical model – by adding interregional factor mobility, and together the three articles earned him the Nobel Prize in Economics in 2008.[40] Krugman (1979, 1980) developed a model wherein countries engage in trade which is welfare-enhancing, even though no comparative advantage is evident. In other words, the models were an extension of the new trade theory, which did not depend on comparative advantage. The challenge of the Krugman (1979) model was that increasing returns to scale implied an imperfect competition market structure, which in turn translated into the problem of providing a trade model with imperfect competition.[41] Fortunately, Krugman could use the model by Dixit-Stiglitz (1977) on monopolistic competition that had been published just a few years earlier (Brakman *et al*., 2009:53-61). Imperfect competition is also dealt with pure external, instead of internal, economies of scale which is a characterisation of new trade theory since it allows geographical economics to use perfect competition as a market structure (Helpman, 1984; Helpman & Krugman, 1985).[42]

---

39 Fujita *et al*. (1999) states that they owe their intellectual debt to regional science and location theory (Neary, 2001:553). According to Scott (2000:487) Krugman (1991a, 1991b) reinvented the inquiry of regional science and spatial analysis in the likeness of geographical economics.

40 Krugman's (1991a) model was in a number of ways an expansion of new trade theory where trade between countries does not rely on comparative advantage and inter-industry trade is replaced by intra-industry trade (Brakman *et al*., 2009:53-59).

41 This was, as previously mentioned, difficult for urban and regional economics.

42 In urban economics, new growth and new trade theory pure external or technological economies of scale are used. With the pure external economies the technological relationship between inputs and output for each individual firm is changed by an increase in industry-wide output. An example of this would be positive information spill-overs for each firm, which would result in higher output for the

The Krugman (1980) model was an important link between the new trade model (in Krugman, 1979) and the core model of geographical economics (Krugman, 1991a). In the 1980 model new trade theory was applied to the phenomenon of industrial agglomeration (Neary, 2004:172). The 1980 model's rationale for intra-industry trade remained the same as in the 1979 model, with some important exceptions. Firstly, in the 1980 model the increase in the market size as a result of freeing up of trade does not induce an expansion in the production scale, although increasing returns to scale at the firm level are evident. Alternatively, the scale of production of each variety remains the same under trade and autarky, and prices remain constant. The consumers can choose more varieties under trade than autarky, and consequently the benefits of trade are due to the love-of-variety effect (Brakman *et al*., 2009:63). Secondly, the 1980 model allows for transport costs on monopolistically competitive goods when nations trade amongst one another (Neary, 2004:172). Thirdly, the uneven distribution of market size in conjunction with transport costs generates a home-market-effect. The effect states that countries will produce goods and export them if they have a relatively high demand for them in their country (Neary, 2001:541). This is due to the fact that home-produced varieties minimise transport costs, whereas imported goods incur higher transport costs. Accordingly, the location of manufacturing firms and industries are relevant and the concentration of economic activity can be a consequence of the model (Brakman *et al*., 2009:63-64; Neary, 2004:172-173).

2.2.4.3 The core model – Krugman's article: Increasing returns and economic geography

In 1991 Paul Krugman published, *Increasing returns and economic geography*, in *The Journal of Political Economy* (Brakman *et al*., 2009:52). Krugman (1991a) had developed a two-region general equilibrium model in a monopolistic competition market structure by expanding his previous work (Krugman, 1979, 1980) by introducing interregional factor mobility (Tabuchi, 1998:334; Brakman *et al*., 2009:86). The writings of Krugman (1991a) draw attention to some new research directions by emphasising the importance of the effects of increasing returns and imperfect competition as basis of determination of location of agglomeration (Scott, 2000:487). Krugman's (1991a) approach can be considered a modern application of neoclassical thought to explain agglomeration, trade and specialisation. Basically, it is micro-economic theory that explains the persistence and existence of

---

entire industry. Consequently, the size of the individual firm does not matter and the market structure can be perfectly competitive (Brakman *et al*., 2009:38).

agglomerations according to the rationale decisions of economic agents (Boschma & Frenken, 2006:275). The fact that it took so long to combine transportation costs, imperfect competition and increasing returns to scale in a general equilibrium framework emphasises the difficulty and the significance of the model (Brakman *et al*., 2009:52). According to Tabuchi (1998:334) the model enables the analysis of the influence of an alteration in the interregional transport costs on the intensity of urban agglomeration. Furthermore, Krugman (1991a) demonstrates that agglomeration takes place with low transportation costs, since firms and workers cluster to reap the benefits of Marshallian externalities, whereas high transportation costs cause firms and workers to disperse, since the dispersion force is dominant relative to the agglomeration force. This outlines an important implication: agglomeration tends to be dominant most of the time, since technological progress and innovation reduce transportation costs.[43] With falling transportation costs, an important transition point is reached when workers and firms realise it is more profitable to cluster in a single region (or city) rather than to spread out over many regions. The transition point relies on the trade-off between increasing returns to scale for firms and the love-of-variety effect of consumers due to clustering on the one hand and interregional transportation costs on the other hand (Boschma & Frenken, 2006:275; Fujita *et al*., 1999b; Neary, 2001).[44]

2.2.4.4 What impact did the model have on the discipline of geographical economics

To understand the impact of Krugman's (1991a) work on the field of geographical economics we refer to the survey of Ottaviano and Thisse (2004) on the theory of agglomeration. They asked the important question of where geographical economics stood in 1990, before the core model was introduced. They noticed that all the elements were already existent in location theory before the model, apart from one critical element (Brakman *et al*., 2009:52-53). Ottaviano and Thisse (2004:2576) summed up the elements originating from existing location theory in the following five points:

(i) The trade-off between different forms of external economies and various types of mobility costs determine the end result of economic space;

(ii) land use, high transport costs and price competition promotes centrifugal movement of production and consumption; consequently

[43] Large cities have emerged around the world while rural areas lost population after the Industrial Revolution. However, Vining, Pallone and Plane (1981) documented that spreading of economic activity has been taking place in most developed countries since 1970 (Tabuchi, 1998:334).

[44] The Krugman model has also shown that it is extendable in a number of ways, to factors such as unemployment and congestion (Boschma & Frenken, 2006:275)

(iii) clustering of firms is to be expected within sizeable metropolitan areas when they sell distinguishable goods and transport costs are low;

(iv) consumers or workers find cities attractive since they provide specialised labour markets and a wide range of final goods; and

(v) urban agglomerations are the result of cumulative processes that include both the demand and supply sides.

Accordingly, the location of economic activity has to be understood as the result of interaction between agglomeration and dispersion forces, a concept designed previously by regional scientists and economic geographers, within a general equilibrium model taking market failures (e.g. imperfect competition) explicitly into account (Ottaviano & Thisse, 2004:2576).

The above-mentioned five points can be regarded as the core of geographical economics (Brakman *et al*., 2009:53). This leads us to the conclusion that the fundamental difference between existing location theories of regional and urban economics and the core model of geographical economics by Krugman (1991a) is that a general equilibrium framework with imperfect competition is provided by the latter, which connects the insights of the former (the above-mentioned points) and allows for the detailed study of the interactions between them (Ottaviano & Thisse, 2004:2576). In other words, the important contribution of geographical economics is to provide a single analytical framework to combine existing elements of location theory (Brakman *et al*., 2009:53; Krugman, 1998:9).

2.2.5 Geographical economics after Krugman (1991a)

As previously mentioned, the Krugman (1991a) model is widely regarded as the starting point of the NEG, and subsequently has stimulated the development of new theorising and empirical work in geographical economics by a number of influential writers (Krugman, 1998:7). Krugman continued to try to create a theory of economic localisation based on increasing returns to scale by means of books (for example, Krugman, 1991b, 1995, 1996a) and a considerable number of articles (such as, Krugman, 1991a, 1993a, 1993b, 1994, 1996b). According to him increasing returns are basically a local and regional phenomenon, to the extent that spatial economic agglomeration as well as specialisation is important enough for economic geography to be accepted as a crucial field within economics (Krugman, 1991b:33). Similarly, Michael Porter (1990, 1994, 1996) emphasised the importance of geographical clustering of industries as a determinant of a nation's global

competitiveness. Porter (1990:791), like Krugman, also argues that evidence exists for classifying economic geography as a central discipline in economics. These statements have led to growing interest in the geographical economics by other economists and consequently more and more research was devoted to the field (Martin, 1999:67).

Research into increasing returns within economics more generally continues to gain more interest. Various forms of increasing returns are implemented by different authors to explain the process of spatial agglomeration. For example, Krugman and Venables' (1996) models focus on Marshallian externalities which are sector-specific spill-overs and are also referred to as localisation economics. These externalities include technological spill-overs (like sharing of information or improvements in innovative technology), a large pool of labour to choose from and intermediate goods demand and supply linkages. These factors lead to the clustering of economic activity locally. On the other hand, other authors on the wider regional level use pecuniary externalities that are transmitted by the market by means of price effects and lead to a nationwide core-periphery pattern of economic development within countries. This entails that factor-market and product-market competition is centrifugal or spreading forces of economic activity, and accordingly transport costs and labour mobility are essential to determine whether agglomeration or dispersion is going to take place. The more immobile labour is and the higher the transport costs are, the more dispersion will prevail over agglomeration, and *vice versa* (Krugman, 1991a, 1991b, Krugman and Venables, 1996; Martin, 1999:68; Brakman *et al.*, 2009:38-43). Another interesting new contribution by Ottaviano *et al.* (2002) is an agglomeration model with linear instead of Dixit-Stiglitz preferences. They found that the fundamental insights remained similar, but by incorporating transport costs the model's results were enhanced (Ottaviano *et al.*, 2002; Neary, 2001:553).

Further variants of the core model of geographical economics include elements of "new growth theory" (Martin, 1999:68). The term has been recognised since the work of Paul Romer (1986, 1990), especially *Increasing returns and long-run growth*, and Robert Lucas (1988). The differences between new growth and neoclassical growth theory lie in the fact that the former attempts to endogenise economic growth and relaxes the assumption of diminishing returns to factors that accumulate (Van Marrewijk, 1999). The combination of the core model with new growth theory focuses either on inter-regional movements of human capital or localised technological innovation as the medium of spatial concentration of economic activity. For example, an endogenous growth model is used by Bertola (1993), in the Romer-Lucas form, to indicate the impact of labour migration and capital mobility,

through localised increasing returns to scale, has on the dynamic process of concentration of economic activity and growth (Martin, 1999:68-69). Furthermore, authors like Audretsch and Feldman (1996), Martin and Ottaviano (1996, 1999), Baldwin (1999) and Walz (1996a, 1996b) show how technological innovation can be embedded in growth models and explain how local research and development spill-overs can induce a process of spatial agglomeration (Neary, 2001:553-554; Martin, 1999:68-69).[45]

Much of the NEG has revolved around the question of regional distribution of economic activity; however, at the same time the same elemental increasing returns models have also been applied to model the development of city systems, an extension of Henderson's (1974, 1977, 1988) ideas through the work of Henderson himself (1996), Arthur (1994a) and Krugman (1993b, 1993c, 1996a, 1996b). To illustrate this, Krugman's models take it that populations cluster in cities as a result of higher wages, opportunities and a greater variety of products that can be found (love-of-variety effect), and at the same time firms cluster in these cities since they provide large markets for their products. These ideas are used, under certain assumptions about external economies and diseconomies of localisation, to obtain various urban geometries like central place systems (of the Christaller and Lösch type), urban land use patterns (Alonso-type), and, most importantly for this dissertation, empirical regularities like the log-linear 'rank-size' distribution of city sizes (Martin, 1999:69; Gabaix, 1999; Fujita, Krugman & Mori, 1999; Fujita & Krugman, 1995; Fujita & Thisse, 1996; Krugman, 1996a).

The explanation of spatial economics, urban systems and location of economic activity can now benefit from geographical economics literature (Kleynhans & Drewes, 2008:141). Geographical economics possesses the necessary mathematical, analytical and econometric tools as well as modelling "tricks" to address the questions of the spatial distribution of economic activity (Martin, 1999:75-76; Fujita *et al*., 1999a:6-7; Kleynhans & Drewes, 2008:141). However, all the abovementioned have not been fully utilised since empirical research in geographical economics to date does not provide much evidence against other location theories, and also has limited evidence on the relevance of "softer" location factors (for example, cultural, institutional, political, environmental and social) in the ascertainment of locational decisions (Brakman *et al*., 2009:520). If theory is not supported by empirical testing and evidence, it will not disappear but it will be much less persuasive (Neary,

[45] However, Martin and Ottaviano (1996, 1999), Baldwin (1999) and (Walz, 1996a, 1996b) used a Romer-Grossman-Helpman style model of endogenous growth, unlike Bertola (1993).

2001:555; Krugman, 1998:16). Fortunately, the increased availability of substantial micro-data sets and increased use of spatial econometrics has encouraged an increase in studies, indicating the relevance of location to a number of economic phenomena (Brakman *et al*., 2009:520; Neary, 2001:553-554).

2.3 Agglomeration and economic growth

The clustering of economic activities or spatial agglomeration and economic growth of a country are inextricable processes (Baldwin & Martin, 2003:3; Martin & Ottaviano, 2001:947). This fact is emphasised by Simon Kuznets (1973), the 1971 Nobel Prize winner, who associated modern economic growth with the spatial concentration of economic activities. Kuznets (1973:248-51) mentioned six characteristics of modern economic growth (of the now developed nations) and highlighted that the rate of structural transformation of the economy needed to be high in order to achieve growth. Structural transformation of the economy to a large extent influenced the location of economic activity. For instance, in the initial years of geographical economics, the economy was agrarian, and therefore endowments in natural-resources, climate and labour supply's relevance were considerable. However, as industrialisation developed during the twentieth century, the relevance of the location of industries and firms, production processes, wages, skills of labour, technological innovation, as well state policy in advancing industrialisation were more important (Aoyama *et al*., 2011:1-8). Furthermore, the strong positive relationship between growth and geographic agglomeration has been documented by economic historians (see for example Hohenberg & Lees, 1985).[46] In the case of the industrial revolution of the 1800s, the economic growth rate in Europe increased substantially. The growth was manifested in an increase in the rate of urbanisation as well as the formation of industrial concentrations that have existed until now. Another example is the recent increased growth in the last 20 years in China. The growth has occurred in conjunction with greater disparity between inland and coastal regions in the country (Martin & Ottaviano, 2001:947). Even in sub-Saharan Africa between 1970 and 1995 urbanisation led to increased Gross Domestic Product (GDP) growth, although negative GDP per capita was shown as a result of rapid population growth in these countries. Urbanisation was accompanied by increased growth in services and industries (World Bank, 2009:58-59). The results of Danny Quah (1997) regarding regional data from Greece, Spain, Portugal and Ireland also indicate the positive correlation between growth and

[46] Especially, the relation of growth and agglomeration to the industrial revolution during the 1800s in Europe was documented (Baldwin & Martin, 2003:3; Martin & Ottaviano, 2001:947).

agglomeration. He found that the two countries that achieved the highest growth rate and converged the most in terms of per capita income (Portugal and Spain), in comparison with the rest of Europe, also encountered the most regional convergence (Baldwin & Martin, 2003:3). Consequently, the link between economic growth and spatial agglomeration materialises at different levels of agglomeration, such as the city and industrial concentrations which can be across regional and national boundaries (Martin & Ottaviano, 2001:947-948).[47]

What is evident out of all these cases is that agglomeration and economic growth happen at the world level. Fujita and Thisse (1996) express the relationship between the two phenomena most adequately as they state that agglomeration can be regarded as the territorial equivalent of economic growth. Another important deduction that can be made is that agglomeration and economic growth are also strongly related to economic development, urbanisation and migration and also that state policy has a role to play in the promotion of these phenomena. The linkages between agglomeration and economic growth on the one hand and economic development, urbanisation and migration and also public policy will be discussed in this section.

2.3.1 Geographical economics and endogenous growth theory

The initial literature on agglomeration and economic growth was an amalgamation of endogenous growth literature with geographical economics (see Martin & Ottaviano, 1999; Baldwin, Martin & Ottaviano, 2001; and Baldwin & Martin, 2003). All the components of Krugman's (1991a) core model are present; however, the fundamental difference is that the emphasis is on the growth of capital, and its mobility. The connection between agglomeration and growth relies on capital mobility. This is an important statement since growth, in principle, occurs as a result of the accumulation of capital. If capital mobility between regions is absent, the incentive for capital accumulation and accordingly growth, which is at the heart of agglomeration, diminishes. In other words, in the absence of migration this means that agents no longer have the incentive to accumulate capital and innovate, which leads to lower agglomeration (Baldwin & Martin, 2003:3-4; Brakman *et al*., 2009:436-437). Baldwin *et al.* (2001) show that their version of the model can explain four familiar stages in economic development, following the industrial revolution. These include: (i) industrialisation of the core, (ii) the growth take-off, (iii) worldwide income convergence, and (iv) rapid expansion

[47] Urban economists (Henderson, 1988; Fujita & Thisse, 1996), growth economists (Lucas, 1988; Romer, 1990) and development economists (Rosenstein-Rodan, 1943; Myrdal, 1957; Hirschman, 1958; Perroux, 1955; Williamson, 1988) have all emphasised the role of cities in economic growth and innovative technology (Martin & Ottaviano, 2001:947).

of trade. There is a strong similarity between models of endogenous growth and that of NEG (Krugman, 1991a) in the sense that they ask questions that are related. The first is how technological innovations impact on new economic activities, and the second asks why these new economic activities are spatially concentrated and how they choose to locate. Accordingly, the process of creation of new activities and the process of location should be considered to be intertwined (Baldwin & Martin, 2003:4; Baldwin & Forslid, 2000:308).

Agglomeration and growth models are characterised by capital as a factor of production and the assumption that capital mobility is interregional and that localised technology spill-overs (knowledge or information) will imply that the location of firms has an impact on the cost of innovation and consequently the growth rate (Baldwin & Martin, 2003:5; Brakman *et al*., 2009:437). The capital mobility assumption is important since it eliminates the possibility of a breakdown in agglomeration because production shifting does not lead to demand shifting as financial gains are retrieved. Therefore capital mobility has a stabilising function where labour mobility on the other hand has a destabilising effect. Furthermore, capital mobility stabilises both the core-periphery and symmetric equilibria since the initial allocation of capital between two regions is constant (Baldwin & Martin, 2003:5).

Baldwin and Forslid (2000) present a model where long-run growth and industrial location are endogenously incorporated – by Romerian (1990) endogenous growth – into Krugman's core-periphery model with intertemporal optimisation in order to explain output per capita. They focus on stability and indicates that growth is a strong agglomerating or centripetal force. However, they show that localised information spill-overs are a strong dispersion or centrifugal force. Accordingly, the Baldwin-Forslid model explains the interaction between economic integration through, for example, the decrease in transportation costs, location of economic activity and economic growth. The model takes into account that economic growth influences location and location influences economic growth (Baldwin & Martin, 2003; Baldwin & Forslid, 2000:307; Brakman *et al*., 2009:438). For instance, effectively all endogenous growth models depend on technical externalities like information spill-overs or production externalities, essentially of the Marshallian type. An empirical study by Eaton and Kortum (1996), for instance, shows that these externalities are associated with spatial distribution of industry and/or research and development activities. Consequently, the presumption would be that spatially concentrating industry and/or research and development activities must be beneficial to growth. In other words, knowledge or information spill-overs

lead to a greater distribution of economic activity which is beneficial to the formation of a strong core-periphery system (Baldwin & Forslid, 2000:307-308).

2.3.2 Economic development

2.3.2.1 Sectoral transformation

As mentioned in the first chapter, spatial transformation, in the form of growth of towns and cities due to urbanisation, is strongly associated with sectoral transformation of countries from an agrarian, to an industrial, to a service-oriented economy (World Bank, 2009:9). According to Henderson (2003:275) the change in sectoral composition away from agriculture to industry and also technology progresses in domestic agriculture, which releases labour from the agricultural sector to migrate to cities, leads to the occurrence of urbanisation. Authors such as Rosenstein-Rodan (1943), Lewis (1954), Perroux (1955), Myrdal (1957), Hirschman (1958), Rostow (1960) and Kuznets (1973) all agree that sectoral transformation or development is a precondition for economic growth. In a broad sense their concepts were mutually supportive as they formed a school of thought which emphasised a more practical and historical and less theoretical approach in terms of economic development. Like any field of analysis, differences were evident on the emphasis and interpretation between their respective works. However, they all shared an important preference for industrialisation as the stimulator of economic growth, which would in turn stimulate all other sectors of the economy. Another important characteristic of their development theories was that, in most cases, they advocated substantial short-term policy intervention by government in order to drive the economy. However, in the long-run they expected government to reduce their intervention to only fulfil a stabilising function since the economy would perform efficiently with competitive market interactions (Cypher & Dietz, 2009:140-141).

2.3.2.1.1 The Lewis dual-sector economy

Sir Arthur W. Lewis (1954) is one of the most important early pioneers of development economics. Lewis's article, *Economic Development with Unlimited Supplies of Labor,* is one of the best-known dual sector models in development economics (Fields, 2004:1; Henderson, 2003:277). Like his development economist counterparts he believed that economic development in less-developed nations could rapidly be promoted. Furthermore, he shared the opinion that industrialisation was essential for less-developed nations since it provided them with the means of escaping poverty and reaching a higher degree of social and

economic progress. Lewis states that developed nations' wages were increasing and consequently widening the disparity with less-developed countries' wages. He was of the conviction that this growing disparity was a direct result of the difference in productive structures prevailing between the two regions. He emphasised the fact that developed nations had large manufacturing and industrial sectors, where a large proportion of the population were employed and relatively small agriculture sectors, where a smaller proportion of the labour force was employed. This was the opposite in less-developed nations, where most of the labour force were situated in rural areas focusing on agricultural activities. He also mentioned that higher wages were paid to labourers in the manufacturing and industrial relative to the agricultural sector in both the developed and less-developed nations, although the disparity in income was narrower between the two in the developed nations. The reasons for this disparity according to Lewis (1954) were that the productivity per worker in the industrial as well as agriculture sectors was higher as a result of the use of intensive capital and advanced technologies. Accordingly the higher average income in the developed nations was a structural function of having more labourers in a productive, higher-income industrial sector relative to less-developed areas (Cypher & Dietz, 2009:151-152).

In reaction to these conditions Lewis (1954:140-141) specifically stated that he wanted to build a dualistic model which can be applied to some but not all economies. In the Lewis model, the essence of labour market dualism is that labourers earn varying wages depending on the sector of the economy wherein they are active (Fields, 2004:3). If the industrial or manufacturing sectors were willing to remunerate at a rate higher than the wage in the agricultural sector, to such an extent that it would cover the costs and discomforts of migrating to industrial regions and also the cost of urban living, then an incentive would be created for agricultural surplus labour to move to urban areas (Lewis, 1954:150; Cypher & Dietz, 2009:153). Consequently, a higher level of productivity is necessitated in the agriculture as well as industries sectors in order for a surplus labour to be created in the agriculture sector as well as a welfare incentive in the industrial sector. These conditions could only be created by technological progress in both sectors. The work of Lewis was subject to critical scrutiny since institutional factors which influence wage determination, such as minimum wage platforms and government labour standards, were absent in the model (Henderson, 2003:277). Furthermore, the influences of capital flight in the developing countries are ignored, although justification can be given for this on the grounds that in the 1950s this was not a significant phenomenon (Cypher & Dietz, 2009:153-158).

2.3.2.1.2 Rostow's stages of growth

Walter Rostow's (1960) work is also notable due to the differences from the early development economists. He focuses on a descriptive approach while stressing the unavoidability and predictability of economic development, since Europe's past development is an indication of what developing nations may experience during their development (Cypher & Dietz, 2009:141). His concept is outlined in *The Stages of Economic Growth* in which he argues that a society undergoes sequential economic steps in its development to modernisation (Mallick, 2005:5). The stage model provided a universal interpretation which he was looking for (Cypher & Dietz, 2009:159). Rostow (1960) recognises five growth stages: (i) traditional society, (ii) preconditions for take-off, (iii) the take-off, (iv) drive to maturity and (v) high mass consumption.

In the first stage, the traditional society, the economic system is pre-scientific and dominated by agrarian activities. Scientific progress happened occasionally, but no system is put in place to introduce scientific knowledge into the production process on a regular basis. Rostow states that traditional society can be found during the time of Newton. The second stage, preconditions for take-off, is characterised by the emergence of new types of entrepreneurs and managers in the public and private sectors, with banks emerging and the rates of investment increasing, especially in infrastructure. Furthermore, modern businesses will be created which will benefit from advanced methods of production. This will initiate a dynamic development. Labour in the primary sector will become less important as a result of new developments in agriculture. An important prerequisite for this phase is industrial revolution (Cypher & Dietz, 2009:159-160; Mallick, 2005:6-7). In other words, this stage can be seen as the incipient stage of urbanisation. Urbanisation starts to accelerate as a result of the rural-urban transformation spurred on by industrialisation and consequently the density of settlements starts to increase (World Bank, 2009:61; Naudé & Krugell, 2003:479). The industries start to cluster in urban areas in order to benefit from increasing returns (essentially information spill-overs and the existence of specialised suppliers) and they want to be in close proximity in order to minimise transportation costs in a time where it was still unconventional to move large amounts of inputs in production and final products.[48] Labour is attracted from rural-agrarian areas by increased wages offered by industries (British success in the international economy induced economic developments that gave remunerative

[48] Since the period does not represent a significant progress in the history of the economy and technology. Productivity growth only increased to some extent (Allen, 2006:1).

possibilities) and a labour surplus in agriculture (as a result of developments in agriculture) and accordingly a gradual process of urbanisation started (see Allen, 2006). The take-off into sustained growth marked the third stage of Rostow (1960). In this stage the dynamic economic growth is sustained (Mallick, 2005:7). It is also characterised by an increase in the productive investment rate (five to ten per cent of national income), emergence of manufacturing sectors with a high growth rate and the creation of a social, institutional and political structure which utilises the drive to expansion (Rostow, 1960:39). Furthermore, resources need to be allocated efficiently and be directed at the modernising non-industrial (foreign exchange earning sectors, social overhead capital, and agricultural earnings) grounded upon the enhanced use of natural resources. Additionally, modernised techniques must be applied to areas that may be conducive to growth, with a high rate of returns on investment (Rostow, 1960:193). Modern techniques in this sense refer to greater use of technology and innovation. In the short-run output per capita can grow by introducing technological progress or capital accumulation. However, capital accumulation is affected by the law of diminishing returns, thus it is only through technological progress that a positive growth in output per capita can be upheld in the long-run (Brakman *et al*., 2009:69). The fourth stage is the drive to maturity. It is a period in which economic growth is adequately high to increase the per capita income of a country. Continual investments take place and new forms of industries like neo-technical industries appear, for example, chemical and electrical industries as well as mechanical engineering. The economy becomes diversified and sophisticated technologically, to such an extent that the society can produce a wider range of products (Mallick, 2005:7; Cypher & Dietz, 2009:163). The impact of increasing returns (due to an even larger market) and the love-of-variety effect can be thought of in this stage. As mentioned before, consumers love a variety of goods and this causes them to cluster in urban areas (Boschma & Frenken, 2006:275; Fujita *et al*., 1999a; Ottaviano & Thisse, 2004:2576; Neary, 2001). Accordingly this period might show ever greater urban development due to increased market demand linkages. The last stage that Rostow (1960) described is characterised by a period where production is for the purpose of consumption. Society lives in prosperity and an abundance of consumer choices is provided (Cypher & Dietz, 2009:163; Mallick, 2005:7). An idea of a system of large well-connected urban structures where rural and urban areas are integrated and has narrow income-disparities is sketched by this stage. In other words, a network of supportive cities and towns is established within a country where

the level of urban primacy is reversed due to a high level of development (World Bank, 2009:231; Henderson, 2003:276).[49] According to Henderson (2003:277) by upper middle income ranges 60 to 90 per cent of a country's population is urbanised. The variation in percentage is due to physical geography, the role of agriculture and national descriptions of urban. However, data comparing different regions from 1960 to 1995 suggest that urbanisation increased in all regions over the world, except developed countries where stagnation in urbanisation and even decentralisation occurred. Accordingly this translates into the fact that at some stage the high level of development is extended to the rural periphery. This is in accordance with the bell-shaped curve of spatial development. Authors such as Williamson (1965) and Wheaton and Shishido (1981) argue that during the initial stages of economic growth a high level of urban concentration along with divergence of the urban-rural wage differential is expected to occur. However, as development proceeds, eventually spatial decentralisation and convergence in the wage differential should occur. In other words, the formation of a core-periphery structure should be followed by phases involving interregional convergence in living standards or development (Ottaviano & Thisse, 2004:49-50).

Rostow's work came under much criticism from all schools of thought in development economics. Empirical evidence showed that the historical development of Britain did not coincide with many developing countries' development. However, Rostow made a meaningful contribution since he made other economists review the development of now-developed nations and demonstrate the gap that had been left as a result of the less-developed world's stagnation, distorted development and economic decay. Furthermore, Rostow opened up a debate on the impact of colonialism on the less-developed world since it may have led to backward socio-economic forces and processes in the colonised countries (Cypher & Dietz, 2009:163-164). The most useful insights that can be taken out of Rostow's work for this dissertation, however, is that countries may not grow precisely as he had predicted, but sectoral transformation away from agriculture into industry along with technological advances in domestic agriculture release labour which in turn leads to urbanisation and growth of important cities and towns that accommodate the growth and development of countries (Henderson, 2003:275).

[49] Urban primacy refers to the dominance of one city over other smaller cities (Henderson, 2003:276).

Subsequently, the early work of development economists typically indicated the desirability of ongoing urbanisation (Henderson, 2003:279). External economies and imperfect competition were most of the time part of the analysis in an implicit form, although these ideas were not continually invoked in mainstream economics at that stage. The theories applied by development economists at that time were interesting not only because they wanted to explain the conditions needed for economic growth in developing countries, but they also focused on the geographical influence on economic development, both in developing and developed countries (Brakman *et al.*, 2009:73-75). Another important fact about the models is that they presume an exogenously assumed condition where productivity of labour in the urban areas exceeds that in the rural areas (Henderson, 2003:277). Empirical evidence through the work of Fay and Opal (1999) and Davis and Henderson (2003) suggests that urbanisation does in fact reflect sector composition and that government policies have an influence on urbanisation essentially through their impact on sector composition. Furthermore, these studies argue that urbanisation occurs in the early and middle stages of development (like Rostow). It is also possible that the path of migration from rural to urban areas can be altered by wage policies (like in the dual economy) or by migration restrictions (such as in planned economies like China) which will be focused on later in this chapter (Au & Henderson, 2006).

2.3.2.2 Human capital and welfare

The influence of the stock of human capital as a prerequisite for regional growth and competitiveness has been accentuated in a large number of studies (Ritsilä & Ovaskainen, 2001:317; Kleynhans & Drewes, 2008:141). Studies conducted by Lucas, (1988), Barro and Sala-i-Martin (1991), Camagni (1995), Glaeser, Scheinkman and Shleifer (1995), Davelaar and Nijkamp (1997), Simon (1998), Forslid (1999), Ritsilä (1999) and Glaesar (2004) all suffice as an example. Both academics and policy-makers have started to focus more on the significance of human capital as a source of economic development at both the regional and national level. The migration process, in particular its selectivity, can lead to human capital being redistributed regionally. A number of factors exist that contribute to the decision to move. These include economic, social and psychological reasons. In the end, positive migration decisions at the individual level translate into large population flows and considerable changes in the regional stock of human capital. The effect of educational attainment on migration is especially interesting (Ritsilä & Ovaskainen, 2001:317-318). Studies suggest that the likelihood of migration is higher if a person has a higher level of

education (see Molho, 1987; Owen & Green, 1992; Ritsilä, 2000). Furthermore, according to Bacolod, Blum and Strange (2010:245) a city's prosperity and growth rely critically on its ability to attract productive labourers, match them with jobs suited for them and also evolve their skills. The attainment of skills was already emphasised in the beginning of urban economics by authors such as Marshall (1890), Vernon (1960) and Jacobs (1969). Cognitive skills and people skills have a positive association with urbanisation and level of education. Strong evidence exists that highly skilled and capable workers' likelihood to migrate to large settlements is higher (Bacolod *et al*., 2010:247-248). Thus, an important variable for economic growth and agglomeration is the initial education level of the population. Many explanations for the relationship between growth and initial human capital levels across countries have emphasised on the productive externalities promoted by schooling. These externalities are particularly persistent in large cities, which support the view that schooling, through spill-overs, spurs on growth (Glaeser *et al*., 1995:140).

Theories on economic growth emphasise that human capital is a prerequisite for the processes of economic growth and development (for example Lucas, 1988; Romer, 1990; Krugman, 1991a). Lucas (1988) states that as a result of the benefits of agglomeration, human capital moves from where it is scarce to where it is abundant. Furthermore, he argued that economic growth relies on the potential to absorb information and create new information, which is directly related to the stock of human capital that exists. However, the cost of gathering information and creating new information is high and accordingly human capital clusters together to minimise the cost of acquisition, creation and dissemination of information. Cities with higher clusters of educated individuals generate more localised spill-overs. It has been shown that these externalities are positively correlated to the location of economic and research and development activities (Sbergami, 2002:1). As a consequence these cities become more productive and attract large portions of the population (Simon, 1998:223-224, Naudé & Krugell, 2003:479-480). This statement coincides with the rational decisions of individuals, who migrate to expanding regions with higher wages and modern jobs. Ritsilä and Ovaskainen's (2001) study is based upon a human capital framework, which in turn is based upon the models of Sjaastad (1962), Weiss (1971) and Schaeffer (1985). These models state that migration is supposed to be a result of differing individual levels of economic utility in various locations. The individual also is assumed to maximise this economic utility. Consequently, migration or relocation takes place if the economic utility from migration is greater than the economic utility of remaining in the same location. An important factor that

influences the economic utility of an individual, and hence his or her decision-making, is the human capital reserve of the specific person. Human capital is heterogeneous and can come in the form of educational attainment (schooling), experience and training. The result of rational decision-making, the positive migration decision, is reached when the utility gain from migrating is in excess of the direct and social costs of migrating (Ritsilä & Ovaskainen, 2001:319). The highly educated tend to urbanise to cities in an attempt to realise their investments in education effectively. The educated migrants raise the educational level in the agglomerated areas and thereby provide new ideas and entice investment in new technological innovation (see, for example, Nijkamp & Poot, 1997). This centralising development movement can induce a recurring "vicious circle" where concentration of human capital can lead to further concentration of human capital (Ritsilä & Ovaskainen, 2001:319). Accordingly, urbanisation can lead to regional concentration of human capital, which can lead to divergence of local economies rather than creating a converging effect in development (Ritsilä & Ovaskainen, 2001:319; Myrdal, 1957; Nijkamp & Poot, 1997). Subsequently, on the one hand, large cities will have higher levels of human capital, higher wages and higher rents per labourer and on the other hand rural areas may just experience the opposite (Naudé & Krugell, 2003:479-480). Two centuries of economic development indicate that spatial disparities in economic activity and income are unavoidable. The implication is that living standards in a country's development diverge before converging. Accordingly, it is imperative that countries institute good policies that stimulate the concentration of economic activity and at the same time aid in the convergence of living standards. The World Bank (2009) stresses in its world development report, *Reshaping Economic Geography*: "The challenge for governments is to allow – even encourage – "unbalanced" economic growth, and yet insure inclusive development" (World Bank, 2009:20). This can be done by narrowing the disparity between rural and urban areas in economic terms. However, what happens when government policies are not necessarily aimed at promoting development of certain rural areas due to political motives? What influence can these repressive agendas have on the natural rate of growth and urbanisation in a country?

2.3.3 Policy implications of a restriction on migration

As mentioned before, economic activities are unevenly distributed across space. The determinants of spatial variations in the patterns of economic activity have traditionally been given in terms of variations in technologies, endowments or policy regime (Ottaviano & Puga, 1998:707). According to Neary (2001) directions for future research in NEG include

empirics and policy. The latter has contemporarily received a lot of attention since NEG possesses the ability to determine the impact on spatial agglomeration (Ottaviano, 2003:665-666; Neary, 2001). As might be expected in this section, policy implications will be discussed regarding the restriction of the mobility of individual agents and how this might affect the spatial pattern of distribution of economic activity and the economic efficiency of regions or cities.

The predominant effect of interregional integration is likely to increase economic efficiency within geographical space. However, if place-based policies reduce stimulus to migrate, then they may reduce the advantageous process of migrating from less productive areas to more productive areas. Economic efficiency requires of labourers to migrate to more productive industries or regions. If economic policy tries to stop that process, efficiency is reduced. Urban economists agree that good urban policy must reflect the fact that people and firms are fully mobile (Glaeser, 2007:27-29). An example of a country that has strongly restricted rural-urban migration is China. Au and Henderson (2006) argues that the intra-sector restrictions have led to insufficient agglomeration of economic activity in both the rural industrial and urban sectors, which in turn has resulted in first order losses in GDP. China's policy instrument is the household registration, or hukou system.[50] Au and Henderson (2006) raise the question whether allowing greater migration to cities would stimulate economic growth by reallocating workers from low marginal product employment in rural areas to high marginal product employment in the urban areas. In addition, they raise a related policy concern regarding the national income inequality which is a major political and domestic social problem (Au & Henderson, 2006:350-351). According to Johnson (2002) since 1978 China's rural-urban income gap has increased and the country's Gini coefficient has substantially risen as a result of restrictions on mobility. Au and Henderson (2006) state that the reason for this is that in both urban and rural sectors unexploited economies of scale exist from spatial agglomeration of resources. Furthermore, they empirically show that due to the restrictions many cities in China are undersized with the unexploited economies of scale leading to productivity losses (Au & Henderson, 2006:350-351). Consequently, agglomeration that spurs on economic growth is lost and discrepancies in economic development occur.

[50] The hukou system is more or less the same as the internal passport system (Au & Henderson, 2006:352; see Chan, 1994 for more information on the system).

Locational agglomeration models contain a recurring argument that regional and urban development is characterised by "path dependence", whereby historical disturbances can alter locational occurrences, and can have long-run cumulative outcomes (Martin, 1999:70). The fundamental competitive equilibrium model implies a unique, optimal and unavoidable locational pattern of industry predestined by the original spatial distribution of resources and factor endowments and transport capabilities (Martin, 1999:69-70). On the contrary, Krugman (1991a, 1991b, 1991d) and Arthur (1994a, 1994b) contend that there is a factor of uncertainty in industrial location and agglomeration which means that various alternative equilibria are within the bounds of possibility. The specific equilibrium spatial pattern of activity that occurs depends on history. This means that the original pattern may be an "accident of history" and not an optimal distribution of economic activity. However, once initiated, forward and backward linkages or self-fulfilling forces mean that the original urban or regional pattern becomes "locked-in" through the cumulative causation process grounded upon increasing returns. This is known as the lock-in effect. The main idea is that temporary policies can have permanent effects on the equilibrium spatial pattern of footloose economic activity (Ottaviano, 2003:671; Brakman *et al*., 2009:463). Accordingly, "illogical" economic rationale can result in sub-optimal but equilibrium distributions (Martin, 1999:70). Furthermore, urban economics also addresses the question of the influence of large temporary shocks (in this case in the form of policy) on agglomeration and the growth of cities. Two views exist on this impact. Firstly, models in line with the Alonso-Muth-Mills tradition (which originated out the Von Thünen model) state that each city has an optimal size. Following a substantial shock it is expected by these models that each city will return to its optimal size. The central idea here is that each city has its own independent level of productivity, which is not influenced by the shock. The result is that a shock will not have a permanent influence. On the other hand, the other view by urban economists is that the productivity of a region or city is a function of the city-size population (or distribution), as well as possibly the interdependencies (positive externalities) with other connected cities. The changes in the level of population of the city will alter the city's productivity and consequently its growth. In this case, the central idea is that a substantial shock has a lasting impact on the city. Geographical economics is in accordance with this view (Brakman *et al*., 2009:230-231). Accordingly in this dissertation it is assumed that a temporary and exogenous shock in the form of a policy can result in moving the economy out of the original, pre-shock, equilibrium. The shock needs to be large in order to have an influence on all locations and also to have the potential to shift the economy from one equilibrium to the next; the shock

also needs to be temporary in order to isolate the impact of the shock; and the shock must be exogenous to ensure that causal links run from the shock to the location of economic activity (Brakman *et al*., 2009:230). An example of the impact of large temporary exogenous shocks is shown by Davis and Weinstein (2002) for the allied bombings of Japan during WW II. In addition, Brakman, Garretsen and Schramm (2004) follow the same approach as Davis and Weinstein (2002) in the case of the Allied bombing of German cities during the war (Brakman *et al*., 2009:230-232). A more policy-relevant example of a shock can be shown by Redding, Sturm and Wolf (2007) who look at impact of the break-up of Germany in 1949 into the GDR (East Germany) and FRG (West Germany) and the subsequent reunification of the two Germanies in 1990 following the fall of the Berlin. It is a large, 41-year, temporary exogenous shock. Redding *et al*. (2007) shows that the shift in location of the FRG's main airport hub from Berlin to Frankfurt can be used as a test for multiple equilibria. If multiple equilibria exist it indicates that permanent locational shifts can occur due to large, temporary, exogenous shocks. The airline industry is a prime test for the applicability of the geographical economics model in this shock since transport costs are naturally relevant in the industry, activity is footloose, and scale economies play a significant role. Before 1949, Berlin was the main hub and during the break-up Frankfurt took over the role from Berlin. Following the reunification Frankfurt remained the largest airport hub in Germany. The conclusion that can be made out of this evidence that multiple equilibria do in fact exist and this confirms that a large, exogenous, and temporary shock can lead to a shift in the optimal equilibrium level to a sub-optimal equilibrium (Brakman *et al*., 2009:241).

It is known that the shock needs to be sufficient in size in order to affect cities and their distribution patterns. Related to the question of this sufficient size is another important policy implication of geographical economics discussed by Ottaviano (2003): the "threshold effect". Policy mechanisms will only have an effect on the equilibrium spatial distribution of footloose economic activity when these mechanisms yield a specific serious mass. Consequently, the level of economic integration or an increase in the freeness of trade between regions can have a significant or no impact. The outcomes rely importantly on the original freeness of trade and on the disparity between the original level and the break or sustain point. As long as freeness is below the break point, the economic activity or firms remain dispersed due to the fact that it is the only long-run equilibrium that can occur. However, as trade is steadily liberalised the freeness of trade passes the break point which induces spatial differences that can become fairly rigid. The reason for this is circular

causality since agglomeration is self-enforcing as it produces rents that tend to keep firms and factors constant (Ottaviano, 2003:670; Brakman *et al*., 2009:464). As Fujita and Thisse (2002) state this effect can be traced back to the "putty clay" effect. Putty clay geography implies that policy interventions of a relatively limited size may have no impact on the spatial location of firms. However, when the policy interventions are substantial and exceed the threshold level, then the pattern of location may change. When this occurs, the forces change the initial condition rapidly, giving rise to immediate geographical shifts. Marginal policy changes are inefficient until the accumulated change remains below a certain threshold. Accordingly after the threshold has been surpassed the influence of policy change can be "catastrophic" (Ottaviano, 2003:670). The benefits of a certain region have a tendency to reinforce themselves, and the choice of location will only be altered if the policy impact is substantial enough to negate the advantages of agglomeration (the so-called agglomeration rent) (Brakman *et al*., 2009:464). Baldwin, Forslid, Martin, Ottaviano and Robert-Nicoud (2003) use an analogy for this process. They state that plate-tectonics form the earth's geography. Even though the primary force is stable, the effects of earthquakes and volcanic eruptions suddenly change the quiescent state and dramatically change the physical geography. Examples of threshold effects are shown in the cases of Baldwin and Krugman (2000) and Kind, Midelfart-Knarvik and Schjelderup (2000). These authors portray an original situation in which firms are agglomerated in an area. They emphasise that if another area wants to attract firms, that region has to provide an incentive through a subsidy that is larger than the agglomeration rents firms benefit from in their present location. Any variation that does not surpass this threshold does not have an influence on the location of firms (Ottaviano, 2003:670).

2.4 Conclusion

In this chapter it was found that the clustering of populations and firms around the world in large cities and metropolises can be seen as the rule and not the exception (Brakman *et al*., 2009:xxiii). Millions of people locate in densely populated cities and readily subject themselves to the unpleasantness of living there. On the contrary vast empty areas exist where few or no people are to be seen. The unequal distribution of population and economic activity is one of the most obvious phenomena of the global economic system (Van Marrewijk, 2012:294). A number of motives exist for people and firms to cluster; however, the economic rationale behind clustering, or agglomeration as it is technically known, can be seen as a prerequisite for all the other motives. The agglomeration of economic activity

occurs at various levels and this raises the questions of *who*, *why* and *where* of the location of economic activity. Economic geographers continue to be gripped by these questions at the international, national and sub-national levels.

One of the main goals of economic geography as a discipline is to explain agglomeration and the location of economic activity. Over the history of the field the focus has changed and different research movements have taken place, which in turn have led to different explanations for the spatial concentration of economic activity (Aoyama *et al*., 2011). The reason for these differing explanations can mainly be found in the fact that the field has multiple origins, enjoys diverse interest from other social sciences and uses various heterodox methodologies. As a consequence the long history of economic geography plays an important role in, firstly, shaping the discipline and, secondly, making the contemporary models and theories more understandable. The reconnection of traditional theories with new phenomena and the determination of the relationship between historical concepts with contemporary debates can be useful in showing that new challenges and questions have long intellectual backgrounds (Aoyama *et al*., 2011).

The long tradition of analysis descending from Von Thünen (1826) all have something valuable to say about the relationship between geography and economics, but each approach has had its limitations. The patterns of land use are explained quite well around the city or an economic centre (for example by Weber, Lösch, Christaller and Hartshorne). However, the existence of a central focus is simply assumed (Fujita & Krugman, 2004:141). The question of why the city was established there in the first place as well as the interaction between these settlements were considerations that were absent in these early models. It was necessary that these limitations be addressed.

The introduction of increasing returns as well as imperfect competition was necessitated in order to explain individual behaviour and decision-making in the models. Although Alfred Marshall articulated increasing returns in 1890, these were only incorporated fully into the discipline by the work of regional scientists like Hoover (1948), Isard (1949, 1953) and Berry and Garrison (1958b) during the quantitative revolution. It was with this revolution that economic geography for the first time was recognised as an actual spatial science. The models were, however, still of partial equilibrium nature and perfect competition was a problem in analysis. During this period, essentially the geometric pattern of the Central Place System was formalised; however, individual firms and consumers' behaviour as well as their

decisions and interactions was not entirely explained by these models. The central place outcome was simply rationalised (Brakman *et al*., 2009:46).

By the 1970s regional science and spatial analysis began to be rejected. Another revolution was initiated during this period in the field of industrial organisation. Theorists for the first time were trying to implement tractable models of competition that could be combined with increasing returns. The work of Dixit and Stiglitz was especially important since their model would revolutionise theoretical modelling in various fields such as geographical economics, industrial organisation, trade and growth theories (Brakman *et al*., 2009:93). Contemporarily the Dixit-Stiglitz model is the foundation of economic theory in economic geography, international trade and economic growth theory. Furthermore, during that period work in line with the ideas of Alonso (1964), Mills (1967) and Muth (1969) on transportation and accessibility dominated the field. Henderson (1974, 1977, 1988) followed these authors with his writings regarding a framework of a system of cities. His work to date remains the basis of research regarding the actual distribution of sizes and various forms of urban areas and external economies are specifically the focus (Tabuchi, 1998:335; Fujita *et al*., 1999b:19; Brakman *et al*., 2009:40).

Although the 1970s signified large developments in economic geography, an essential ingredient was still absent. Models had evolved and incorporated both increasing returns and imperfect competition; however, the application of these elements was still not in a general equilibrium framework. On the one hand the regional and urban economics lacked micro-economic foundations of individual behaviour within a general equilibrium structure. On the other hand the old and new trade and growth theories about the micro-economic foundation and general equilibrium exist, however it lacked the incorporation of geography or when it was incorporated it was not sufficiently related to individual behaviour of agents (Brakman *et al*., 2009:76). Fortunately, the most recent revolution in economic geography was the rise of the New Economic Geography led by the model of Paul Krugman (1991a). The central contribution of the NEG lay in creating a framework in which aspects of mainstream economics, in particular general equilibrium models and rational individual behaviour and decision making, are implemented to model the trade-off between forces of agglomeration (external economies) and forces of dispersion (diseconomies such as congestion).

Krugman's (1991a) core-periphery model has until now been described as the core model of geographical economics. The work of Dixit and Stiglitz is combined with Krugman's (1979,

1980) previous new trade theory by introduction of interregional factor mobility. When transport costs are low, workers and firms tend to cluster in profitable regions in order to reap the benefits of external economies rather than to spread out over many regions. Krugman's model made a lasting impact on the field of geographical economics since he was the first to implement increasing returns and monopolistic competition into a general equilibrium framework. Consequently, the main contribution of geographical economics is to provide a single analytical framework to combine the elements of existing location theory.

Krugman's work inspired other writers and stimulated the development of new theorising and empirical work in geographical economics. New variants of the core model are still emerging by including new growth and new trade theory. The models furthermore try to explain the development of city systems through increasing returns. To illustrate this Krugman's models show that the population cluster in cities because of higher wages, the love-of-variety effect, and the larger market for firms to sell their products within. These ideas are applied to obtain empirical regularities like the log-linear "rank-size" distribution of cities. Consequently, the explanation of the location of economic activity, urban systems and spatial economics can now benefit from geographical economics literature since it possesses the needed analytical, mathematical, econometric tools and "modelling tricks" to address the spatial distribution of economic activity.

In numerous instances the Krugman model and new geographical economic literature have been applied in order to explain the relationship between economic growth and agglomeration. The initial literature on agglomeration and economic growth has combined endogenous growth literature with geographical economics. All the elements of the core model of geographical economics are present; however, the focus has shifted towards capital mobility. This is important since growth occurs as a consequence of accumulation of capital. If capital mobility is absent between regions, the incentive for capital accumulation and consequently growth disappears. In other words, in the absence of migration, agents no longer have the incentive to accumulate capital and innovate, which results in lower levels of agglomeration. Accordingly agglomeration can be considered the territorial counterpart of economic growth (Fujita & Thisse, 1996). However, agglomeration and economic growth can also have a significant impact on the economic development of a country.

The growth of towns and cities due to urbanisation is strongly associated with the sectoral development of a country. The shift of the sectoral composition away from agriculture, to

industry and also technological processes in domestic agriculture releases labour in the agricultural sector to migrate to industries in cities. This fact is also stressed by Sir Arthur W. Lewis (1954) who is one of the most important development economists. He, like most of these economists, shared the view that industrialisation was essential for less-developed countries to escape poverty and reach higher levels of social and economic progress. The incentive created by higher productivity and accordingly higher wages in the urban areas, where industries are located, as well as a surplus labour in the agricultural sector, due to technological progress, would lead to urbanisation and consequently growth of cities and towns which would have a positive effect on economic growth. Another approach was followed by the development economist Walter Rostow (1960) who described a process where countries go through certain stages in their economic development. He also emphasised that sectoral transformation would occur in the development of a nation and the desirability of urbanisation in economic growth. Although these development economists' approaches may differ, they all come to the conclusion that sectoral transformation and the migration of workers to more productive areas in a country are essential prerequisites for economic growth and consequently development.

Human capital has also been shown to be of significant importance for regional growth and competitiveness. Academics and policy have both shown an increasing focus on the importance of human capital as a driver of economic development at both the regional and national levels. Studies suggest that educated people are more likely to migrate to regions with high levels of human capital, since they gain form proximity to each other. A city's prosperity and growth crucially rely on its ability to attract productive labourers with the necessary skills. Skills have a positive correlation with level of education as well as urbanisation. Educated people migrate to higher wage urban areas in order to realise their investments in education. The educated migrants raise the educational level in agglomerated areas and accordingly create an incentive for research in new technology. This leads to a recurring phenomenon where human capital moves to a further concentration of human capital. However, rural areas can be left behind in terms of human capital levels, wage levels, and rents per labourer. It is imperative for governments to institute policies that are aimed at the development of rural areas in order to narrow the discrepancy in living standards with urban areas. However, political motives may lead to a situation where the development of rural areas is ignored.

The policy implications regarding the restriction of the mobility of individual agents can affect the spatial pattern of distribution of economic activity, and consequently the geographical efficiency of a country. The incentive for labourers to migrate from less productive areas to more productive areas is taken away. Increasing returns in urban areas may be left unexploited and this may result in undersized cities. This may lead to insufficient agglomeration and it can have a direct impact on economic growth. Furthermore, historical shocks can alter locational occurrences and can have long-run cumulative outcomes. Temporary policies can have a permanent effect on footloose economic activity. Accordingly, economic rationales in the form of restrictive migration policies can result in a sub-optimal equilibrium distribution of cities. Geographical economics, and accordingly this dissertation, coincide with the view of urban economists that the productivity of a region is a function of the city-size population as well as the interdependencies with other connected cities and towns. An alteration of the level of population of a city can alter a city's productivity and consequently its growth.

Chapter 3: The history of the concentration and movement of people and economic activity in South Africa

3.1 Introduction

It is important to give a historical analysis of the concentration and movement of people and economic activity in South Africa. A number of reasons can be given for this. Firstly, agglomeration is a key driver of urbanisation and industrialisation. Urbanisation takes place as a country moves its sectoral composition away from agriculture into clustered industries and furthermore technological innovation in domestic agriculture releases the labour from agriculture to migrate to large settlements (Henderson, 2003:275). Secondly, the settlement pattern of the African population varied significantly from the other racial groups of the nation. Africans to a large extent did not experience the same rate of urbanisation as the latter groups and the reasons should be looked for in historical aspects of the country (Gelderblom & Kok, 1994:67). Lastly, and this coincides with the second reason, history plays a key role in the certain equilibrium spatial pattern of economic activity. Accordingly, illogical economic policy can result in sub-optimal but equilibrium distributions (Martin, 1999:70).[51] South Africa's experience of the movement and concentration of people and economic activity has been unusual in a number of ways. Until the 1990s government intervention was substantial in the process. Initially interventions were aimed at accelerating a certain form of temporary urbanisation, and subsequently it was aimed at the restriction of the movement of people to cities. A large number of policy mechanisms, institutions and laws were used to impact on the mobility of households which included racially discriminatory government measures on people's ability to choose where to settle, to own land, regulation of their employment and education and training system (Turok, 2012:4).

This chapter will consequently look at the factors that have influenced the contemporary distribution of economic activity in the country. Historical development will accordingly be delved into. The first nature aspects will firstly be discussed followed by a chronological order to be further followed regarding development or in many cases under-development of economic activity in South Africa.

[51] See section 2.3.3 in Chapter 2.

3.2 Development of South Africa

3.2.1 The impact of first nature aspects and the start of urbanisation and industrialisation

According to Ottaviano and Thisse (2004:2564) the uneven distribution of people and economic activity has two possible explanations at a high level of abstraction. Firstly, spatial imbalances can be regarded as a consequence of the uneven existence of natural resources across geographical space. This is known as "first nature" and refers to characteristics that are naturally given by factors such as the climate and raw materials to a specific region. However, first nature does not entirely explain the clustering of economic activities and people when physical geography does not matter significantly. This leads to the second explanation known as "second nature" which refers to the economic forces that drive centripetal motion of people and economic activity. Second nature is a result of people's actions to enhance first nature advantages. Ellison and Glaeser (1997, 1999) express the same view and provide two reasons for agglomeration. First, they contend that the spatial concentration of United States of America's (USA) industries occurs due to the natural cost advantages as a consequence of differences in physical geography. Ellison and Glaeser (1999:315) concluded that more or less 20 per cent of the noticed clustering of industries could be explained by variables representing natural advantages. Second, the existence of increasing returns and other external economies of scale induces agglomeration.[52] They call this the role of economics in geography. Similarly studies conducted by Diamond (1997); Haaland, Kind, Midelfart Knarvik and Torstensson (1999); Acemoglu, Johnson and Robinson (2001); Black and Henderson (2003); Sachs (2003); and Nunn and Puga (2007) all emphasise the relevance of physical geography as well as concentration effects examined in geographical economics.

The point that is stressed here is that physical geography and also economic rationale play a role in the location of population and economic activity. According to Fujita and Mori (1996) the reason why ports and transportation nodes regularly become the sites for large cities is that they serve as linkage in the market potential function of a location and consequently creates an incentive for city growth. The assumption that can be derived is that, in many cases, the first nature aspects of the location of economic activity may play an even greater role in the initial distribution of activity than second nature considerations. Also the aspects

[52] Geographical economics models revolve around the second reason (Brakman *et al*., 2009:72)

of first and second nature are not opposing forces. Gallup, Sachs and Mellinger (1998:132) recognise the fact that they can work in tandem since the cost advantages of the initial differentiated geography create the city but the agglomeration economies allow the city to expand even after the cost advantages have diminished. Fujita and Krugman (2004:147) state that natural characteristics like rivers and ports matter. The natural advantages of a location, such as the availability of a harbour of river, generally have an "inaugural" or "catalytic" role. This role entails that the possibility of an emerging new centre to establish itself in a certain location rather than another location in the area is heightened. However, when the new settlement becomes established, self-reinforcing processes such as increasing returns lead to the growth of the settlement, and this continues until the initial advantages of the location become unimportant in relation to the self-reinforcing advantages of agglomeration.

South Africa demonstrates a similar pattern of settlement since first nature geography played a significant role and can be considered the root of uneven distribution of economic activity and population. The cities of Cape Town and Durban first emerged in the 1600s and 1700s as trading posts on the trade route between Asia and Western Europe (Naudé & Krugell, 2005:88). Later on a number of smaller settlements arose due to the change from cultivation to stock farming; however, urban settlement was predominantly restricted to the four harbour towns of Cape Town, Durban, Port Elizabeth and East London.[53] The majority of the inhabitants of South Africa were engaged in subsistence agriculture, except those farmers who produced wool and wine for export purposes. Until 1870 the country was relatively unimportant in terms of the world economy (Gelderblom & Kok, 1994:68). However, this changed with the discovery of diamonds in 1867 and gold in 1884 in the interior of the country (Naudé & Krugell, 2005:88). This discovery led to, as Gelderblom and Kok (1994) states, an "explosion of economic activity" in the new urban centres of the Witwatersrand (Johannesburg and Pretoria)[54] and Kimberley. This "Mineral Revolution"[55] led to fast industrialisation and extensive national and global migration, especially to the Witwatersrand (contemporarily known as Gauteng).[56] The explosion of economic activity necessitated the creation of a substantial workforce. The mining output grew rapidly; however, the demand for labour that existed held it back and could not be instantly satisfied (Wilson, 1972; Gelderblom & Kok, 1994; Turok, 2012). It is, however, not possible to determine one single

53 East London was the only river port at that stage (Floyd, 1960).
54 See Naudé and Krugell (2005).
55 See Yudelman (1984).
56 Johannesburg in the Gauteng Province is to date the largest city in South Africa (Turok, 2012:4).

reason why employers could not employ enough labour from the rural areas to satisfy the demand. A number of factors influenced the problem of lack of labour migration including the internal structure of African societies, bad working conditions as well as low wages (Gelderblom & Kok, 1994:69-71; Turok, 2012:5).[57]

Indeed, the mining explosion induced rapid industrialisation and urbanisation in a number of ways. As soon as the surface deposits of minerals had been extracted by small diggers, the excavation of deep underground deposits required machinery, especially steam engines; a substantial labour force; and also the provision of credit in society. The mining activities also stimulated the development of support industries and services. For example, chemicals (for explosives in mining activities), civil and mechanical engineering industries and services were required (Turok, 2012:5). Furthermore, the harbour towns evolved from being service and stop-over centres, to being ports which handle valuable commodities and railways were provided for inland mining activities (Naudé & Krugell, 2005:88). This period is regarded by many as the move from Rostow's (1960) traditional society to preconditions for take-off (see section 2.3.2.1.2; Fair, 1982:48-51; Houghton, 1973:7-18; Haasbroek, 1985). According to Gelderblom and Kok (1994:68) the economic explosion initiated a period of political, social and economic transformation. Consequently, South Africa quickly became a modern capitalist society. The increase in economic activity resulted in an increased demand for labour, which in turn resulted in large-scale urbanisation in South Africa. Accordingly, the first nature aspects of the country influenced the modern day agglomeration of economic activity in the six "large"[58] cities, namely Johannesburg, Pretoria, Ekurhuleni (East Rand), Cape Town, Durban and Port Elizabeth (Naudé & Krugell, 2005:87-88).[59]

However, the essence of urbanisation and also the capitalist development in the country was characterised by two features of society which were present from the early stages. First, the rich among the white population held a monopoly over political power and secondly, land ownership was organised on a racial basis. The consequence of the latter pattern was that the rural areas were separated into two parts. On the one hand, there were the areas

[57] See Gelderblom and Kok (1994) for an extensive explanation of these factors.

[58] The World Bank (2000:128) characterises cities according to their population. Less than 0.5 million people constitutes a small city; large cities are between one and five million; and megacities exceeds five million people.

[59] The other two mentioned cities (East London and Kimberley) are contemporarily also substantial in size; however, the centripetal or second nature forces of the other major cities (e.g. increasing returns and external economies) may have been greater than these forces in these cities later on in development.

predominantly owned by white farmers; and on the other hand the reserves owned by the Africans. The majority of the African labour forces of the new towns were supplied by the reserves (Gelderblom & Kok, 1994:68). However, early on most African labour was provided by African states on a temporary basis, which meant that a recurring migration pattern was established where migrant labour would travel to the mines for a summer to work, before returning home again (Wilson, 1972). The reason for this was that the development of towns also led to an increase in demand for agricultural products to sustain the labour force. Accordingly some authors, like Bundy (1979) and Murray (1976), contend that African farmers inside the reserve preferred production on the farms for the market over the unpleasantness of working in the mines due to the low wages, as well as the precarious and appalling circumstances. This may also be an explanation for the initial shortage in labour supply experienced by the mines (Gelderblom & Kok, 1994:70). On the other hand an incentive for migrant workers (from Botswana, Mozambique, Zimbabwe, Malawi and Zambia) to work in the mines was created by mining companies and the colonial government due to the fact that a limited labour supply was offered within South Africa. They invested in a number of schemes to secure labourers for longer periods and from further away. Corporate agents offered these migrants fixed contracts and arranged wages to encourage their movement to the mines (Turok, 2012:5).

Bundy (1979) furthermore contended that the South African supply of black labour only increased after the collapse of agricultural production. He argued that the central reason behind the collapse was due to interventions by the state at that time. White farmers were afraid of the competition posed by African farmers and mine owners demanded more African labour. Whites had political power and consequently the state acted against African farmers. Although recent historical research indicates the limitations of Bundy's (1979) arguments, the fact remains that it cannot be disputed that in the past government intervention had its role in the under-development of African agriculture.[60] Actions and institutions were established that put African farmers at a disadvantage to white farmers. For instance, agricultural marketing boards discriminated against the products of African farmers (Van der Horst, 1971:311). Furthermore, credit provision was available to white farmers but not to African farmers and discrimination was evident in terms of education, infrastructure and also technical advice (Gelderblom & Kok, 1994:70-71). Consequently, South African black migration was forced

[60] Examples of this included the Land Act 27 of 1913 and Land Act 18 of 1936, which made squatter farming illegal (Gelderblom & Kok, 1994:71).

to a certain extent into unpleasant, dangerous and low-wage mining activities. According to Yudelman (1984) by the beginning of the 20th century there were an estimated 100 000 African workers living in mining compounds which were introduced by the government to keep the workers on site.[61] By 1910 the number of Africans in these compounds exceeded the total white population of the Witwatersrand (Turok, 2012:5). The compounds were one of the first examples of the local authorities making provision for separate residences for Africans. However, no general law existed in South African legislature at that time regarding the right of ownership and occupation of property in urban areas.[62] However this all changed when the Lagden Commission proposed the ideological vision of segregation of Africans. The commission prescribed a separate residential location for the occupation of Africans who would only be entitled to move temporarily into towns if they were employed there. Specific legislation started to ratify this ideological vision in the forms of The Native Reserve Locations Act in the Cape (1902), the Native Locations Act in Natal (1904) and also the Orange Free State Municipal Ordinance of 1903 (Davenport, 1991:1). This was the beginning of a long line of discrimination against the occupation of urban areas by Africans in South Africa instigated by Lord Milner's regime who feared that African urbanisation would lead to disorderliness in a time characterised by industrial growth and fearfulness of health dangers (Baines, 1990:76-77).

The end of the nineteenth century marked a period when white agriculture also went through a crisis. Eventually it resulted in thousands of Afrikaners migrating to urban areas (Gelderblom & Kok, 1994:74). According to Wilson (1971) five factors influenced this movement. First, no more land was available for growth of agriculture in South Africa due to the rapid population growth of the Afrikaners. Second, the Roman-Dutch legal system allowed farmers to sub-divide their land amongst their children. This sometimes led to the inefficient use of land. Third, rinderpest epidemics affected the cattle in a disastrous way. Fourth, the Anglo-Boer War resulted in the "scorched earth" policy. The policy entailed that British forces burnt farms. Lastly, the effects of the drought of 1896 had devastating consequences for agricultural produce. Booysen (1985:16) furthermore points out that medical and educational facilities as well as the promise of a fixed income played a role in urbanisation.

---

61 Later on these developments would be significant in South Africa (Turok, 2012).

62 Excluding the Orange Free State who implemented such a law in 1893 (Davenport, 1991:1).

Apart from the White and African urbanisation the Coloured and Indian populations in cities also increased. The connection of the Witwatersrand by means of railway lines to Maputo, Durban and Cape Town also stimulated large urbanisation of these groups. Other towns were also connected as well as to the countryside and ports like Cape Town and Durban were expanded to cope with rapid immigration and commercial activity. The British brought a large number of indentured Indians to Durban in 1860 in order to work in sugar plantations. The Indians, however, quickly found opportunities in the trading sector and established themselves in urban areas, especially Durban. The urbanisation of Coloureds on the other hand was brought about by John Phillip who fought for the rights of slaves in the Cape that eventually led to the emancipation of a substantial portion of the group in 1828 and 1834. The released Coloured non-landowners clustered in urban areas and mission stations to attain an income (Booysen, 1985:18; Turok, 2012:5-6; Fair & Browett, 1979:268). However, what was evident from these early stages in South Africa's development was that all groups were in fact part of the urbanisation process (Booysen, 1985:15).

3.2.2 The establishment of the Union of South Africa

The British colonial government wanted to command the entire region in order to eliminate the possibility of competition with their mines and enhance industrial expansion. They wanted to enforce their laws and regulations and this created tension within the Afrikaner bureaucracy. This resulted in the Anglo-Boer War in 1899 to 1902 between the British colonies and the Afrikaner states. The consequence of the war was that South Africa for the first time was united into a single state under British authority. The Union of South Africa was established in 1910 (Turok, 2012:6). Thus the colonial government acquired what they wanted. Browett and Fair (1974) considered the establishment of the Union as the beginning of the phase of industrialisation. This was not in line with the view of Houghton (1973) and Haasbroek (1985); however, what is important is that the discovery of diamonds and gold brought about rapid sectoral transformation of the economy and it was also the beginning of actual urbanisation in South Africa. According to Booysen (1985:16) 52 per cent of the white population in South Africa was urbanised in 1911. However, he argues that it needs to be noted that most of these whites were migrants from Europe and non-Afrikaners. However, according to Bos (1992:222-223) from 1911 onwards this started to change. From 1911 to 1921 the annual growth rate of the white urban population was 3.01 per cent, whereas the annual growth rate of the rural white population was only 0.15 per cent over the same period.

The demographic urbanisation was strongly associated with an increase in economic activities in cities (Haasbroek, 1985:12). The first industrial census that was taken just after unification indicated that most of the industrial activities were concentrated in Cape Town, Port Elizabeth, Johannesburg and Durban. Four years after unification the First World War (WW) ensued and made the imports of goods difficult. However, this resulted in an increase in demand for locally manufactured goods, especially from gold-mining activities, which attracted industrialists (Bos, 1992:222). Furthermore, the mines required cheap unskilled labour to support the skilled British migrant workers. In other words this period was characterised by increased output in the country (Haasbroek, 1985:12).

The impoverished and unskilled Afrikaners who came from the rural areas were unwilling to do the same work as Africans (Haasbroek, 1985:12). However, they were both in competition for unskilled job opportunities (Van Jaarsveld, 1982:177). Like the African migrants the Afrikaners were extremely poor[63] and they entered into a foreign cultural environment directed by English-speaking people. Furthermore, another similarity was evident between the Africans and Afrikaners. Both groups strongly resisted proletarianisation, a process in which independent farmers lost their autonomy for a dependent position as worker for mining-employers (Gelderblom & Kok, 1994:75-76).

The reaction of the government to the urbanisation and poverty of Afrikaners was, however, different to the reaction of urbanisation and poverty to the Africans. Afrikaners were able to use their political capacity to ensure that their interests were more valuable than their African counterpart's interest (Gelderblom & Kok, 1994:76). The white workers felt threatened by the cheap black labour and consequently demanded protection by means of the channels of party politics, labour unions and strikes (Haasbroek, 1985:12). Initially, as Yudelman (1984) indicates, the government and capital (that is mining and industrial organisations) was in a "symbiotic relationship" which led to the forceful suppression of the strikes and labour unrest in 1922. However, after the Pact Government of Hertzog (a coalition of the Labour Party and the National Party) came to power in 1924, the state reacted by taking steps to resolve the poor white problem (Gelderblom & Kok, 1994:76-77). The government and capital, in their symbiotic relationship, did not react by reimbursing the white workers who were part of the strikes in 1922; the mines increased their profitability with the lower wages and the increased government income through taxation of the mines was used to enhance industrialisation.

[63] According to Van Jaarsveld (1982:168) the majority of the poor white population were Afrikaners. This is indicated by an estimate in 1921 that Afrikaners constituted 90 per cent of the poor whites.

They rather opted to discriminate against the African workers by not employing them in industries, although an increasing demand for labour existed in the industries as well as in mining, by giving white workers job reservations in the industries through the so-called civilised labour policy. This policy entailed that the white workers would receive favourable employment opportunities in industries, which in turn meant that African "uncivilised" labour would be forced to work in the precarious mines which offered extremely low wages (Gelderblom & Kok, 1994:76-77; Haasbroek, 1985:12-13).[64] Furthermore, the government neglected the improvement of infrastructure in African neighbourhoods (Haasbroek, 1985:13). It would not end there though. African industrial workers were regarded as guests by the local authority and required passes to move into urban areas on a temporary basis.[65] This aspiration of government of "influx control" was summarised by the Stallard Commission of 1921 which stated that the native should only be permitted to move into urban areas, which was fundamentally the "white man's creation", when he was inclined to enter and attend to the needs of the white and should withdraw himself from the urban area when he ceased to do so (quoted in Wilson, 1972:3). The ideal of segregation had by this stage evolved into a reality and the Union's Department of Native Affairs took it for granted. As a result the Natives (Urban Areas) Act 21 of 1923 was promulgated (Davenport, 1991:2-7). Although this Act, according to Davenport (1991:7), was not a harsh measure in comparison with the standards of its later amendments, it still had significant spatial implications.[66] The Act was the catalyst for the increase in restrictive controls regarding the movement and control of Africans in South Africa (Wessels & Wentzel, 1989:24).[67]

Consequently, the state had made a discriminatory decision not to employ and allow the movement of black Africans although it would have led to faster urbanisation and

---

64 Although the Afrikaner's political power played a major role this was not the only manner the government solved the Afrikaner unemployment problem. The public works programmes developed the skills of the Afrikaners which later on had a significant influence in the development of commerce and industry (see Abedian & Standish, 1985). Furthermore, organisations such as the Reddingsdaadbond and the emergence of Afrikaner capitalist concerns (SANLAM, Federale Volksbeleggings) contributed to the employment and upliftment of the group (Gelderblom & Kok, 1994:77-78).

65 The "pass" system was invented to discourage labourers from failing to comply with their employment contracts. Passes date back to the 1760s when slaves were required to carry passes in order to move between rural and urban areas (Turok, 2012:6).

66 The Act empowered local authorities to allocate land for Africans in certain locations, or to house them, or required their employers to do so. Furthermore, it prevented whites from acquiring land in these locations. However, they still were not entitled to own land (Davenport, 1991:7; Booysen, 1985:21).

67 For example, the Colour Bar Act restricted Africans from practising skilled trades. This protected the skilled White workers from wage competition (Turok, 2012:7).

consequently industrialisation.[68] Instead, large numbers of unemployed Afrikaners were employed by the government and flocked towards urban areas (Gelderblom & Kok, 1994:77). The result was that between 1924 and 1933 the number of white unskilled workers on the railways increased from 9.5% cent to 39%. Contrarily, the number of African unskilled workers decreased over the same period from 75% to 49% (Giliomee, 1979:151). Furthermore, 74.7% of the white population were urbanised by 1946 (Bos, 1992:223), whereas the percentage urbanised Africans at the same time was 21.6% (Hattingh & Olivier, 1985:30).

Gold mining remained central to the South African economy for decades and was a major source of export revenue (Turok, 2012:5). In the 1930s the mines were estimated to contribute 50 per cent of the government taxation revenues and around 50 per cent of the country's people were directly or indirectly dependent on the mines for their livelihoods (Yudelman, 1984). Further expansion of gold mining and manufacturing in urban areas during the 1930s and 1940s occurred and this made urbanisation and industrialisation even more desirable. Industries required more labour power and the rural population needed the job opportunities and incomes that were available (Turok, 2012:8). In other words the centripetal forces of agglomeration were getting stronger. Although the Urban Areas Act of 1923 prevented Africans from owning land in white urban areas and restricted their movement through influx control, some local municipalities did not always endorse the state's plan of segregation, township creation and slum clearance. Some municipalities were more moderate and others were more extreme in their application of this plan and consequently this resulted in uneven outcomes in various places (Baines, 1995). In effect the government policies were not always effectively translated into practice (Booysen, 1985; Parnell & Mabin, 1995; Gelderblom & Kok, 1994). Although discrimination was still evident, this resulted in the annual growth of 6.3% of African urbanisation between 1921 and 1936 (Wessels & Wentzel, 1989:27).[69]

Many municipalities started to provide formal housing in the townships for African families in acknowledgement of the need of industries for a more stable workforce and the dreadful circumstances in the inner-city slums and mining compounds. Subsequently, the amendment of the Urban Areas Act that was implemented in 1937, although impacted on by the official

[68] Although the government and capital knew it would be advantageous in economic growth and development.

[69] The annual growth rate for Whites in the same period was only 3.6% (Wessels & Wentzel, 1989:27).

rhetoric which identified urban Africans as temporary visitors, officially accepted that Africans who had been born in urban areas would not have to withdraw from these areas if they were unemployed (Wessels & Wentzel, 1989:40). Consequently, the period before and during WW II marked a relative relaxation of segregation policy in South Africa (see Davenport, 1991:12-15).

The outbreak of WW II, according to Kotzenberg (1973:139), was the central stimulus of industrialisation which led to an increased demand for manpower, commodities of all types and sea transport to serve the population as well as to supply war equipment.[70] Whatever argument existed for the stricter implementation of segregation, by the start of the War these policies no longer contributed to the needs of the economy (Davenport, 1991:12). The need for urbanisation of the Africans outweighed the political pressure to contain it until the end of WW II (Turok, 2012:8). This was significant for economic growth in South Africa. According to Houghton (1973:16-17) the period between 1933 and 1945 marked the take-off stage in the country's economy and the fact that more Africans entered into the modern sector was an important reason for this occurrence. Other reasons for the initiation of this stage included the depreciation of the rand which stimulated growth in mining activities; the introduction of Yskor; the inflow of foreign investments into the country; the accumulation of capital increased rapidly; and government intervention and industrialisation (also see Davenport, 1991:12).

The rapid urbanisation of Africans before and during the war also had significant socio-economic consequences. The living standards were considerably higher in the urbanised areas; however, a lot of the African workers succumbed to competition in the labour market. This gave rise to unemployment and in many cases the new urbanites did not know how to handle their financial affairs. In addition the housing, health and educational standards were appalling and family disintegration occurred along with increased crime. The white population also did not take the increased competition for employment and the composition of neighbourhoods out of different races kindly. Furthermore, African workers demanded higher compensation and African labour unions were formed, although they were illegal according to the Industrial Conciliation Act of 1924 (Wessels & Wentzel, 1989:54-55).

The Smuts government was aware of the problems the African urbanisation posed and even the preoccupation with the war did not distract them from giving attention to post-war reform

[70] See Chapter 4 of Wessels and Wentzel's (1989) report for further discussion.

(Davenport, 1991:12; Wessels & Wentzel, 1989:57). Accordingly, the Smit Committee was appointed in 1941 to address the socio-economic issues by means of investigating alternative approaches, excluding only the use of wage increases. The committee suggested feasible solutions to the problem and acquired valuable information regarding the standards of living of Africans.[71] The average wage of the group per month was not even enough to provide the basic living essentials such as food, clothes and housing.[72] Malnutrition, famine and sickness in conjunction with a shortage of housing were causes for great concern. However, the government was reluctant to implement actions that were suggested by the committee. As a matter of fact they actually reverted to existing legislation (Wessels & Wentzel, 1989:57-63).[73]

In 1945 the Native Urban Areas Consolidation Act no. 25 was ratified which contained the entire 1923 Act and all ensuing amendments of it. The Act was based upon the improvement of the living conditions of urban Africans, more efficient administration of urban areas and application of the 72 hour-clause and qualifying Section 10 rights (Booysen, 1985:21). According to Section 10 no black South African could stay longer than 72 hours in a designated area, unless the person had lived there for more than 10 or 15 years from birth depending on the number of employers that person had, was given permission by a labour officer or was a migrant worker. The Act was, however, ineffective in achieving its goal (Wessels & Wentzel, 1989:64).

Confrontations over the pass laws started to occur and tension increased between the government and African urban communities. Tension was also spurred on by the housing crisis of 1944 to 1947 and ongoing strikes by Africans in order to get wage increases (Davenport, 1991:13-14; Wessels & Wentzel, 1989:65). The mining strikes of 12 to 16 August 1946 where more or less 60 000 to 70 000 Africans participated probably had the greatest influence on White voters since it may have created the idea that the government was incompetent in addressing the problems in the labour market. It was evident that African urbanisation needed to be investigated on the national level and this led to the appointment of the Fagan-commission. The goal of the commission was to examine and give a report on the state of African urbanisation wherein advice and solutions are given for the problem (Wessels

[71] The suggestions included the subsidisation of households and the acknowledgment of labour unions (Wessels & Wentzel, 1989).

[72] War-time shortages made the matter worse – it pushed the price of food and transportation (Posel, 1991:20).

[73] Later commissions with the same goal were the Native Affairs commission and Broome-commission (Wessels & Wentzel, 1989:66).

& Wentzel, 1989:65-66). Fagan proposed that a substitute for influx control in the form of a Union-wide network of labour bureaux should be provided, which was also suggested by the Smit Committee. The suggestion was also made that African labour could be stabilised by encouraging temporary workers in urban areas to bring their families with them.[74] Fagan thought that the urbanisation of migrant labour could not be forced out of existence by legislation, or terminated by administrative mandates, or forced on employers. However, he argued that it would be advantageous to experiment with the idea in the new Orange Free State (OFS) goldfields. Accordingly he suggested that the power of the pass laws should be reduced in order to encourage contact between an urban worker and his rural origins. He published the report on the eve of the general election of 1948. The whites wanted to keep the ruling for generations and Africans were regarded as a threat to this ideal (Turok, 2012:8).

3.2.3 Apartheid and decentralisation

Segregation policies intensified when the Afrikaners' National Party (NP) came to power and apartheid was adopted as official government policy in 1948 (Turok, 2012:8). With the coming to power four basic principles regarding the urbanised Africans were enshrined in legislation according to Booysen (1985:22): Firstly, influx control would be implemented more effectively and severely; secondly, slums and mining compounds would be cleared; thirdly, races would be separated by neighbourhoods and groups would be allocated to certain areas; and lastly the provision of services and housing would be the obligation of the employer. The NP's argument was that South Africa could not be considered a single nation, but comprised four racial groups, namely Africans, Indians, whites and coloureds. Each race would be compelled to live in different locations according to their race. This translated into different regions of South Africa as well as different neighbourhoods within urban areas (Turok, 2012:8).

3.2.3.1 1950s and 1960s

As noted in the Section 3.2.2 the number of urbanised Africans increased rapidly during the 1930s and 1940s which was spurred on by opportunities in these urban areas as well as rural poverty. Between 1936 and 1946 the growth of the urban African population was 57.2 per cent. Contrarily, the white urban population had grown by 31.5 per cent. Accordingly one of the first priorities of the new regime was to find a solution for the "urban native problem" (Posel, 1991:19-20). The first of these "solutions" was the implementation of the Population

[74] The Native Representatives had the same idea (Davenport, 1991:15).

Registration Act of 1950. The Act formalised the racial classification and prescribed an identity card for everyone over the age of 18. The second fundamental aspect of Apartheid was the Group Areas Act of 1950. The act prescribed the racial composition for all the residential areas, and later on played a significant role in the removal of people who were not entitled to live in an area (Turok, 2012:8). Furthermore, the establishment of the Bantu Self-Government Act of the 1950s assigned Africans to their respective so-called "Bantustans" or homelands according to their ethnical groupings. The government created these homelands to function as independent states and was aimed at keeping whites and non-whites in separate locations (Choe & Chrite, 2014:83). However, urban resettlement policies were not initially (that is the 1940s and 50s) part of the Apartheid strategy (Posel, 1991:21).

Influx control, however, remained essential since it was a means of regulating the movement of the African population. The Pass Laws Act of 1952 intensified the restriction of the movement of Africans by necessitating them to carry a passbook if they were over the age of 16 (Choe & Chrite, 2014:83). According to the NP it formed a large part of the solution since it allowed them to restrict the growth of the urban African population and also eradicate urban unemployment. Consequently, the urban labour market would be restructured according of the 'urban labour preference principle'. The main idea behind this was that the urban employers' contact with the rural workers, or "tribalised" Africans, would be restricted and as a consequence the employer would be forced to use the alternative of a smaller urbanised, or "tribalised", labour pool. In turn this would result in labour turnover being cut and reduction of unemployment in the urban areas. The demand for rural workers in the urban areas would diminish which resolved the problem of urbanisation of Africans on the one hand and the shortage of workers on rural farms would be taken care of since deflected workers would seek job opportunities in the farmlands on the other (Posel, 1991:21).

In other words the urban labour preference principle was a strategy based upon the ideological and administrative permanent separation of the "detribalised" urban African population and the "tribalised" rural dwellers. However, the policy contrasted with economic realities since individual behaviour and preferences were not taken into account. This entailed that some employers had to employ urban individuals who did not have the needed job requirements, for example in the form of skills and training (Posel, 1991:22-24). As noted in Chapter 2 a city's growth and prosperity depend on its ability to attract and engage productive labourers in jobs suited for them (Bacolod *et al*., 2010:245). The importance of the attainment of skills is also stressed by urban economists (Marshall, 1890; Vernon, 1960;

and Jacobs, 1969). Increasing returns to scale are to a certain extent constrained in this process and accordingly agglomeration by means of industrialisation and urbanisation may be impaired. Subsequently, economic growth and development are adversely affected.

Although urbanisation and industrialisation may have been greatly impacted on by the system of apartheid, Houghton (1973:17-18) contends that the period ranging from 1945 (the end of WW II) to about 1993 can be considered the fourth stage of economic development labelled the drive to maturity. His argument, however, is based on speculation since by the time of the writing of his work 1993 had not yet come to pass. He, however, supported the argument that the beginning of the stage was due to the fact that the government started to implement more strategies to influence the geographical distribution of economic activity by means of industrialisation. In the preconditions to take-off stage government intervention was already evident regarding the growth of industries. The Law on Customs and Excise of 1914 (amended in 1925), the establishment of the National Industrial Corporation in 1919, and direct entry to industry with Evkom in 1923 and Yskor in 1928 were all examples of these interventions. The next stage, the take-off, marked the beginning of government intervention and industrial development aimed at the geographical distribution of industries. The founding of the Industrial Development Corporation in 1940 and the appointment of a Social and Economic Advisory Council were the most important institutions during this stage (Bos, 1992:224-225). These policies were, however, mostly aimed at decentralising industrial activity and the cost of unequal development would later have a significant impact on the South African economy (Naudé & Krugell, 2005:89). Subsequently the most important reports, strategies and implementations will be discussed in order to determine the decentralisation of industrial activity.

3.2.3.1.1 The Tomlinson report of 1954

During the first half of 1950 the cabinet considered the appointment of a commission to investigate the possibility of socio-economic development in the African regions. The cabinet later on approved the appointment of this commission in April 1950. The cabinet's view was that the white population was obligated to retain stewardship over Africans, but these Africans were responsible for the control of their own affairs. Furthermore, they believed that African urbanisation needed to be stopped and that the geographical concentration that occurred needed to be rectified by shifting industries towards the population and not the other

way around (Bos, 1992:235; Wessels & Wentzel, 1989:126-128).[75] Consequently, a commission under the leadership of Professor Frederik Rothmann Tomlinson was officially appointed on 22 November 1950 by the secretary-general, Dr Gideon Brand van Zyl (Wessels & Wentzel, 1989:128).

The Tomlinson Commission was given the instruction to investigate the possibility of rehabilitation of the native areas and write a comprehensive report regarding this matter. The ultimate goal of this report would be to promote the development of independent native structure which was based upon effective socio-economic strategies (Wessels & Wentzel, 1989:129; Houghton, 1956). After four years of intensive research by the Commission a complete report was released. The focus of the report was the development of industries in the African areas with the objective of providing employment opportunities to the unemployed. According to surveys the Africans were more productive in the industrial sector than in agrarian activities (Bos, 1992:236). The Tomlinson Commission also believed that the policy of separate development was a prerequisite for the national development of the Bantustans and the Bantus would have their own inalienable territory in which individual and societal, economic, political and social growth and development would occur. Separate development would also be advantageous to the white population since they would be able to live in a free and harmonious society without racial conflict. Furthermore, the whites would be able to perform their tasks as care-taker of the Africans. According to the report the people of South Africa had to choose between eventual total apartheid or eventual total integration between the Bantus and Whites (Wessels & Wentzel, 1989:143-144; Houghton, 1956).

3.2.3.1.2 Decentralisation policy of 1960

Although the Tomlinson commission emphasised the importance of the promotion of separate development of the African territories some years passed before the implementation of the proposals. Investigations were launched in 1959 in order to determine where industries could be established on the borders of the African territories. Furthermore, these territories were examined for the availability of factors of production. Based upon the results of these investigations, decentralised areas which would be efficient were identified and proposed. However, the investigations also emphasised the important role large already established cities would play in the future regarding development in South Africa. In addition it was realised that rural development would only take place if supporting strategies were

[75] This was strongly in conflict with the natural evolution of the distribution of industries according to spatial theory.

implemented. It was also realised that industrial development would be an imperative process in order to achieve the desired pattern of racial segregation (Bos, 1992:240).

The government was well aware of the fact that the decentralisation policy would have to be implemented gradually since immediate implementation would result in extensive demands which might result in unintended effects. However, the first official decentralisation policy was ratified in 1960. The policy aimed to encourage the establishment of white industries near the borders of the African territories. The decentralisation of economic activity by means of decentralisation of industries was an attempt of the state of the Republic of South Africa (RSA) to create increased employment opportunities near the labour market, which were the homelands. On the other hand it was also an attempt to oppose the excessive clustering of economic activity and was aimed at balanced territorial growth (Bell, 1973; Bos, 1992:240-241). According to Wellings and Black (1984:5) the function of the growth points was firstly to provide Bantus or the excessive labour supply in the periphery with employment opportunities by means of commuting to the manufacturing industry near the growth centres and secondly the exploitation of the benefits of the surrounding areas through some type of employment and income multiplier instrument.

The emphasis of policy was initially on bordering areas of the Bantustans. According to Palmer (1980:36) bordering areas can be defined as a place where in which industrial development takes place through White control and inventiveness, and are situated in such a manner that African workers from the Bantustans can maintain their residences and readily move to their location of employment. Later on a permanent committee for establishment of industries in the bordering areas was established. The permanent committee was the first government institution established whose aim was to address the establishment of industries according to the decentralisation policy. In addition, the function of the committee was to determine the growth points in bordering areas and coordinate and support these points. Support came in the form of incentive schemes which included financial support, allowances regarding income taxation, provision of basic services such as water and electricity and subsidisation of these points in order to increase the wage share of Bantus (Bos, 1992:241-244).

3.2.3.1.3 The Agency system of 1968 and the Law on Physical Planning of 1967

According to Kotzenberg (1973:148) the incentive schemes did not deliver the results which were expected. The border areas were unsuccessful in attracting investment and the influence

on development of the Bantus was insignificant. Accordingly the response to the problem of the deteriorating Bantustan sub-economies was to introduce the Agency system in 1968. It allowed white capital investment into the Bantustan areas although it was in conflict with the ideological view of separate development in 1968 (Wellings & Black, 1984:5).[76] According to the Agency system, courtesy would be given to white entrepreneurs to establish themselves in the homelands. The introduction of the system was due to government's realisation that manufacturing industres were an essential component of economic growth (Bos, 1992:248-250).

At the same time the Law of Physical Planning 88 of 1967 was implemented and signified a new era in the implementation of industrial decentralisation programmes. The law used negative measures in order to manipulate the establishment of industries and control urbanisation in the Republic. The justification for the promulgation of this law was that industrial development was too slow in large industrialised areas, especially in the Pretoria-Witwatersrand-Vereeniging-area (Bos, 1992:245; Wellings & Black, 1984:5).[77] The reasoning behind the law was that the development of industrial activities around the large industrial areas needed to be predominantly white industries which are place-based, while the predominantly black industries were non-place-based due to current economic conditions and would not contribute to the development of the Pretoria-Witwatersrand-Vereeniging (PWV). Consequently, these black industries needed to develop in decentralised areas (Bos, 1992:245).

The Acts intensified restrictive policies regarding the establishment of industrial activities. In terms of Section 2 of the Act, without preceding written consent by the Minister, no proclamations might be made regarding industrial land if the land was not zoned in an area that was preserved for industrial development; no land that was zoned for industrial development might be sub-divided; and no industrial settlement established (Bos, 1992:245; Turok, 2012:9; Oranje & Merrifield, 2010:33). Section 3 also restricted the extensions to existing industries (Wellings & Black, 1984:5). The result was that positive measures in the form of incentive schemes were combined with negative measures in the form of restrictive Acts (Bos, 1992:247).

---

[76] It was also proposed 15 years earlier by the Tomlinson Commission (Wellings & Black, 1986:138).

[77] The Port Elizabeth-Uitenhage, Cape Town and Bloemfontein areas also showed minimal growth in the industrial sector (Wellings & Black, 1984:5).

3.2.3.2 The 1970s

It appeared by this stage (by using official statistics as guideline) that government policy had been successful in reducing the rate of urbanisation of Africans in South Africa. In the period from 1951 to 1960 the rate of African population growth was a mere 4.5% annually. After 1960 the annual growth rate decreased to 3.6%. Furthermore, according to estimates there was a net out-migration of 750 000 Africans from the urban areas between 1960 and 1980 (Gelderblom & Kok, 1994:96). This was mainly a result of the implementation of resettlement policies between 1960 and 1980 which forced people to move to their designated group areas or Bantustans. At the very least two million people were forcibly removed during this period (Turok, 2012:9). Not only the African but also the Indian and coloureds were expelled from certain areas. By 1972 more or less 45000 coloured- and 28000 Indian families were moved as a result of Apartheid legislation (Booysen, 1985:22). An example of a brutal resettlement policy was the Black Homeland Citizenship Act of 1970. The Act permitted the removal of citizenship from Africans living in the country and accordingly they became citizens of one of the ten Bantustans, depending on their ethnic classification (Turok, 2012:9). According to Booysen (1985:30) the percentage of urbanised blacks by 1970 was 33.1% and by 1980 it was 32.9%. In contrast 87% of the white population had already been urbanised by 1970 (Oranje & Merrifield, 2010:33). The goal was to secure the existence of a demographic majority of whites and at the same time augment the economic, political and spatial riddance of non-whites (Turok, 2012:9). Two contradictory trends had developed. On the one hand blacks left the urban areas (mostly through forced removals) and on the other hand employment opportunities became concentrated in urban areas as a result of the ineffectiveness of the decentralisation policies. Accordingly the location of employment of and residences of blacks were separated and the settlements that developed in the homelands had little economic potential (Gelderblom & Kok, 1994:99).

3.2.3.2.1 The Riekert and Franzen Commissions

The precise impact of the previous years' decentralisation policy was difficult to determine. However, it was obvious that it had important consequences for job creation within the metropolitan areas on the one hand. On the other hand it had a minimal influence on industrial development in the homelands. Many metropolitan employers closed down their industries, established in less restrictive urban areas, employed races other than Africans, and resorted to capital-intensive rather that labour-intensive methods, rather than decentralising their industries (Wellings & Black, 1986:138). The ineffectiveness of these policies was

evident. According to Naudé and Krugell (2005:89) unequal development had a particularly strong influence on the productivity of the manufacturing sector. By 1970 South Africa's manufacturing sector was relatively advanced and diversified. However, the period following this marked a decrease in production and employment. Reasons for this included: the decrease in gold exports and prices, the reduction in global demand for commodities from the 1980s, the debt crisis of the 1980s, the devaluation of the Rand, sanctions imposed against the country, shortages in foreign exchange, and capital and skill shortages (Bell, 1987; Naudé & Krugell, 2005:89).

The result of the decrease in manufacturing productivity pointed to the need to evaluate the existing industrial decentralisation policies. This led to the appointment of an interdepartmental commission in 1970 under the leadership of the chairman P.J. Riekert. The objective of the commission was to give coordinated advice regarding the steps that needed to be taken in order to efficiently apply the decentralisation policies and the establishment of industries in decentralised areas. The White Paper on industrial development was published by the commission in 1971. The Paper made recommendations regarding the method that could be followed to make industrial decentralisation more streamlined and provided ways in which the institutional framework of the policy could be improved. The main idea was still that industrial activities should be shifted towards the periphery; however, in this process the needs of the existing industrial centres should also be taken into account. In addition it was noted that in order to achieve successful decentralisation the authority would need to prepare the Bantustans for industrial development by means of provision of facilities and infrastructure. In other words, an environment conducive to industrial growth needed to be developed beforehand in the form of adequate provision of housing, education, recreational and social facilities (Bell, 1987; Bos, 1992:250-253).

In 1972 another report on industrial decentralisation appeared, carried out by the Franzen Commission. The report contained various objections of the decentralised industrialists. The objections included that the decentralisation increased their risks of doing business and this led to the discouragement of other industries to establish themselves there. Furthermore, the incentive schemes were not efficient enough in ensuring the welfare of these industries. The commission derived the conclusion that decentralisation induced large indirect costs for industries and this led to the questioning of the validity of decentralisation (Moolman, 1984:70-71).

3.2.3.2.2 The National Physical Development Plan of 1975

All previous measures regarding the decentralisation of industrial activity had been planned and put into action by the government; however, they were not the result of an intensive planning process and could be considered "focused ad hoc measures" (Oranje & Merrifield, 2010:34). However, in 1975 the Department of Physical Planning instituted the National Physical Development Plan (NPDP). The NPDP was the first step towards a coordinated and overarching physical development plan in South Africa (Bos, 1992:256). The plan was a consequence of a report put together by a committee of the Prime Minister's Planning Advisory Council who were given the task of examining and giving recommendations on the connection between various levels of government regarding physical planning (Oranje & Merrifield, 2010:34).

The NPDP stated that South Africa needed to be sub-divided into 38 planning areas in addition to the four large urban areas. The homelands were excluded from the plan since they were regarded as independent states and not part of the country (Oranje & Merrifield, 2010:34). Strategies were implemented for the development of existing metropolitan areas, planned metropolitan areas, growth points and main settlements (Bos, 1992:260). Furthermore, the plan included development "axes" which would connect inland metropolitan areas and planned metropolitan areas with existing and proposed ports and/or important mining or industrial activities (Oranje & Merrifield, 2010:34). The plan was an extension of the Group Areas Act since the latter was a demarcation of separate development, whereas the NPDP could be considered the implementation of separate development in South Africa (Bos, 1992:258). According to Oranje and Merrifield (2010:34) an essential component of the plan was the influence it would have on the private sector by the end of the 70s and early 80s. It appeared that the Botha government wanted to put a wedge between economic relations between the periphery (homelands) and the core (white urban areas). This would force the private sector which was primarily run by the urban areas to make Grand Apartheid an economic reality.

3.2.3.2.3 The Carlton summit and the Small Business Development Corporation of 1979

By 1979 government still believed the decentralisation process was unsatisfactory in its implementation. According to them the desired results had not been obtained due to sub-standard incentive schemes. The periphery was still unable to attain the needed growth of industries as well as employment opportunities. This all led to a summit on 22 November

1979 between the central government and the private sector which is historically known as the Carlton Summit. The summit revolved around issues such as the territorial development of the homelands, the establishment of a developmental bank for Southern Africa and the promotion of small enterprises. The summit proposed the confederation of Southern African states and suggested prospects of greater regional integration and development. The recommendations regarding small businesses would lead to the establishment of the Small Business Development Corporation (SBDC) of 1979 (Bos, 1992:265-266; Oranje & Merrifield, 2010:34).

The SBDC was a unique partnership between the state and the private sector. The shares of the corporation were divided equally between the two parties and the government's was basically handled by National Development Committee. The establishment of the SBDC was the reaction to the appeals for a national organisation which would be able to provide financial support to small businesses. The goals of the SBDC were to promote small businesses of all racial groups, provision of financial support, development of infrastructure of small businesses, provision of advice to the businesses they financially supported and most importantly the decentralisation of small businesses (Bos, 1992:266-267). It was evident that the government wanted to move in a new direction; however, what was also apparent was that decentralisation would still be advocated strongly although little or no success had been achieved in previous years. On the one hand the forces of agglomeration were pulling economic activity together and on the other hand government intervention was trying to disperse it.

3.2.3.3 The 1980s

Since the late 1970s the apartheid state started to encounter a persistent and intensified crisis of legitimacy. The crisis of political legitimacy was spurred on by the economic depression at that stage and solutions to this crisis were starting to disappear. The government started to experience increased opposition[78] from the townships and Africans and international hostility towards the country escalated. The weight of sanctions against the country in conjunction with the economic depression started to take a hold (Cobbett *et al*., 1985:1; Wellings & Black, 1986:144). Employers started to question the validity of the restrictions on the movement of black people and other associated procedures. Firms started to bypass laws in terms of the contracting of black employees since it was administratively costly and too

[78] According to Turok (2012:11) the black populations started to become more militant and townships became ungovernable. The living standards also deteriorated, resulting in even more conflicts.

complicated (Turok, 2012:11). Despite all this pressure the ruling group was still attempting to create justifications based upon long-term strategies with the hope that it would resolve the political conflict and allow the re-structuring of the economy (Cobbett *et al*., 1985:1; Wellings & Black, 1986:144).

It had become inevitable for the South African government to incorporate Africans into a single national state by extending their political rights. However, the goal of new policies and implementations was to ensure that the real power still belonged to the ruling classes (Cobbett *et al*., 1985:1). Consequently, industrial decentralisation had to be reorganised by the government during the 1980s. It was, however, unexpected that the state would react by increasing expenditure on decentralisation and simultaneously following firm monetary policies in an economy that was constrained by the recession (Wellings & Black, 1986:144).[79] A reason for this, according to Hyslop and Tomlinson (1984:115), was that, although Bantustans at that time still had restrictions in terms of their rights and privileges, a trend had been established with the incorporation of coloureds and Indian minorities into the political sphere when restrictions on these groups were lifted. The whites feared that the tendency might lead to the inclusion of blacks later on. They posed a threat to the ruling white population and accordingly emphasis in policies was still on the separate development of the blacks. However, for this to be possible the incentives, in the form of development of the Bantustans' economic base, and disincentives, in the form of influx control, would have to be strengthened (Hyslop & Tomlinson, 1984:115).

3.2.3.3.1 The Good Hope summit and establishment of the Development Bank of Southern Africa

Although decentralisation policy had intensified over its history, strong clustering of industrial activities was still evident in the metropolitan areas. Around 75% of industrial production and employment was provided by large metropolitan areas in South Africa. The PWV-area contributed 52% of this production and 45.7% of total employment in the country by 1978. Furthermore, the PWV-area employed 3.1 million Africans. However, unequal distribution of economic activity was clear when the industrial production of this area (with its 3.1 million employed Africans) was compared with the industrial production of East London, Richards bay, Pietersburg and Bloemfontein who had 4.1 million Africans employed. The combined areas contributed only 4.7% of industrial production in the country.

[79] For instance the National budget allocated 138% more towards industrial decentralisation in the 1985 to 1986 fiscal year than the previous year (Wellings & Black, 1986:144).

According to the government it was imperative that new proposals be made regarding the development of decentralised industries in order to raise the living standards of Bantus (Bos, 1992:283-284).

In response the Good Hope summit was held on 12 November 1981 in order to consider new proposals for industrial development (Bos, 1992:284). The Good Hope summit was primarily held to find ways in which private sector investment in the Bantustans could be increased (Tapscott, 1990:235).[80] The result of the summit was that in April 1982 the White Paper in terms of the development of industries was published (Bos, 1992:284-285). This so-called Good Hope Plan or Regional Development Strategy for Southern Africa included eight new development regions in addition to the 38 of the NPDP and included the four metropolitan areas as well as the various Bantustans. The indication of development axes was also absent. Alternatively, a set of eight de-concentrated growth points in metropolitan areas and 20 decentralised points for industrial development were recognised (Oranje & Merrifield, 2010:34). This suggests that it was recognised that the concentration of economic activity was needed in economically conducive growth points, and not necessarily near or in the homelands (Wellings & Black, 1986:138). The White Paper also emphasised the importance of coherent regional development framework for Southern Africa as a single entity where not only industrial development but also development of the primary and tertiary sectors were necessitated to achieve sustainable economic development. According to the paper the development of the other sectors would stimulate the growth of industries. Furthermore, it was recognised that each region had its own development characteristics which translated into different requirements and possibilities for development. Consequently, regional development would only take place if the country was viewed as a single unit sub-divided into large development regions which stretched across various independent states (Bos, 1992:285). Some were very sceptical about this new plan that shifted away from the ideological idea of separate development and believed it was an attempt to dismiss the responsibility for failures of the decentralisation plan. For instance, Tapscott (1990:235) thought that it was an attempt of the state to be removed from the "scene of the crime" that was apartheid policies. Accordingly, some thought the initiative was politically and not economically motivated.

[80] The Good Hope summit signified a formalised regional development plan for South Africa. It was first suggested in the NPDP of 1975 (Tapscott, 1990:233).

The Development Bank of Southern Africa (DBSA) was initially proposed by the Carlton summit in 1979 (Tapscott, 1990:232). However, with the implementation of the Good Hope plan increasing private sector involvement was emphasised and this led to the establishment of the DBSA in 1983. The aim of the Bank was to mobilise funds through the private sector for the sustained development of the growth points (Wellings & Black, 1986:138-139). Membership to the Bank was permitted to any independent state in Southern Africa. Consequently, South Africa and the independent states of Transkei, Bophuthatswana, Venda and Ciskei (TBVC) were members of the Bank. The institution was funded by shared capital, capital market loans as well as contributions by members to the development fund. The DBSA would then in turn finance development of less-developed regions as well as development organisations such as the SBDC (Bos, 1992:304-305). Although authorities stated that the Bank was a multinational organisation and by no means a political instrument, pundits were suspicious.

3.2.3.3.2 The Kleu Report of 1983

In 1977 the Minister of Economic Affairs appointed a research group under the leadership of Dr. S.J. Kleu as chairman. The group consisted of representatives of both the private and public sector and was given the task of formulating an industrial development strategy as a component of the whole economic policy programme. Furthermore, the group had to formulate a plan that would prevent the problems of previous industrial strategies from recurring. In 1983 the group presented the Kleu Report (Bos, 1992:306; Van Zyl, 1983:30).

The report's envisaged strategy was mostly in line with mainstream economic ideas based on the market system. For this very reason it was criticised for being too conventional and not taking into account the present market conditions in the form of increasing international hostility as well as the need for sectoral transformation support of especially the underdeveloped Bantustans. However, the report emphasised an important fact regarding industrial strategy in South Africa. According to the report the strategy in the country had now come to a stage where certain choices could have lasting impacts and thus the correct decisions would have to be made in terms of industrialisation (Van Zyl, 1983:31; Fine & Rustomjee, 1996:197).

The report pointed towards prerequisites for industrial strategy to be successful in South Africa. Industrial competitiveness of the country would need to be addressed and the government's ability to evaluate various policy choices would have to be improved.

Consequently, government realised that they would have to obtain greater knowledge of the factors influencing competitiveness such as technology, domestic market structure and competitiveness and competition from abroad. In other words, the report stressed the need for government to critically analyse area-specific industrial factors that had led to the failure of development, especially in the Bantustans. In the past it was never a focus of government to collect industry-specific information involving the independent states (Bos, 1992:306-308; Van Zyl, 1983:31). The report also acknowledged that the analysis of competitiveness must be directed to individual sectors of the economy. In other words, industrial strategy needed to be supported by the other sectors and consequently those sectors also needed targeted subsidisation (Van Zyl, 1983:32).

The report moved away from the apartheid ideological perspectives and emphasised that specific measures should not be disregarded as possible options (Van Zyl, 1983:32). According to Fine and Rustomjee (1996) a relaxation of state intervention was detected which could be attributed to the global inclination towards liberalisation. For example, the report acknowledged the fact that southern Africa should be seen as a single entity and that all the states were interdependent. More importantly, it was recognised that unless decentralised industrialisation was contributing productively to development in a region, it would not lead to economic growth and development in the country (Bos, 1992:308). However, this did not mean that decentralisation was not a focus. In fact it was proposed that decentralisation policy be enhanced (Fine & Rustomjee, 1996:198).

3.2.3.3.3 White Paper on industrial development of 1985

The White Paper on industrial development was established in 1985 as a result of subsequent discussions of the Kleu Report by the Industrial Advisory Committee (Fine & Rustomjee, 1996:197). The Paper emphasised the necessity to develop a more productive manufacturing sector, a better technological policy, re-orientation of industry from import substitution to export promotion and a stronger and more effective decentralisation effort in order to achieve improved industrial development (Wellings & Black, 1986:142; Bos, 1992:308-309). The improvement of industrial competitiveness was the key phrase and reconstruction of the industrial sector was the building block in this strategy. However, the policies that were adopted to affect the reconstruction process in order to achieve international competitiveness put a lot of strain on an economy that was already experiencing pressure as a result of the recession (induced by monetary policy), political unrest and large capital outflows.

Consequently, the reconstruction of the industrial base would have significant spatial implications for the location of economic activity (Wellings & Black, 1986:142).

Since the late 1970s and early 1980s increased investment and incentive schemes did have a substantial impact on decentralised industries. According to Wellings and Black (1986:142) the period from 1972 to 1978 manufacturing employment in decentralised areas had only increased from 12.3% to 12.9%, whereas from 1978 to 1984 it had increased from 12.9% to 19.3%. However, the pace of decentralised industrial development was still not sufficient in solving the unemployment problem in less-developed areas since the population of South Africa was rapidly increasing (Bos, 1992:310; Dewar *et al*., 1986:365). It was a fact that many industries were attracted to the periphery although the metropolitan areas provided the benefits of agglomeration. However, this attraction to locate in remote areas was driven by the extremely cheap labour. For instance, according to Wellings and Black (1986:142) Isithebe, a growth point 100 kilometres from Durban, offered 45% lower wages than Durban for machinists in the textile industry. This was a result of the surplus labour and unemployment (due to influx control and forced removal of Africans from the cities) in these areas and consequently the living standards and welfare of the majority of the people living in the periphery were still appalling (Wellings & Black, 1986:142). As a result the White Paper emphasised that the industrial sector development should not enjoy more attention than any other sector in the economy. The primary sector should also be seen as an essential component in the development of the country (Bos, 1992:313). However, capital investment at that stage was hampered by the political and economic crisis and state expenses were increasing. Furthermore, Kleu's ideological view had not materialised in the form of greater integration between the state and the private sector (Fine & Rustomjee, 1996:198).[81] It was evident that the decentralisation of economic activities in conjunction with restriction of urbanisation was not successful and would never be successful in the development of all groups in South Africa (Gelderblom & Kok, 1994:90).

3.2.3.3.4 The Urbanisation Strategy of 1986

In 1985 the government recognised that relatively little progress had been made in terms of the economic development of the less-developed areas in South Africa, especially in the homelands (Dewar *et al*., 1986:365). The enormous growth of informal settlements in the 1980s was a good indication of the shortcomings of previous government policies with regard

[81] This was a direct result of the exhaustion of incomes from the mining price increases in the early 1980s as well as adverse reactions towards apartheid (Fine & Rustomjee, 1996:198).

to decentralisation and urbanisation. The circumstances and location of these settlements revealed the inefficiency of the country's economy to provide them with the basic platform for development. The settlements had shifted away from their rural status; however, they could not be considered truly urban in nature (Mabin, 1991:40). As a consequence government officials realised that they could no longer ignore the need for the permanence of the urban African population (Gelderblom & Kok, 1994:91). The President requested that the President's Council should advise him regarding a strategy which could aid in the orderly urbanisation of Africans in order to address the social, economic and physical problems that the country faced. The President's council responded with a report and in 1986 a White Paper based upon their advice was promulgated (Bos, 1992:321; Hindson, 1987:589).

The Urbanisation Strategy of 1986 was based upon 'orderly' or 'positive' urbanisation which effectively meant that government recognised that the racial and regional inequalities in the country were a direct result of misplaced government intervention in the economy. The racially-biased government policies of apartheid essentially in the form of influx control, the establishment of homelands and industrial decentralisation were the main causes of poverty, inequality and unemployment. The regime resulted in the misallocation of the factors of production in a regional and sectoral fashion which prevented the South African economy from reaching its potential in terms of levels of output, welfare and employment. Consequently, the government realised that the prejudice and government mismanagement had led to the contemporary urban crisis (Hindson, 1987:588).

The result of this realisation of government culminated in the abolition of the influx control system (Gelderblom & Kok, 1994:91; Murray, 1987:311). However, the repeal of influx control did not mean that complete freedom of movement was granted to Africans (Mabin, 1991:41). The orderly urbanisation set out in the Urbanisation Strategy accepted that African urbanisation was inevitable; however, it attempted to restructure the process within a regional planning framework. The framework was made up of nine new development regions which replaced the Bantustans and was intended to economically re-integrate these less-developed regions (Hindson, 1987:589). The White Paper, however, still emphasised the need for decentralisation in order to control the economic, political and social impact it might have on the metropolitan areas and large urban areas (Dewar *et al*., 1986:365).

The orderly urbanisation reflected a form of decentralised apartheid in which the market forces played a greater role within the regional planning framework grounded upon the idea

of racial segregation and dispersion of population and industry. Consequently, the abolition of influx controls was replaced with new measures to restrict the movement and settlement of Africans in areas other than the Bantustans. An example of this was the Aliens Act which restricted Africans from permanently resettling and obtaining citizenship rights in areas not specifically designated to them (e.g. their homelands).[82] Furthermore, the orderly urbanisation strategy entailed the use of fiscal measures to promote the concentration of labour-intensive and unskilled industries away from large urbanised areas. These measures were consequently used to reduce urbanisation in metropolitan areas and also direct new urbanites to racially ordered communities on the periphery of metropolitan areas (Hindson, 1987:589-591). Although African urbanisation was seen as desirable it was still based upon certain assumptions and conditions. The African population needed to be primarily accommodated in de-concentrated points outside the core metropolitan areas. Accordingly the industrial decentralisation could still be seen as the mechanism in the promotion of development in the periphery and a means of keeping Africans away from the metropolitan areas. In other words it was an essential component of the reformulated apartheid policy (Dewar *et al.*, 1986:366).

The White Paper of orderly urbanisation was to a great extent in line with the Urban Foundation's plea for a positive urbanisation strategy in South Africa. In the build-up to the White Paper on Urbanisation, the Urban Foundation publicly opposed the racially-biased measures that had been used up to that stage. However, state planners contended that the Urbanisation strategy needed to be based upon the regulation of urbanisation since unregulated urbanisation would result in overcrowding which would in turn result in the deterioration of living standards in the established residential areas. They argued that government intervention was necessary to address the market failures that would occur if masses of the population suddenly moved to urban areas (Hindson, 1987:593). Although apartheid was still embedded in this strategy and racial laws such as the Group Areas Act, the Aliens Act and the Population Registration Act remained intact, the Urbanisation strategy and abolition of the influx control could be seen as the start of the elimination of Grand Apartheid in South Africa (Mabin, 1991:41; Hindson, 1987:593; Murray, 1987:311).

[82] Further measures included trespass, squatter and health laws (Hindson, 1987:590).

3.2.4 Post-Apartheid South Africa

By the late 1980s there was a strong national and international contention that the apartheid system could not be justified and would never work. Infrastructure as well as social and physical circumstances had deteriorated in the African settlements and this spurred on political opposition to the system. Furthermore, the economic and social pressure created by apartheid was ever mounting. The government recognised that South Africa needed to be directed onto a new path in order to avoid the civil and political turmoil that it faced. They had no choice but to release political prisoners such as Nelson Mandela and to unban the African National Congress (ANC) and other opposition parties in 1990. For a period of four years the country experienced social unrest and political violence until the first democratic elections were held on the 27th of April 1994. This marked a new dispensation led by the ANC which was characterised by increased national unity and social stability (Turok, 2012:11-12). However, the ANC was left with a country with large differences in the standards of living between racial groups, unequal distribution of economic activities in different regions, people unable to obtain basic services and quality of living, and large-scale public debt (Oranje & Merrifield, 2010:35).[83]

3.2.4.1 Urbanisation

Although the democratisation of the country brought new political freedom which was regarded as the premise of inclusion and economic welfare, the impact of apartheid in the form of restrictive movement and decentralisation policies on urbanisation in South Africa was massive (Bhorat, Hirsch, Kanbur & Ncube, 2014:2). Table 3.1 indicates that from 1951 to 1991 the urban share of the population had only increased by a mere 10.4%. Consequently, the level of urbanisation had been artificially impaired by apartheid legislation (Collinson, Tollman & Kahn, 2007:78). As could be expected after the abolition of restrictive measures regarding urbanisation such as the influx control and Group Areas Act an increase in rural to urban migration in the country was evident (Turok, 2012:12; Boraine, Crankshaw, Engelbrecht, Gotz, Mbanga, Narsoo & Parnell, 2006). With the freedom of mobility of Africans reinstated the benefits of agglomeration started to take effect. Table 3.1 indicates that between 1991 and 1996 the total urban population in the country increased by an estimated 7.8%. However, during the following five-year period from 1996 to 2001 there was an increase of only 3.2%. An explanation for this may be that the abolition of restrictive

[83] The country had a low growth rate and high levels of unemployment and inequality (Naudé & Krugell, 2005:89).

measures on African movement substantially increased urbanisation in the initial stages; however, a poverty trap was created by apartheid for a large proportion of the poor African population residing in the former homelands which made it difficult for them to urbanise and escape their situations (Choe & Chrite, 2014:84).[84] The unemployed family members were accordingly forced to remain in densely populated rural areas where they remained aliens to employment opportunities (Collinson *et al*., 2007:78; Murray, 1987:312). Furthermore, the expectation that labour migrants would be replaced by permanent settlement in urban areas did not materialise. Reasons for this may be that remittances to rural areas did not decrease after the emergence the 'new South Africa'. The increase in the social pension and the execution of the National State Subsidy Scheme for development of housing in the decentralised regions may have reduced the attractiveness of the opportunity of higher income in cities for Africans (see Posel, 2003, Bremner, 2000 and Collinson *et al*., 2007).[85] Another reason for the temporary nature of migration might be that Africans are socially bound to their households of origin. For example, in a volatile and insecure labour market and in a country marked by increasing unemployment, these households provide (perceived) 'insurance' for job-seekers, a safe house to take care of children, as well as a place for retirement (Posel, 2003:17).

All these factors reinforced the African peripheral pattern and segregation of urban poor residential areas created by the apartheid policies. Despite the fact that administrative officials implemented plans for the integration of segregated metropolises in the early stages of democratisation these mechanisms and procedures failed to alter the effects of apartheid (Bremner, 2000:88; Bond, 2003:42).[86] According to Bond (2003) the main urban policies initially used include the Housing White Paper and Water and Sanitation White Paper of 1994; the Urban Development Strategy of 1995; the Municipal Infrastructure Investment Framework of 1997 and 2001; the Local Government White Paper of 1998; and the Energy White Paper of 1998. According to Bremner (2000:87) these attempts by the newly-elected ANC government have not generally been place-based policies and practices that counter the existing apartheid spatial geography or recognise the changing social patterns of urban areas

[84] Bond (2003) describes the post-apartheid era as "class apartheid". It refers to the fact that although structural racism had been removed since 1994 under-development and inequality still persisted for Africans.

[85] According to Posel (2003:8) in 1993 rural African households held 32.6% migrant labourers and in 1999 it increased to 35.8%.

[86] Such as the Interim Strategic Framework, Integrated Framework and Integrated Metropolitan Development Plan, implemented by means of the Development Facilitation Act 67 of 1995 (Bremner, 2000:88).

in post-apartheid South Africa. On the contrary, as mentioned in the previous paragraph, policies such as the National State Subsidy Scheme and investment projects in the under-developed regions have to a certain extent reinforced the phenomenon called "displaced urbanisation" (Murray, 1987:312; Bond, 2003:41; Oranje & Merrifield, 2010:35).[87]

Table 3.1: Urban share of population in South Africa, 1911-2011

Year	National Population (thousands)	Urban share of population (%)
1911	5,973	18.2
1921	6,927	19.8
1936	9,588	25.8
1951	12,671	35.2
1960	15,994	37.9
1970	23,311	38.5
1980	29,208	42.5
1991	38,012	45.6
1996	40,580	53.4
2001	44,819	56.6
2011	51,771	62

Source: Turok (2012) and Statistics South Africa (2012).

As can be seen in the previous paragraph, the first-decade post-apartheid urban policies were aimed at rectifying the problem of apartheid based upon the ideas of non-discrimination, reconstruction and development of the previously disadvantaged rural population and was not necessarily aimed at the development of urban settlements in order to accommodate the rural population. However, the last few years have marked a shift from overcoming apartheid to a developmental state which aspires to induce urban innovation. This aspiration by the state is envisaged in the Breaking New Ground strategy (initiated by the Department of Housing) which is based upon the idea of integrated and sustainable human settlements and also more recently the call by the National Development Plan (NDP) for an Integrated Urban Development Framework (IUDF) which acknowledges that South African cities can be successful drivers of economic growth by means of enhanced spatial efficiency and social

[87] The phrase referred to the clustering of Africans in rural slums in the homelands and on the peripheries of large cities due to apartheid policies (Murray, 1987:312).

inclusion (Boraine *et al*., 2006; Department of Cooperative Governance and Traditional Affairs, 2014). Another important event was the creation of the South African Cities Network (SACN) which is considered by some the reason why new urban policy documents do not expect local government to be the only representative of urban management. *The State of South African Cities 2004* delivered by the SACN breathed new insights into post-apartheid city transformation. The report suggests that it is imperative for cities to be well-functioning and sustainable in order to achieve economic growth (Boraine *et al*., 2006:261,264, 281).

3.2.4.2 Industrial policy

After 1994 it was expected that the government would institute large-scale, government-controlled plans and interventions in the South African economy. This, however, did not materialise since massive economic growth was needed to fund the reconstruction and development. The country's growth rate had been around 3% which would not suffice in the task of rectifying the socio-economic problems of the country. The lack of progress that had been made in terms of the apartheid space economy caused concerns for a number of departments including the Transport, Housing and Constitutional Development departments. They argued that the absence of coordination between government investment and expenditure as well as public and private sector investment and expenditure was not improving the spatial diseconomies caused by apartheid, but in some instances even exacerbating them (Oranje & Merrifield, 2010:35).[88] Accordingly industrial policy in post-apartheid South Africa has reflected growing application of planning regarding Local Economic Development (LED) (Rogerson, 1997:175).[89] LED refers to the phenomenon when settlements (or cities) become more entrepreneurial as economic development becomes more localised in terms of the shift of responsibility for development planning from national to sub-national or local levels. As a consequence the region or locality determines the coordinated effort to enhance the creation of economic activities which in turn translates into development. In other words the cities or municipalities become more independent (Rogerson, 1997:177). According to Rogerson (2008:307) the application of a more decentralised bottom-up method to economic development induces a strong foundation for economic growth by means of productive investment, improvement of social overhead capital and innovative technology, through developing key characteristics of each locality.

---

88 As noted in the previous section in terms of urban policies.

89 LED initiatives were also abundant during apartheid due to the decentralisation which was linked with the ideological vision of separate development. The LED initiatives after apartheid are considered as a "second wave" of policies aimed at urban reconstruction and development (Rogerson, 1997:176).

As can be thought of in the new democratised South Africa, regional industrial strategy abandoned apartheid's industrial decentralisation instruments directed at forcing industries to the periphery in order to achieve the ideal of separate development (Phalatse, 2001:167; Crush & Rogerson, 2001:85). In the initial stages of democracy the government instituted an 'a-spatial' industrial policy (Naudé & Krugell, 2005:90). This meant that the new industrial strategy revolved around the identification and promotion of Spatial Development Initiatives (SDIs) and Industrial Development Zones (IDZs) instead of the Bantustan industrial nodes (Phalatse, 2001:167; Naudé & Krugell, 2005:90).[90] The SDIs strategy was formulated in 1995 and implementation subsequently followed in 1996. From that stage onwards the government advocated the programme with great conviction (Crush & Rogerson, 2001:85; Oranje & Merrifield, 2010:36). The SDI programme aims to enhance investment, industries and other economic activities along a network of specified transport corridors. The problems of lacking infrastructure will consequently be removed by the SDIs which in turn will provide a basis for increased private sector investment and involvement in the country. However, the SDIs have not delivered the desired outcomes in the form of economic transformation envisaged by the programme, excluding the success achieved in the Maputo-Pretoria-Gaborone-Walvis Bay Development Corridor (Naudé & Krugell, 2005:91; Crush & Rogerson, 2001:85; Oranje & Merrifield, 2010:36). A further problem experienced by new industrial policy, according to Phalatse (2001:167), was that only a few of the previous Bantustan industrial nodes are situated within the SDIs and IDZs. Accordingly the outcome of the new industrial policy has in many cases led to the removal of incentives (in the form of investment) from the previous apartheid-created decentralised industrial zones which in turn induced de-industrialisation in these areas. This intensified rural under-development in these areas in the form of unemployment and poverty (Phalatse, 2001:167).[91] The implementation of the IDZs was also delayed by government. Only in 2001 was the plan put in action by the announcement of IDZ incentives.[92] The focus of these industrial zones is to enhance exports and accordingly they are situated near major transport nodes such as ports and airports (Naudé & Krugell, 2005:91).

---

90 Another initial plan was the National Spatial Development Framework (NSDF) which was proposed by the Reconstruction and Development Programme (RDP) in 1995. The plan was directed at efficient investment in reconstruction of underdeveloped regions in the country, although funding was lacking (Oranje & Merrifield, 2010:36).

91 See Crankshaw (2012) for an extensive review of the impacts de-industrialisation can have on social and economic aspects.

92 See Naudé and Krugell (2005) for further discussion of these incentives.

Another instrument that emerged during the late 1990s was the National Spatial Development Perspective (NSDP) as a result of the failure of the state to coordinate infrastructure and development expenditure in the other initiatives implemented, especially the National Spatial Development Framework (NSDF). The crux of the NSDP was that investment should be restricted to areas with economic growth potential and places characterised by high poverty. However, the idea of the NSDP was challenged by the government at that stage since it was in conflict with the assumption that the poor lived in rural areas and economic growth was restricted to the urban areas. The NSDP was consequently perceived in a negative manner. After successive revisions of the NSDP in 2002, 2003 and 2005/6 the 2007 updated version was given a much more positive outlook. This was as a result of optimism at that time in the country and the availability of hundreds of billions of Rand for investment in infrastructure. The principles of the document were still the same as in the initial version in the sense that it supported a forced national spatial development initiative; however, it was toned down to a certain extent and was far more specific regarding spending on investment (Oranje & Merrifield, 2010:36-37).

In the second decade of democratisation the application of LED initiatives intensified with the Department of Provincial and Local Government (DPLG) releasing a document which determined a specific framework that local authorities should follow in terms of LED. The document promotes an aspiration of strong and inclusive local development by using local economic opportunities in the sense that real potential and competitive advantages are exploited. Moreover, the document stressed that local needs and the contribution to the national development objectives were priorities. The release of the document led to a number of policy developments. These included the Department of Trade and Industry's (DTI) Accelerated and Shared Growth Initiative (ASGISA), the National and Industrial Policy Framework (NIPF), the NSDP, the Integrated Small Enterprise Development Strategy, as well as the Regional Industrial Development Strategy (RIDS) (Rogerson, 2008:308-309). The RIDS was a result of the attempts of the DTI to implement a regional development initiative. The strategy determined that the insufficient economic growth in the peripheral regions of the three largest metropolitan areas was due to regional inequality as a consequence of apartheid legacy. The aim of the strategy was accordingly to achieve balanced development by inducing state promotion of economic development in small towns and rural areas. The DTI argued that this approach could be implemented across the country, but made no mention of

the difficulties it would face in making these regions economically sustainable (Oranje & Merrifield, 2010:37).

More recently government and administrative officials have recognised that South Africa has struggled to achieve its development objectives in terms of industrial policy. The Green Paper on National Strategic Planning, for instance, indicates that in the future collaboration will be essential between State and private sector in order to achieve the country's development objectives. Furthermore, the Paper identifies the lack of coordination between National, Sub-national and local authorities has slowed down the effort (Oranje & Merrifield, 2010:37). The NDP of the National Planning Commission (2011:244) has also acknowledged that South Africa's spatial planning system has been slow to develop and coordination has been inadequate. However, the emergence of the NIPF and successive Industrial Policy Action Plans (IPAPs) since 2007 has shed new light on the implementation of industrial policy in South Africa. These strategies intend to enhance labour-absorbing industrialisation, improve the competitiveness of these industries and also stimulate employment and growth (DTI, 2007; DTI, 2013).

3.2.4.3 Effects of decentralisation and restrictive urbanisation on development

Although Bos (1992:217) speculated that South Africa would enter the phase of high mass consumption from 1993 onwards, it can be argued based upon Table 3.1's estimates, that the country as a whole may not have entered or was still in the process of entering a strong form of high mass consumption at that stage due to the impact of apartheid on the level of African urbanisation. The white population of the country may have moved into a phase of high mass consumption since they have reaped the benefits of large cities. However, South Africa as a whole (including all racial groups) still has to undergo changes in order to satisfy a large growing population. In Chapter 2 it was noted, according to Henderson (2003:227), that by upper middle income 60 to 90% of the population is urbanised in most countries. However, he also outlined the important fact that at these high levels stagnation of urbanisation and even decentralisation may occur. At some high level of clustering of economic activity the high level of development of the cities spills over to the rural periphery inducing the growth of smaller settlements. Accordingly, divergence in development is expected between urban and rural areas in the beginning of urbanisation and later on with high levels of urbanisation and development spatial decentralisation occurs which in turn converges the disparity between urban and rural development (as noted by Williamson, 1965 and Shisido, 1981 in Chapter 2). Table 3.1 indicates that by 2001 the urban population of the country still did not

pass the threshold of 60%. However, according to Boraine *et al*. (2006:264) four of the nine SACN cities'[93] annual growth was 1.4% from 1996 to 2001 which is less than the 21 largest urban centres population growth. This suggests that secondary cities are starting to develop since some of them have growth rates of up to 8% per annum. At the same time some of the SACN cities are growing rapidly (Boraine *et al*., 2006:264). Accordingly, growth of large cities and natural decentralisation was an occurrence during 1996 to 2001 and this indicates that South Africa is moving to an integrated core-periphery system. By 2011 the urban population accounted to 62% which is past the threshold.

According to Bhorat *et al*. (2014:3) three outcomes typify the structural changes: Firstly, mining activities constituted 11% of Gross Domestic Product (GDP) in 1994 and had decreased to 5% in 2012. Secondly, manufacturing accounted for 19% of GDP in 1994 and was marginally lower in 2012 at 17%. Lastly, the financial and business sector constituted 17% of GDP in 1994 and has subsequently increased to 24% of GDP in 2012. This indicates that South Africa has moved from being historically reliant on mining and resources to a more service-oriented economy. Accordingly the period from 1993 (according to Bos, 1992:217) may in fact only be considered as an extension of the drive to the maturity phase since the structural transformation by 1994 was not wholly inclusive of the entire population. Furthermore, the contemporary under-performance of the industrial sector may in fact be a result of decentralisation policies during apartheid in tandem with post-apartheid spatial policies which have been inefficient (see Bhorat *et al*., 2014 and Naudé & Krugell, 2005).[94] According to Rogerson (1997:177) economic restructuring across Western Europe and North America has been associated with de-industrialisation and a corresponding increase of service or information industries. Accordingly, the same has happened in South Africa. The policies aimed at redistribution have in fact led to the industrial activities and the corresponding employment opportunities it provides declining. This is due to the manufacturing sector being more labour-intensive than the service industry and consequently unemployment and poverty in the country have grown.

3.3 Conclusion

In South Africa first nature geography can be considered the root of uneven distribution of economic activity and population. Ports such as Cape Town, Port Elizabeth, East London and

[93] The SACN cities include Johannesburg, eThekweni, Cape Town, Ekurhuleni, Tshwane, Nelson Mandela Bay, Buffalo City, Mangaung and Msunduzi (Boraine *et al*., 2006:265).

[94] However this cannot be considered the sole reason for industrial underperformance (see Bhorat *et al*., 2014). It can be considered a supplementary reason for poor performance of the industrial sector.

Durban were initially the main bearers of economic activity and population. However, this changed with the discovery of diamonds in 1867 and gold in 1884 in the interior of the country. This led to the rapid movement of people to the sites of Kimberley and the Witwatersrand. Along with this, rapid movement industrialisation was sparked in the country. The mining activities as well as industries required a substantial workforce to sustain and accordingly large migrant labour to these sites of economic activity was stimulated. The increased demand for labour in industries and mining activities sparked large-scale urbanisation. South Africa quickly became a modern capitalist society. The first nature aspects accordingly influenced the modern day agglomeration of economic activity in large cities including Johannesburg, Pretoria, Ekurhuleni (East Rand), Cape Town, Durban and Port Elizabeth.

In the initial stages racial discrimination was already evident in the country since the white population held a monopoly on political power and land ownership was organised on a racial basis. This to some extent forced African labour into the mining activities which were characterised by precarious and unpleasant circumstances and provided low remuneration. The migrant labour system posed a problem for the state at that time and accordingly mining compounds or hostels were built in order to house Africans. The compounds were one of the first attempts of the local authorities to make separate residences for Africans. However, no general law existed at that stage regarding the right of ownership and occupation of property in urban areas. This changed when the Lagden Commission proposed the ideological vision of segregation of races.

The Union of South Africa was established in 1910. This was the beginning of the movement of the white Afrikaner to urban areas (at that stage the urbanised whites had been Europeans). The demographic transition was strongly associated with the increase of economic activities in cities. Increased demand for labour resulted due to the output of mining activities rising and this attracted industrialists. The newly-urbanised Afrikaners were in competition with the Africans for unskilled employment opportunities and both groups were extremely poor. The reaction of the government to the poverty problem of Afrikaners was, however, different to that of the Africans as a result of the white Afrikaners' political capacity. Accordingly, job reservations were granted to white workers in the industries by means of the so-called civilised labour policy. This to some extent forced Africans to work in the precarious and low-wage mines. Furthermore, the government neglected infrastructure development in the African neighbourhoods and African industrial workers were regarded as guests which in turn

required them to carry passes in order to move into urban areas on a temporary basis. The ideal of segregation had evolved to a reality and as a result the Union's Department of Native Affairs promulgated the Natives' (Urban Areas) Act of 1923. The Act was a catalyst for the increase of restrictive measures regarding the movement and control of Africans.

The preferential employment opportunities along with the restrictions on African movement resulted in rapid urbanisation of the white population. However, African urbanisation was also evident but was restricted to the minimum. Although the Urban Areas Act of 1923 prevented Africans from owning land in urban areas and restricted their movement, the application of the Act was not always efficient since some municipalities did not endorse the idea of segregation. This led to unequal outcomes in various places and consequently African urbanisation still occurred but at a minimal rate. This led to the amendment of the Act in 1937, which permitted Africans who were settled in urban areas to stay there. Consequently the period before and during WW II marked a relative relaxation of segregation policy in South Africa.

The outbreak of WW II was a stimulus for industrialisation which led to an increased demand for workers. This was a significant period for economic growth in the country and was considered the take-off stage in the country's economy. An important reason for this occurrence was that more Africans entered the modern sector due to the relaxation of segregationist policy to serve the war effort and enhance production.

After the War restrictions on the movement of Africans were again intensified. This resulted in large-scale opposition by Africans to the pass laws and Africans stated to organise mass uprisings for increased wages in mines. This created the idea that the white government at that time were incompetent and unable to resolve the problem. This led to the appointment of the Fagan Commission who was given the task of giving advice and reporting on the problem. The Fagan Report proposed that many of the restrictions on Africans should be mitigated; however, the report was only published on the eve of the 1948 elections. The election was already swaying towards a new and illogical regime of Stallardism. The whites wanted to keep the ruling for generations to come and the Africans were perceived as a threat to this ideal.

In 1948 when the Afrikaners' National Party (NP) came to power apartheid was implemented as the official government policy. The segregationist policies intensified and influx control along with the clearance of mining compounds and slums ensued. The NP contended that

South Africa could not be considered a single nation, but comprised four racial groups, namely Africans, Indians, whites and coloureds. Each race was compelled to live in areas according to their race. This created different racial regions in South Africa, most notably the Bantustans, and also segregated neighbourhoods within urban areas.

The perceived problem of African urbanisation, according to the NP, needed to be rectified. This led to the implementation of the Population Registration Act as well as the Group Areas Act in 1950. These Acts led to a prescribed racial composition for all residential areas, and later on played a role in the removal of large groups of people “not entitled” to stay in urban areas. The Bantustans were also created and were regarded as independent states. The aim of these homelands was to separate the urban white population from the rural non-whites in South Africa.

The ideological vision of apartheid appeared to have materialised and would have lasting spatial implications in the future of South Africa. Along with the restriction of the movement of Africans by means of influx control the government realised that it would be difficult to keep the African population away from urban areas without an incentive to keep them there. Accordingly decentralised policies were implemented during the span of apartheid to help keep the Africans in their homelands or at least keep their migration to employment areas temporary. Accordingly the decentralised industrial policies aimed to develop industries on the borders of the white urban areas and in the rural areas. The shifting of industries towards the population away from urban areas and not the other way around was unnatural at the time of the development of the country. These decentralised policies were not always effective and resulted in uneconomical outcomes. Development of these peripheral industries was troublesome and most of the time did not aid in the development of the rural Africans.

A large number of industrial decentralisation policies were implemented from 1954 till the end of apartheid. Although the apparent aim of these policies had been to develop the homelands and the periphery, it could be considered politically motivated. The homelands and decentralised areas remained impoverished and struggled to make a productive daily living. Racial injustice and the fear that Africans would flock to urban areas kept these policies intact, at least until the late 1980s. The ineffective application of these policies in conjunction with rising political, economic, social and international pressure eventually overcame the government.

The Urbanisation strategy of 1986 and the repeal of the pass laws was the first step in the disbandment of the apartheid system. Later on it became apparent that the racially-based system could not and would never work. In 1990 the unbanning of opposition political parties, including the ANC, as well as political prisoners such as Nelson Mandela marked the beginning of a new dispensation. Eventually in 1994 the first democratic election was held in the country.

As was to be expected the democratisation of the country resulted in large numbers of African people urbanising. However, the period from 1996 to 2001 marked a relative decline in the rate of urbanisation. The impact of apartheid was evident on the rural Africans as they struggled to escape from their disastrous economic situations. They could not locate to economic opportunities since they had become bound to their poverty in a number of ways. The new ANC government started to implement policies to rectify and reconstruct these under-developed rural areas. However, investment in these areas is costly and accordingly policies were ineffective in most instances due to lack of funding and inefficient mechanisms for implementation. In the last decade urban policies have, however, enjoyed greater optimism than their predecessors. The new policies have shifted their focus from reconstruction and development towards the growth of South African cities in order to stimulate inclusive growth and enhance spatial efficiency.

Industrial policy after apartheid has abandoned apartheid's industrial decentralisation instruments directed at forcing industries to the periphery. Industrial strategies have reflected Local Economic Development (LED) initiatives which entail the shift of authority to from the National to sub-national or even local municipalities. The LEDs lead to entrepreneurial cities that compete for mobile investments and accordingly use their economic characteristics to induce local economic growth and development. The LED planning strategy has meant that Spatial Development Initiatives (SDIs) and Industrial Development Zones (IDZs) have replaced the Bantustan industrial nodes. However, these strategies have not delivered the desired results and consequently many other industrial development instruments have emerged since the late 1990s. These policies have, however, also struggled as a result of a lack of coordination between the National, sub-national and local spheres of government. In addition the ties between the private and the public sector have not been strong enough to produce rapid transformation and development in specified areas.

Consequently, even after the democratisation of the country, the apartheid legacy has left lasting spatial implications that are costly and cannot be easily resolved. Furthermore, the following arguments can be made regarding the development of the country as a whole:

(i) The unnatural application of decentralisation policies implies that decentralisation occurred prematurely in South Africa due to ideological visions of separate development. Henderson (2003) as well as development economists such as Williamson (1965) and Shisido (1981) implies that decentralisation should occur when urbanisation and clustering of economic activity in large cities are sufficient enough to deliver spill-overs to the periphery. This would lead to greater regional integration, economic growth and development in the country.

(ii) The restriction of movement of Africans and decentralisation policies during apartheid removed the incentive for Africans to urbanise and consequently to join in the benefits of agglomeration. South Africa as a whole was accordingly deprived of greater economic growth and development since agglomeration was forced to a suboptimal level. Consequently, the country, with all groups included, may only now be entering an age of high mass consumption.

(iii) The forces of agglomeration were restricted during apartheid which led to urban areas losing economic efficiency. Furthermore, the decentralised industrial nodes were in many instances uneconomical and not very productive sites. However, the incentive of rural workers to move from less productive regions to more productive regions was impaired. Accordingly the lack of agglomerating forces was not strong enough to firstly create sustainable economic growth through flourishing industries and secondly the fact that these forces may not have been satisfied entirely in large cities resulted in a lack of natural dispersion or decentralisation which in turn meant that a strong interdependent core-periphery spatial system could not be established.

The impact that apartheid had on urbanisation, segregation and development in South Africa was immense. Although all racially suppressive policies were removed after democratisation an imprint on South Africa's urban system was made that cannot be easily removed. As stated by Martin (1999:70) history plays a major part in agglomeration and 'irrational' economic decisions can lead to a sub-optimal equilibrium spatial pattern of economic activity. Once this pattern has been established processes of cumulative causation grounded upon increasing returns lock in the effect of a large exogenous shock. The 'irrational'

economic decision in the form of apartheid left the socio-spatial geography of the countries' cities with a legacy that will most likely remain for many years to come. Although residential segregation, disconnected areas between races, peripheralisation of the African population as well as substantial distances between African households and their workplace had to a certain extent diminished after 1994 as a result of the forces of agglomeration inducing increased urbanisation, the impact of apartheid to this date on the distribution of economic activity is still evident (Bremner, 2000:87). During apartheid the incentive was removed from people to move from less-productive regions to more productive cities. Economic efficiency was accordingly reduced and this can be considered as a root of contemporary developmental issues in South Africa. Accordingly this dissertation is aimed at determining the spatial efficiency of the country in terms of the population of its settlements. The balance between forces of agglomeration and dispersion in the country may have been largely altered by its history. Chapter 4 will describe the data and analyse it in order to give an answer to the question.

Chapter 4: South Africa's city size distribution

4.1 Introduction

Chapter 2 argued that economic growth and development have a place-dependent character. Chapter 3 made the argument that initially first-nature geography determined settlement patterns in South Africa and subsequently the growth of these settlements into large towns and cities can be attributed to forces of agglomeration pulling population and economic activity together. Apartheid policies, however, aimed to influence the spatial distribution of people and economic activity, separating whites and blacks within urban agglomerations (suburbs and townships) and between the large towns and cities and the rural (homeland) periphery. This sets the scene for the question that this dissertation seeks to answer: what has been the balance between agglomeration and dispersion forces, and what does it say about the persistence of the population and economic activity in South Africa?

Although the geographical economics literature explains the sources and scope of agglomeration economies it does not usefully describe city-size distributions. The reason for this is that the literature generally revolves around questions regarding the formation, size and growth of individual cities and is less concerned with the interdependencies between them. In other words, geographical economics cannot determine the unequal distribution of population and economic activity in the components (metropolitan areas, large cities, towns, service centres) of a core-periphery urban system (Brakman *et al*., 2009:299).

However, the unequal distribution of population and economic activity is in practice characterised by a remarkable regularity in terms of patterns across space and also in terms of interaction between centres (Van Marrewijk, 2012:294). This empirical regularity is known as Zipf's law for cities. The first discovery of the power law dates back to Auerbach (1913) when he suggested the idea that the size distribution of urban centres in a country can be estimated by a Pareto distribution (Gabaix, 1999:739; Soo, 2005:239). Ever since Auerbach's suggestion many refinements have occurred most importantly that of George Zipf (1949). Accordingly the term "Zipf's law" was adopted to refer to the notion that city sizes are in accordance with a Pareto distribution (Soo, 2005:239).

Thus this chapter sets out to examine the spatial distribution of the population in South Africa over the period 1911 to 2011, using census data and estimating Zipf's law. The following section provides a brief overview of the literature explaining and testing Zipf's law. Section

4.3 presents the South African data. Section 4.4 reports the results of the estimation of Zipf's law and its interpretation. The final section is a summary and conclusion.

4.2 An explanation of Zipf's law

Zipf's law is a remarkable case of the rank-size rule (Brakman *et al*., 1999:185; Brakman *et al*., 2009:301). The rank-size rule determines that the size of the second largest city should be half the size of the largest, the third largest city should be a third the size of the largest and so forth (Gabaix & Ioannides, 2004:2344). The rule is believed to be relevant to all countries (Anderson & Ge, 2005:757).The rank-size rule fits Zipf's law when "the number of cities with populations greater than S is proportional to $1/S$ or S^{-1}" (Gabaix, 1999:739). In other words the law states that the Pareto exponent q will be equal or close to one indicating a negative linear relationship when regressing the logarithm of the rank given to a city (e.g. largest=1, second largest=2) on the logarithm of the population of that corresponding city (Naudé & Krugell, 2003:483; Anderson & Ge, 2005:757; Benguigui & Blumenfeld-Lieberthal, 2007:649). Zipf's law can be denoted by the following equation:

$$\ln Size_m = \ln(c) - q \ln Rank_m \quad (1)$$

where c is a constant, $Rank_m$ is the rank of city m, $Size_m$ is the size of a city m in population and q is the estimated coefficient or Pareto exponent that gives the slope of the supposedly log-linear relationship between the city rank and city size (Gabaix, 1999:740; Brakman *et al*., 2009:301). If q equals one then Zipf's law occurs and this means that the largest city by population is exactly m times as large as the m^{th} largest city.

In most cases, however, the exponents produced by the regression are not equal to one and consequently Zipf's law does not hold (World Bank, 2009:52). For example, Rosen and Resnick (1980) conducted a rank size regression for 44 countries. Their estimated coefficient differed from 0.81 to 1.96 across the countries. Furthermore, Soo (2005) estimated that the average coefficient for 75 countries was approximately 1.11 which is greater than the prediction of unity by Zipf's law.[95] This has led to many writers questioning the empirical validity of Zipf's law. The law accordingly enjoys scepticism from a number of authors and can only be considered a rough approximation (Ioannides & Overman, 2003:128). However, the significance of the power law lies in the fact that it provides a benchmark for analysis of

[95] Brakman *et al.* (2009:302) also states that for small samples the estimation is biased and inconsistent. In addition the predicted standard errors underestimate the true standard errors which in the past have led many researchers to wrong conclusions regarding the Pareto exponent.

city size distribution (Duranton, 2006:543). The reason for this is that it follows from Gibrat's law which suggests that cities grow in parallel (World Bank, 2009:52). According to Gabaix (1999:741) when it is assumed that cities follow homogeneous growth processes[96] (in other words Gibrat's law) then the city size distribution will conform to Zipf's law in steady state. However, and more importantly for this dissertation, this does not mean that policies cannot influence the size of a city and its economic performance. Good and bad policies can move cities up and down their country's urban hierarchy (measured in population) (World Bank, 2009:52). Furthermore, Córdoba (2008) emphasises that Zipf's law holds when all cities, regardless of size, have identical expected growth rates (in other words Gibrat's holds for the mean) and certain conditions are applied to the variance in city sizes.[97] Importantly, Gabaix (1999) also emphasises the fact that deviations to the Pareto exponent can be explained as a deviation in Gibrat's law (Gabaix, 1999:756; Duranton, 2006:543).

In brief, the paragraph above suggests that Zipf's law can be used as an effective benchmark for city-size distributions (Duranton, 2006:543). This allows one to derive certain probabilistic interpretations of the city size distributions (Gabaix, 1999:752).[98] For instance if q is smaller than one it means that cities' sizes are more even and if q equals zero all city sizes are similar (Brakman *et al*., 2009:301-302). This could also suggest a greater level of suburbanisation in large secondary cities when q is smaller than one (Soo, 2005:240). Another suggestion could be that there are too few secondary and small cities (Gabaix, 1999:756). In the case that q is larger than one then the largest city is more than m times as large as the m^{th} largest city which implies that the urban system is more clustered than the law predicts and a substantially larger primate city is evident (Brakman *et al*., 2009:301-302). In other words an outlier is presented in the form of the capital which has a bigger size than Zipf's law would predict. This, however, would not be as surprising since the primate city is a special object, driven by unique forces such as political factors (Gabaix, 1999:755-756).

Due to the fact that the growth of cities is a balancing act between forces of agglomeration and forces of dispersion, Zipf's law can also be usefully applied to analyse this balance and examine the persistence (and efficiency) of the distribution of people and economic activity (World Bank, 2009:52). Naudé and Krugell (2003:483) indicate that when cities are

96 For example they have a common mean and common variance in growth rate (Gabaix, 1999:741).

97 These conditions include the idea that the elasticity of substitution is constant between goods; the positive externalities are assumed to be constant across goods; or an intricate combination of preferences and technologies is used (Córdoba, 2008:179).

98 Gabaix (1999:752), however, notes that even if Zipf's law is confirmed by the Pareto exponent exactly, the rank-size rule will only be considered an approximation.

characterised by either constant returns to scale or external economies are balanced by the diseconomies of an urban system, then Zipf's law would hold. This implies that when q is larger than one, it may indicate that forces of agglomeration were greater in the largest cities in a country since a larger share of the population is concentrated in large urban areas where they benefit from increasing returns and positive externalities. However, in the case that the q is smaller than one, it may indicate that the large cities did not stimulate agglomerating forces enough or that urban diseconomies led to a more even distribution of population in settlements. In addition, Naudé and Krugell (2003:484), state that when q is smaller than one, cities offer urbanisation economies instead of localisation economies that are offered when q is larger than one. According to Brakman *et al*. (2009:40) localisation economies, also known as Marshall-Arrow-Romer (MAR) externalities, and urbanisation economics (also referred to as Jacobs externalities) are both location-specific spill-overs, in other words firms and industries have to locate near one another to benefit from these externalities. The distinction between the two can be made on the grounds that MAR externalities entails that knowledge spill-overs occur between firms of the same industry and with Jacobs externalities knowledge spill-overs happen between different industries. The implication that MAR externalities have for cities is that their growth is spurred on by specialisation, whereas in the case of Jacobs externalities growth of cities are driven by diversity (Brakman *et al*., 2009:40, 280).

A concrete visual example is provided by Gabaix (1999:739-740) for 135 American metropolitan areas. He orders the cities of the United States of America (USA) by population size. The largest city, New York, is given a rank of one, the second largest, Los Angeles, is given a rank of two, and so forth. After that he draws a graph with the logarithm of the rank of the cities on the y-axis and the log of population on the x-axis for each corresponding city. The result is shown in Figure 4.2.

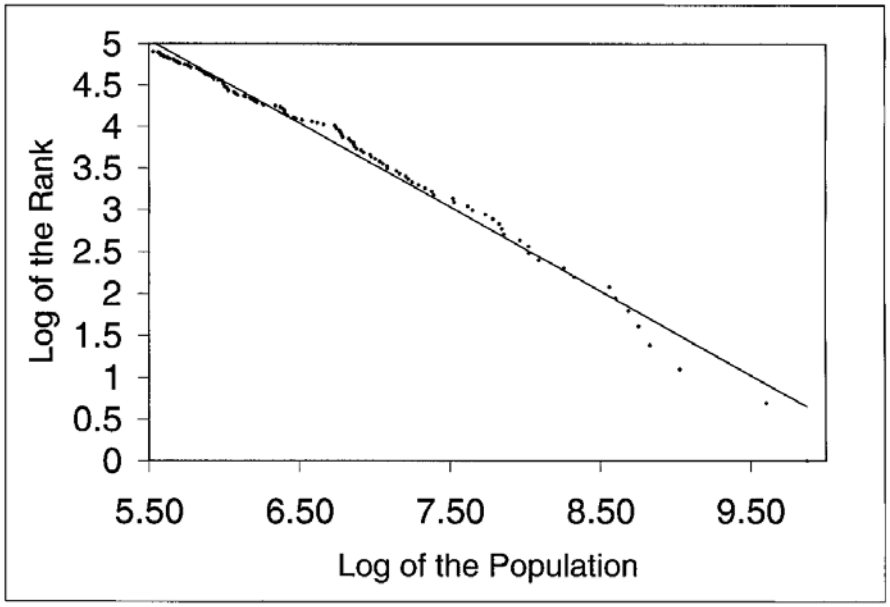

Figure 4.2: Log of population versus log of rank of the 135 US metropolitan areas in 1991 Source: Gabaix (1999).

A clear feature of Figure 4.2 is the fact that the plot is a near straight line. In addition the slope of this line is approximately minus one. This translates into the fact that Zipf's law is evident here since the size of a city is greater than S proportional to S^{-1}. In addition the strength of this result is supported by supported by the results of a simple regression model with a R^2 value of 0.986 and a standard deviation of 0.010 (Gabaix, 1999:739-740).[99]

Finally, Soo (2005) highlights the important fact that q is not constant over time. Brakman *et al*. (2009:303) state that when q is not stable over time it enables one to provide a theoretical explanation for a deviation from Zipf's law. According to Gabaix (1999:744) Zipf's law is a good approximation of the steady state distribution. He takes it even as far as to say that the law is the only steady state distribution (see Section III in Gabaix, 1999). Therefore, if cities grow stochastically and Gibrat's law is followed, in the limit the distribution will converge to Zipf's law. This view is also supported by the work done by Brakman *et al*. (1999). They estimated q for the Netherlands for 1600, 1900 and 1990. They came to the conclusion that industrialisation increased the value of the Pareto exponent as the city sizes started to increase (in other words the forces of agglomeration were strong) in the Netherlands since the value of

[99] However Gabaix (1999:740) declares that the regression is not fully appropriate. According to him Monte-Carlo simulations suggests that the true value of the exponent is underestimated by .05 on average, and also the standard deviation by more or less .1. However, even with the small corrections the Pareto exponent remains around 1.

q was 0.55 in 1600 and subsequently rose to 1.03 in 1900.[100] Importantly for all three periods the rank-size distribution held since at least 96 per cent of the variance in city sizes was explained by the rank-size distribution (Brakman *et al*., 1999:189).

4.3 South African data and the evolution of the urban system

To estimate the rank-size rule for South Africa and examine the balance between agglomeration and dispersion forces requires data of the size of the population per place. For this study it was gathered from the 1911 and 2011 censuses. The population data were transcribed from the published 1911 census and the 2011 data were gathered digitally from Statistics South Africa (2013).[101] The population data are at magisterial district level and hold benefits for the analysis. According to Rosen and Resnick (1980) Zipf's law holds better when agglomerations are used in the model. The use of individual cities in most cases leads to a Pareto exponent statistically different from 1 (Soo, 2005:240). The term city is accordingly an inaccurate label since metropolitan areas typically consist of many urban components (Gabaix, 1999:739). Therefore in this analysis the use of magisterial districts that encompass a greater range of urban components or agglomerations should deliver a better estimation.

Another advantageous factor contributing to the value of the analysis is the time-frame that is used. The censuses are 100 years apart and this long time period means that most of the historical forces contributing to the city size distribution in South Africa are taken into account. By 1911 most of the large contemporary urban agglomerations, like Cape Town, the PWV-area, Durban, East London, Bloemfontein and Port Elizabeth had already existed. This means that the urban system in the country had to a certain extent already moved away from the extensive margin of urban system evolution (the increase of the number of cities) towards the intensive margin (the increase of size of existing cities) (Córdoba, 2008:178).[102] As can be expected the city populations in 1911 were also substantially smaller and predate the rapid urbanisation of all groups in the country as well as the first racially biased policies. As mentioned in Chapter 3, no general law was established at that time regarding the right to ownership and occupation of property in urban areas. The Land Act was only ratified in

---

100 In 1990 the *q* value had decreased to 0.72 as a result of the Netherlands getting a more even city size distribution due to the movement towards a post-industrial era (Brakman *et al*., 1999).

101 The data were provided by Professor Waldo Krugell (supervisor of this study) at the North-West University, Potchefstroom Campus.

102 The latter case is more appicable since new cities are not commonly created in a well-developed urban system (Córdoba, 2008:178). Anderson and Ge (2005) refer to the intensive margin as the process of city size shrinkage or growth and the extensive margin as the city birth process.

1913. The 1911 census was also the first census after the establishment of the Union of South Africa and the only one that provided extensive coverage of the African population. Later censuses did not take into account the African populations in the Bantustans and segregated areas.

After 1911 the first racially-biased policies were implemented. Later apartheid policies were pursued and restrictions on the movement of Africans and decentralisation policies intensified. By 2011 apartheid and its segregationist policies had been abandoned. The analyses of the 1911 and 2011 censuses consequently enable one to derive strong conclusions regarding the impact that racial injustices, especially restrictive measures on urbanisation and decentralisation policies, had on the South African urban system.

A constraint of using the magisterial district-level data was that in 1911 there were fewer census districts than the 2011 census enumeration areas. Two options exist regarding this particular problem. Firstly, one can aggregate the 2011 census data to match the 1911 districts or secondly one can disaggregate the 1911 data to match the 2011 magisterial districts. In this dissertation the 2011 census data were aggregated to correspond with the 1911 boundaries. Disaggregation the other way around would require assumptions about the distribution of the population in 1911 that cannot be supported by relevant sources or data.

The use of the 1911 boundaries does, however, result in spatial units of analysis that require more careful interpretation. The largest census district in 1911 was Zoutpansberg but this encompassed a large area of the modern day Limpopo Province and cannot be considered a major urban agglomeration. Aggregating the 2011 data to this boundary thus results in a large area that has urban and rural characteristics. Interpretation of changes in the distribution has to take account of this fact that one is not simply observing towns or cities.

Table 4.2 shows that the smallest district with a population of 2834 was Port Nolloth. In 2011 Sutherland had the smallest population with 3921 people. The largest census district in 1911 was Zoutpansberg (336075) and in 2011 it changed to Pretoria (City of Tshwane Metropolitan Municipality) with a population of around 5.6 million people. The average size of magisterial districts increased from approximately 29030 in 1911 to 252424 in 2011.

Table 4.2: Description of smallest, largest and average population of magisterial districts in 1911 and 2011

	N	Minimum	Maximum	Mean	Std. Deviation
Total population in 1911	205	2834.00	336075.00	29029.9805	35516.33744
Total population in 2011	205	3921.00	5589846.00	252423.6683	590801.2171
Valid N (listwise)	205				

Source: Author's own calculations.

Table 4.2 indicates how the natural fertility of South Africa, immigration and urbanisation have risen in South Africa. The South African urban system undoubtedly evolved in terms of size. This is important since it will have a direct effect on the city-size distribution and consequently the test of Zipf's law.

4.4 An application of Zipf's law to 1911 and 2011

Using an Ordinary Least Squares (OLS) log-linear regression on the place-level data the following result for the estimation of Zipf's law was obtained for 1911:

(2) $\ln(S) = 13.318 - 0.783\ln(R)$ (SE = 0.020)

where S represents the size of the municipality in population and R represents the rank of the municipality. SE indicates the standard error of the estimation.

With R^2 = 0.879.

Thus the value of q exponent is 0.783 for South Africa for 1911. The R^2 is 0.879 and means that 88% of the variance in the population size of places is explained by the rank of the place in the distribution. This is not as high as other studies found in other countries. However, the value is high enough to interpret the Pareto exponent and apply Zipf's law as a benchmark for analysis. The q value of 0.783 suggests that the city-size distribution is more evenly spread and that the cities were of similar size. Furthermore, this suggests that these settlements offered urbanisation rather than localisation economies. This is not surprising since the country was still in its initial stages of industrialisation and forces of agglomeration were not as strong in settlements. One can argue that the unequal spread of first nature resources had a large impact on the fact that the rank-size distribution does not necessarily hold as strongly.

The abundance of minerals in different places within the interior and growth of transportation nodes at the coasts of the country resulted in people migrating to many different areas in the country and not only a few specific regions. However, it was evident from this early stage that first nature geography had already created a strong basis for the development of contemporary large urban areas in South Africa. The data suggest that cities such as Johannesburg, Cape Town, Pretoria, Durban, Germiston, Bloemfontein, Harrismith, Port Elizabeth, East London and Kimberley were already in 1911 substantially-sized settlements in South Africa.

However, the unequal distribution of population in South Africa and the evenness of the spread were evident in the fact that 101 of the 205 settlements had a population of less than 20000. Furthermore, 83 settlements had a population of 20000 to 50000. Only 21 settlements had a population of 50000 or more. Table 4.3 indicates the 21 largest settlements in South Africa in 1911. As noted, the primate city has a substantial role to play in the Pareto exponent. The primacy ratio can be used to determine the relative size of the primate city. To calculate the primacy ratio one takes the sum of the population of the five largest cities in a country to determine the ratio of the size of the largest city (Brakman *et al*., 2009:304). According to Brakman *et al*. (2009:304) in most countries the primacy ratio is greater than 50%. The primacy ratio in 1911 for the Zoutpansberg was 32.4%. This also gives an indication why the Pareto exponent is lower than one. The value of the exponent was decreased by the equal distribution of population in the five largest places and the relatively small nature of the primate district.

Table 4.3: 21 largest agglomerations with 50000 or more people in 1911

Zoutpansberg	336075
Johannesburg	241131
Cape	187831
Pretoria	157444
Lydenburg	114757
Kingwilliamstown	106474
Krugersdorp	103473

Boksburg	86922
Umlazi	75856
Waterberg	73901
Durban	72512
Potchefstroom	69360
Rustenburg	64902
Germiston	64805
Kimberley	64352
Middelburg	64251
Engcobo	62055
Inanda	59222
Bloemfontein	58451
Alexandra	52578
Glen Grey	50597

Source: Author's own calculations.

Figure 4.3 plots the log of population on the log of rank for the 205 census districts in 1911. The red line in the graph indicates the slope that would be evident if Zipf's law held in this specific case. The blue line indicates the actual gradient of the distribution. As can be seen the slope of the blue line is flatter than the red line due to the fact that the Pareto exponent is lower than one. However, there are a number of conclusions that can be drawn regarding the graph using the red line (Zipf's law) as point of reference. The upper-end of the graph firstly indicates that a substantial number of large places are below the predicted Zipf's law size. This suggests that too many large agglomerations of similar size were evident at that stage. Another conclusion that can be derived is that many small settlements existed as indicated by the lower-end of the graph. A distinct curve arises at the lower end of the graph that is higher than the Zipf's law line. This is an indication of too many small settlements which are larger than predicted by Zipf's law and also do not meet the rank-size rule. The suggestion can be made that the Pareto exponent was largely dragged below minus one by the similar size of the

largest places. However, the exponent's value was pushed up a bit by the large number of small settlements that were larger than Zipf's law would predict. However it is still safe to say that South Africa had a more even distribution of city sizes in 1911.

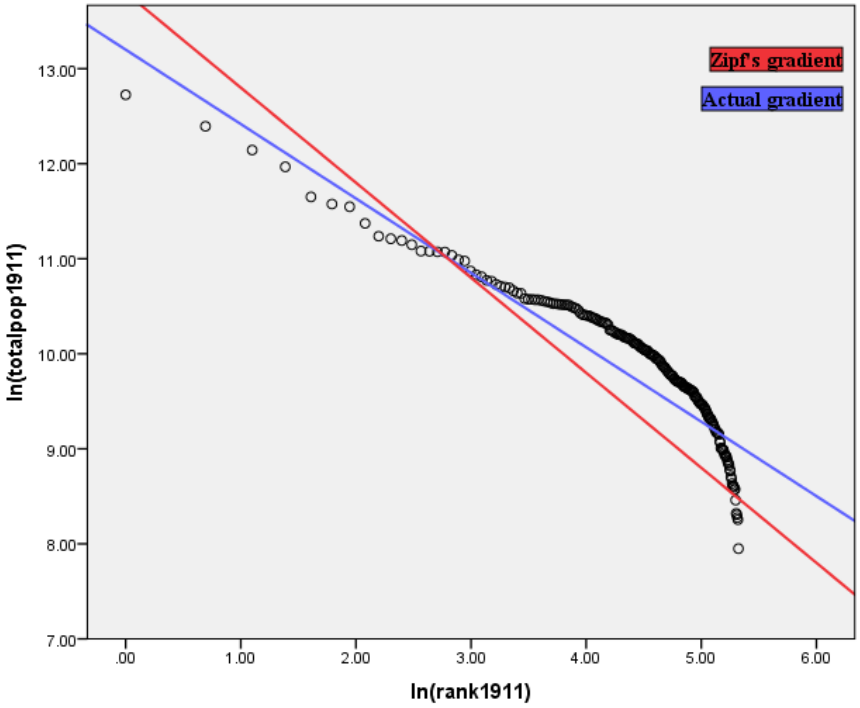

Figure 4.3: Logarithm of size versus logarithm of 205 magisterial districts in 1911 in South Africa

The results of the regression model for the 2011 magisterial districts were as follows:

(3) $\ln(S) = 17.280 - 1.356\ln(R)$ (SE = 0.039)

With R^2 = 0.859.

The result of the OLS regression suggests that the value of the q exponent is minus 1.356. This suggests that apartheid era spatial policies did not impair the forces of agglomeration in South Africa. Furthermore, the large cities are larger than Zipf's law predicts and there is a larger concentration of urban agglomeration. However, an argument can be made that the Pareto exponent is misleading due to the fact that the rank size distribution does not necessarily hold as strongly since only 86% of the variance in the size of cities is explained by the rank size distribution. This suggests that one needs to delve deeper into the nature of the city size distribution in order to find an explanation for the large value of q.

A first step that we can take is to look at the primacy ratio evident in 2011 according to the data. The reason for this is to establish whether the primate city is the driver of the high value of the Pareto exponent. Using the data for the five largest magisterial districts in 2011 it was found that the magisterial district of the City of Tshwane Metropolitan Municipality (Pretoria), which is the primate city, had a primacy ratio of 33.4%. This indicates that the primate city in the country is too small and that the high q exponent may in fact be driven up by the small primate city. As mentioned before, the rank size rule states that the second largest city should be half the size of the largest city, the third largest city should be a third the size of the largest and so forth. If we apply this rule the second largest magisterial district, the Zoutpansberg District, should have a population of 2794923. However the population of the Zoutpansberg is actually 3636653 which is much larger than the rank size distribution would predict. The third largest magisterial district, Krugersdorp, should have a population of 931641. However, it grossly exceeds this value with a population of 3170252. Table 4.4 indicates the discrepancy between the predicted populations of the largest magisterial districts and their actual population, using the primate city as reference in the rank size distribution.

Table 4.4: Discrepancy between actual population and the predicted population if rank size rule is applied to 2011

Magisterial district	Actual population	Predicted population according to the rank size rule
Pretoria	5589846	Primate city as reference
Zoutpansberg	3636653	2794923
Krugersdorp	3170252	1863282
Cape Town	2757961	1397462
Umlazi	1572554	1117969
Potchefstroom	1524961	931641
Heidelberg	1460243	798549
Lydenburg	1241470	698731
Inanda	1223558	621094
Johannesburg	1203100	558985

Source: Author's own calculations.

Figure 4.4 also demonstrates the large number of secondary cities that are similar of size and are above the slope of Zipf's law if it were to hold for South Africa (look at the curve in the middle-end of the graph). The graph also indicates that the size of the first and second magisterial districts is in fact smaller than Zipf's predicted size. An interesting feature of the graph is that the middle and lower ends of the graph are more in line with Zipf's law's predicted population indicated by the red line. This may suggest that South Africa is in fact

converging to Zipf's law in the cases of the medium and smaller-sized towns. The large number of smaller settlements (indicated by the downward curve in the lower-end of the graph) is below the red line of Zipf's. This indicates that these settlements are smaller than Zipf's law predicts. These settlements may have lowered the value of the Pareto exponent; however, it was insignificant in lowering the value of q substantially.

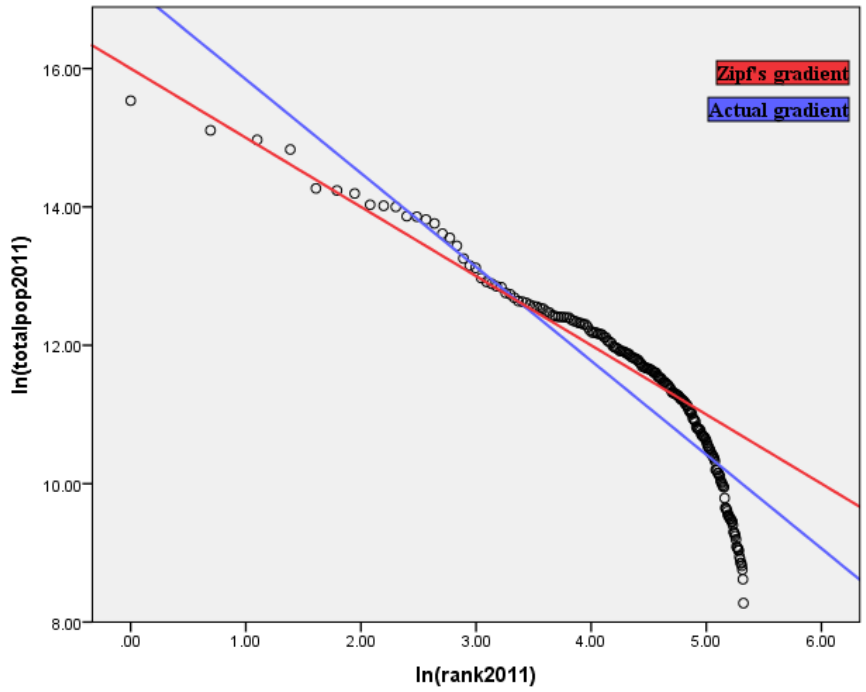

Figure 4.4: Logarithm of size versus logarithm of rank of 205 magisterial districts in 2011 in South Africa

Another test about whether this large q is not only a post-1994 phenomenon would be to perform a similar estimate with data from the 1996 population census. Krugell (2014:13) finds q = 1.25. Agglomeration was not only an apartheid-era phenomenon, but has since continued to the point of over-concentration.

4.5 Conclusion

The conclusion that can be drawn from the test of Zipf's law in South Africa over the period 1911 to 2011 is that urban agglomeration did in fact happen in spite of the efforts of Apartheid policy-makers. However, one needs to carefully consider the nature of this agglomeration since it gives an indication of the impact that restrictive urbanisation policies and industrial decentralisation policies had on the country. The primate and second largest magisterial districts are smaller than Zipf's law predicts. Furthermore, a large number of secondary cities are larger than Zipf's law predicts and do not suffice in the rank size rule.

The conclusion that can be drawn from this is that growth of the primate and second largest districts was inhibited between 1911 and 2011 and that their economic performance was affected; however, it also had an impact on a large number of secondary cities. Apartheid spread out urbanisation and increased the number of secondary cities in the country. This raised the urban agglomeration in the secondary districts which in turn raised the value of the Pareto exponent in 2011. In addition the large number of smaller settlements with smaller populations than those predicted by Zipf's law may suggest that either the rapid urbanisation in conjunction with the fact that industrial development nodes changed after apartheid stimulated the decrease in size of the rural areas in terms of population.

One also has to keep in mind that magisterial districts cover larger areas. For instance, the districts encompass the townships formed outside large urban centres as a result of the urban migrant labour created by the influx control and other restrictive measures. The decentralised industries formed by industrial decentralisation policies are also taken into account by the large magisterial districts. Consequently it can be said that natural magisterial growth for secondary districts actually increased since continued investment before and after apartheid in the periphery of urban centres created a disincentive for the population to move to the metropolitan agglomerations.

Another component that may have influenced the development of large secondary areas in South Africa may be geographical in nature. If the uneven distribution of settlements during 1911 is used as reference point then it can be argued that first nature aspects also had a significant influence of modern day uneven distribution of population and wealth. The large number of secondary urban agglomerations and their growth can be regarded a result of first nature aspects and it can be said that apartheid intensified their growth and formation considering that the majority of the population (Africans) could not afford the move to primate cities. Consequently, apartheid did have an effect on agglomeration in South Africa. However, it did shift the equilibrium spatial distribution of the country to another level. The restrictive measures on urbanisation and industrial decentralisation policies resulted in the spreading of population and economic activity throughout the country. The consequence is that Zipf's law does not hold and that the rank-size distribution of the country was altered. However, it can be seen that a convergence towards Zipf's law is occurring regarding medium and small-sized towns. In other words, the country may in fact be moving to a new steady state after apartheid. The question arises whether this occurrence will spill over to the largest cities in the country? The answer for this question can only be speculated upon. Only

time will tell if South Africa will move to Zipf's in the future. Martin (1999) stated that 'irrational' economic policies can alter the equilibrium spatial pattern in a country to a sub-optimal level. This observation is directly applicable to South Africa and it may take many years before we reach a new equilibrium spatial distribution of population and economic activity.

Chapter 5: Summary, conclusion and recommendations

5.1 Summary

The dissertation set out to determine the spatial persistence of agglomeration in South Africa and the impact that apartheid, especially restrictive measures on urbanisation and industrial decentralisation policies, had on the distribution of the population in the country. Firstly, literature regarding economic geography was discussed in order to get a better understanding of agglomeration and the nature of this phenomenon. The discussion of the historical development of economic geography explained how geographical economics or the new economic geography came to being as a discipline. The essence of the issue of geographical economics is to explain the structure of economic agglomeration across space. In other words it strives to answer the question of who produces what, why and where.

According to geographical economics the location of economic activity has to be understood as the result of the interaction between forces of agglomeration and dispersion within an urban system. Furthermore, a strong positive relationship exists between agglomeration and economic growth and this happens in conjunction with urbanisation. However, state policy has a role to play since agglomeration and growth are strongly related to economic development, urbanisation and migration and consequently the growth and economic performance of cities.

The historical development of South Africa was subsequently discussed and the geographical, economic and political factors that might have influenced the spatial distribution of population and economic activity were taken into account. First nature geography played a large part in the uneven distribution of population and wealth initially. However, the spatial distribution of the country was also significantly influenced by policies that aimed to deconcentrate and decentralise a large proportion of its population. The implementation of apartheid, most importantly restrictive measures against the movement of Africans and industrial decentralisation, resulted in the forces of agglomeration in the country being inhibited since an incentive was removed from people to move from less productive to more productive areas. Accordingly, apartheid led to unexploited economies of scale in large urban centres which in turn led to substantial productivity losses within the economy. Geographical economics is in agreement with the view that changes in the level of population of a city will alter the city's productivity and consequently its growth. Accordingly large temporary shocks

like apartheid can have a lasting impact on the distribution of wealth and population of a country.

Literature of geographical economics provides a comprehensive understanding of the sources and scope of agglomeration economies. However, it cannot be effectively applied to explain the city size distributions since the field does not revolve around the question of how urban systems are interconnected. Zipf's law for cities and the rank size distribution can effectively be applied to derive probabilistic explanations regarding the distribution of wealth and population within a country.

In the analysis census data were used regarding the population of 205 magisterial districts in 1911 and 2011 in South Africa. The time frame for analysis allows for an effective comparison of the results using Zipf's law for 1911 and 2011. The reason for this is that 1911 predates intensive segregationist policies in the country and 2011 falls within a new democratic dispensation within the country. Accordingly the large temporary shock in the form of apartheid could be isolated and the impact of restrictive measures on urbanisation and industrial decentralisation policies could be analysed. The results gathered from the ordinary least squares log-linear regression of the log of population on the log of the rank of the country indicate that the Pareto exponent obtained is minus 0.783 for 1911 and minus 1.356 for 2011. Accordingly Zipf's law does not hold for South Africa in either of these periods. However, the exponent indicates that a more even city-size distribution is evident in 1911 and that in 2011 more urban agglomeration is evident in the country. This might suggest that agglomeration in the country was not affected by apartheid policies. However, if one investigates the reason for the high value of the Pareto exponent further, one can come to a conclusion regarding the impact that apartheid had on agglomeration in the country. The primate and second largest magisterial districts are substantially smaller than Zipf's law predicts. Furthermore, a large number of secondary cities are larger than Zipf's law predicts and accordingly this raised the Pareto exponent. Too many small settlements exist that are smaller than Zipf's law predicts. The implication of this is that apartheid altered the equilibrium spatial distribution of the population by inhibiting the growth and economic performance of the primate and second largest cities. Furthermore, a large number of secondary cities were created that are similar in size, suggesting that the restrictive measures on urbanisation and industrial decentralisation policies spread out the population and economic activity of the country. In other words, the disincentive created by apartheid policies created a sub-optimal equilibrium spatial distribution in the country.

5.2 Conclusion

The aim of the study was to determine the spatial persistence of South Africa and to determine the impact that had on the distribution of the population in the country. The results clearly showed that agglomeration prevailed over Apartheid dispersion, but at the cost of a sub-optimal spatial distribution of population today. On the one hand, the few large metropolitan cities may be too large, imposing costs of congestion and diseconomies of scale. On the other hand, there are too many small settlements that act as rural poverty traps. The conclusion is that policy can influence economic geography and this has important implications for current industrialisation and rural development policies. Policies like the National Development Plan (NDP) would do well to heed the World Bank's (2009: xxi) warning:

> "…economic growth will be unbalanced. To try to spread out economic activity is to discourage it. But development can still be inclusive, in that even people who start their lives far away from economic opportunity can benefit from the growing concentration of wealth in a few places. The way to get both the benefits of uneven growth and inclusive development is through economic integration".

5.3 Recommendations

A number of recommendations for future research can be made. On the technical side of the empirical analysis it may be possible to pursue refinements in the form of different estimators of the regression model. In the analysis a simple Ordinary Least Squares (OLS) log-linear regression was used. However, some authors like Soo (2005) suggest that the Hill-estimator is a better method in estimation. Soo (2005) used both estimation methods in his analysis of Zipf's law on 73 countries. He found that in using OLS he rejected Zipf's law for 53 out of the 73 countries. However, when using the Hill-estimator he rejected Zipf's law 30 times out of the 73 countries. Although the Hill (maximum likelihood) estimator has its problems (see Gabaix and Ioannides, 2004:2349-2350) it may deliver more conclusive results than OLS regarding the data. Other methods for estimation also exist and can be applied such as the Gabaix-Ibragimov method and the Kratz-Resnick standard errors.

It may also be possible to try and define true agglomerations, instead of using administrative units like the magisterial districts. According to Brakman *et al*. (2009:303) when Zipf's law is compared in studies in a different period the choice of sample size is also very important.

They state that two approaches can be followed in this regard: firstly, the use of a fixed number of cities or secondly a threshold level can be used where cities are not included if their populations are too small. Accordingly, the number of magisterial districts used in the analysis can be decreased from 205 to, say, the 100 largest magisterial districts.

As mentioned in the study the R^2 was in both instances not as high. The goodness of the fit of the rank size distribution was not good and consequently the results may be biased to a certain extent. A better construction of the agglomerations might yield better results and consequently more conclusive arguments can be made.

It is, however, essential that researchers continue to study the rank size of cities or districts in the country. It may in the future indicate whether South Africa has moved to a new steady state regarding the equilibrium spatial distribution of population and economic activity. This can be used to advise policy-makers regarding the spatial efficiency of their policies (for example industrial policies). The evaluation of these policies and their effectiveness will indicate which steps need to be taken next by plannners to ensure the sustainabilty of spatial growth in South Africa. Contemporarily a fault in policy lies in the fact that there is a lack of coordination and integration between the national, sub-national and local levels of government. If this problem can be rectified the predominant effect of inter-regional integration can increase economic efficiency within the geographical space in the country (Glaeser, 2007).

Bibliography

Abedian, I. & Standish, B. 1985. Poor whites and the role of the state: The evidence. *South African Journal of Economics*, 53(2):93-108.

Acemoglu, D., Johnson, S. & Robinson, J.A. 2001. The colonial origins of comparative development: An empirical investigation. *American Economic Review*, 91(1):1369-1401.

Albion, R. 1939. The rise of New York port, 1815-1860. New York: Charles Scribner's Sons.

Allen, R.A. 2006. The British industrial revolution in global perspective: how commerce created the industrial revolution and modern economic growth. Oxford: Oxford University Press.

Alonso, W. 1964. Location and land use. Cambridge, MA: Harvard University Press.

Anas, A., Arnott, R. & Small, K.A. 1998. Urban spatial structure. *Journal of Economic Literature*, 36(1):1426-1464.

Anderson, G. & Ge, Y. 2005. The size distribution of Chinese cities. *Regional Science and Urban Economics*, 35(1):756-776.

Aoyama, Y., Murphy, J.T. & Hanson, S. 2011. New concepts in economic geography. London: Sage Publications Ltd.

Arthur, W.B. 1994a. Urban systems and historical path dependence. (*In* Arthur W.B., *ed*. Increasing Returns and Path Dependence in the Economy. Michigan: Michigan University Press. pp. 99-110).

Arthur, W.B., *ed*. 1994b. Increasing returns and path dependence in the economy. Michigan: Michigan University Press.

Audretsch, D.B. & Feldman, M.P. 1996. R&D spillovers and the geography of innovation and production. *American Economic Review*, 86(3):630-640.

Auerbach, F. 1913. Das gesetz der bevölkerungskonzentration. *Petermanns Geographische Mitteilungen*, 59(1):73-76.

Au, C.C. & Henderson, J.V. 2006. How migration restrictions limit agglomeration and productivity in China. *Journal of Development Economics*, 80(2):350-388.

Bacolod, M., Blum, B.S. & Strange, W.C. 2010. Elements of skill: Traits, intelligences, education, and agglomeration. *Journal of Regional Science*, 50(1):45-80.

Baines, G.F. 1990. The origins of urban segregation: Local government and the residence of Africans in Port Elizabeth c. 1835-1865. *South African Historical Journal*, 1(1):70-76.

Baines, G. 1995. The politics of welfare: The provision of housing and services in New Brighton, Port Elizabeth. *Kronos*, 22(1):87-114.

Baldwin, R.E. 1999. Agglomeration and endogenous capital. *European Economic Review*, 43(2):253-280.

Baldwin, R.E. & Forslid, R. 2000. The core-periphery model and endogenous growth: Stabilizing and destabilizing integration. *Economica*, 67(1):7-24.

Baldwin, R., Forslid, R., Martin, P.O.G. & Robert-Nicoud, F. 2003. Economic Geography and Public Policy. Princeton, NJ: Princeton University Press.

Baldwin, R. & Krugman, P.R. 2000. Agglomeration, integration and tax harmonization. Discussion paper no. 2630. London: Centre for Economic Policy Research.

Baldwin, R. & Martin, P. 2003. Agglomeration and regional growth. Discussion paper no. 3960. London: Centre for Economic Policy Research.

Baldwin, R.E., Martin, P. & Ottaviano, G.I.P. 2001. Global income divergence, trade and industrialisation: The geography of growth take-offs. *Journal of Economic Growth*, 6(1):5-37.

Barnes, T.J. 2000. Inventing Anglo-American economic geography, 1889-1960. (*In* Sheppard, E. & Barnes, T.J., *eds*. A companion to economic geography. Blackwell, Oxford: Blackwell Publishing Ltd. pp. 11-25).

Barnes, T.J. 2001. On Theory, history and anoraks. *Antipode*, 33(1):162-168.

Barnes, T.J. 2011a. Spatial analysis. (*In* Agnew, J. & Livingstone, D., *eds*. The Sage handbook of geographical knowledge. London: Sage. pp. 380-391).

Barnes, T.J. 2011b. The quantitative revolution and economic geography. (*In* Leyshon, A., Lee, R., McDowell L. & Sunley, P., *eds*. The Sage handbook of economic geography. London: Sage. pp. 39-52).

Barnes, T.J. 2012. Reopke Lecture in Economic Geography: Notes from the underground: Why the history of economic geography matters: The case of central place theory. *Economic Geography*, 88(1):1-26.

Barro, R. & Sala-i-Martin, X. 1991. Convergence across states and regions. *Brookings Papers on Economic Activity*, 1(1):107-158.

Becker, R. & Henderson, J.V. 2000. Intra-industry specialization and urban development. (*In* Huriot J.M. & Thisse, J.-F., *eds*. The Economics of Cities: Theoretical Perspectives. Cambridge: Cambridge University Press. pp. 138-166).

Bell, T. 1973. Industrial decentralisation in South Africa. Cape Town; London: Oxford University Press.

Bell, T. 1987. International competition and industrial decentralisation in South Africa. *World Development*, 15(10-11):1291-1307.

Benguigui, L. & Blumenfeld-Lieberthal, E. 2007. Beyond the power law - a new approach to analyze city size distributions. *Computers, Environment and Urban Systems*, 31(1):648-666.

Berry, B.J.L. & Garrison, W.L. 1958a. Alternate explanations of urban rank-size relations. *Annals of the Association of American Geographers*, 48(1):83-91.

Berry, B.J.L. & Garrison, W.L. 1958b. The functional bases of the Central Place Hierarchy. *Economic Geography*, 34(2):145-154.

Berry, B.J.L. & Horton, F.E. 1969. Emergence of urban geography: methodological foundations of the discipline. (*In* Berry B.J.L. & Horton F.E., *eds*. Geographical Perspectives on Urban Systems. New York: Prentice Hall. pp. 1-19).

Bertola, G. 1993. Models of economic intergration and localised growth. (*In* Torres F. & Giavassi, F., *eds*. Adjustment and Growth in European Monetary Union. Cambridge: Cambridge University Press. pp. 159-192).

Bhorat, H., Hirsch, A., Kanbur, R. & Ncube, M. 2014. Economic policy in South Africa: Past, present and future. Development Policy Research Unit Working Paper 201401. University of Cape Town.

Black, D. & Henderson, J.V. 2003. Urban evolution in the USA. *Journal of Economic Geography*, 3(1):343-372.

Bond, P. 2003. The degeneration of urban policy after apartheid. (*In* Harrison, P., Huchzermeyer M. & Mayekiso, M., *eds*. Confronting fragmentation: housing and urban development in a democratising society. Cape Town: University of Cape Town Press. pp. 40-56).

Booysen, J.J. 1985. Bevolkingsverhoudinge binne Suid-Afrika. (*In* Van Jaarsveld, F.A., *ed*. Verstedeliking in Suid-Afrika. Pretoria: Universiteit van Pretoria. pp. 15-27).

Boraine, A., Crankshaw, O., Engelbrecht, C., Gotz, G., Mbanga, S., Narsoo, M. & Parnell, S. 2006. The state of South African cities a decade after democracy. *Urban Studies*, 43(2):259-284.

Bos, D.J. 1992. Ruimtelike implikasies van nywerheidsontwikkeling. 78th ed. Wetenskaplike bydrae van die P.U. vir C.H.O.

Boschma, R.A. & Frenken, K. 2006. Why is economic geography not an evolutionary science? Towards an evolutionary economic geography. *Journal of Economic Geography*, 6(1):273-302.

Brakman, S., Garretsen, H. & Schramm, M. 2004. The strategic bombing of German cities during WWII and its impact on city growth. *Journal of Economic Geography*, 4(1):20-218.

Brakman, S., Garretsen, H. & Van Marrewijk, C. 2009. The new introduction to geographical economics. 2nd ed. New York: Cambridge university press.

Brakman, S., Garretsen, H., Van Marrewijk, C. & Van den Berg, S. 1999. The return of Zipf: towards a further understanding of the rank-size curve. *Journal of Regional Science*, 39(1):183-215.

Bremner, L.J. 2000. Post-apartheid urban geography: A case study of greater Johannesburg's rapid land development programme. *Development Southern Africa*, 17(1):87-104.

Browett, J.G. & Fair, T.J.D. 1974. South Africa, 1870-1970: A view of the spatial system. *The South African Geographical Journal*, 56(2):111-120.

Bundy, C. 1979. The rise and fall of the South African peasantry. London: Heineman.

Camagni, R.P. 1995. The concept of innovative milieu and its relevance for lagging regions in the 1990s. *Papers in Regional Science*, 74(1):317-340.

Chan, K.W. 1994. Cities with invisible walls. Hong Kong: Oxford University Press.

Chisholm, G.G. 1889. Handbook of commercial geography. London: Longman, Green & Co.

Choe, C. & Chrite, E.L. 2014. Internal migration of blacks in South Africa: An application of the Roy model. *South African Journal of Economics*, 81(1):81-98.

Christaller, W. 1933. Die Zentralen Orte in Süddeutschland. Jena: Gustav Fiscger Verlag.

Cobbett, W., Glaser, D., Hindson, D. & Swilling, M. 1985. South Africa's regional political economy: A critical analysis of reform strategy in the 1980s. Seminar Paper. Johannesburg: University of the Witwatersrand African Studies Institute.

Coetzee, S.F. & De Coning, C. 1992. An agenda for urban research: South Africa in the 1990s. Paper presented at the Workshop on Urban Research Agenda for Southern Africa, June 1992. Johannesburg: University of the Witwatersrand.

Collinson, M.A., Tollman, S.M. & Kahn, K. 2007. Migration, settlement change and health in post-apartheid South Africa: Triangulating health and demographic surveillance with national census data. *Scandinavian Journal of Public Health*, 35(69):77-84.

Córdoba, J.-C. 2008. On the distribution of city sizes. *Journal of Urban Economics*, 63(1):177-197.

Crankshaw, O. 2012. Deindustrialization, professionalization and racial inequality in Cape Town. *Urban Affairs Review*, 48(6):836-862.

Crush, J. & Rogerson, C.M. 2001. New industrial spaces: Evaluating South Africa's spatial development initiatives (SDI) programme. *South African Geographical Journal*, 83(2):85-92.

Cypher, J.M. & Dietz, J.L. 2009. The process of economic development. 3rd ed. New York, NY: Routledge.

Davelaar, E.J. & Nijkamp, P. 1997. Spatial dispersion of technological innovation: A review. (*In* Bertuglia, C.S., Lombardo, S. & Nijkamp, P., *eds*. Innovative Behaviour in Space and Time. Berlin: Advances in Spatial Science, Springer-Verlag. pp. 17-40).

Davenport, R. 1991. Historical background of the apartheid city to 1948. (*In* Swilling, M., Humphries P. & Shubane, K., *eds*. Apartheid city in transition. Cape Town: Oxford University Press. pp. 1-18).

Davis, J. & Henderson, J.V. 2003. Evidence on the political economy of the urbanisation process. *Journal of Urban Economics*, 53(1):98-125.

Davis, D.R. & Weinstein, D.E. 2002. Bones, bombs and breakpoints: The geography of economic activty. *American Economic Review*, 92(1):1269-1289.

Department of Cooperative Governance and Traditional Affairs **see** South Africa. Department of Cooperative Governance and Traditional Affairs.

Dewar, D., Todes, A. & Watson, V. 1986. Industrial decentralization policy in South Africa: Rhetoric and practice. *Urban Studies*, 23(1):363-376.

Diamond, J. 1997. Guns, germs, and steel: The fates of human societies. New York: W.W. Norton.

Dixit, A. & Stiglitz, J.E. 1977. Monopolistic competition and optimal product diversity. *American Economic Review*, 67(1):297-308.

DTI **see** South Africa. Department of Trade and Industry.

Duranton, G. 2006. Some foundations for Zipf's law: Product proliferation and local spillovers. *Regional Science and Urban Economics*, 36(1):542-563.

Duranton, G. & Storper, M. 2006. Agglomeration and growth: A dialogue between economicsts and geographers. *Journal of Economic Geography*, 6(1):1-7.

Eaton, J. & Kortum, S. 1996. Trade in ideas patenting and productivity in the OECD. *Journal of International Economics*, 40(1):251-278.

Ellison, G. & Glaeser, E.L. 1997. Geographic concentration in US manufacturing industries: A dartboard approach. *Journal of Political Economy*, 105(1):889-927.

Ellison, G. & Glaeser, E.L. 1999. The geographic concentration of industry: Does natural advantage explain agglomeration? *American Economic Review, Papers and Proceedings*, 89(1):311-316.

Fair, T.J.D. 1982. *South Africa:* Spatial Frameworks for Development. 93rd ed. Johannesburg: Juta.

Fair, T.D.J. & Browett, J.G. 1979. The urbanisation process in South Africa. (*In* Herbert D. & Johnson, R.J., *eds*. Geography and the urban environment. New York: John Wiley & Sons. pp. 259-294).

Fay, M. & Opal, C. 1999. Urbanisation without growth: Understanding an African phenomenon. Washington, DC: The World Bank.

Fellmann, J.D. 1986. Myth and reality in the origin of American Economic Geography. *Annals of the Association of American Geographers,* 76(3):313-330.

Fields, G.S. 2004. Dualism in the labor market: A perspective on the Lewis model after half a century. *The Manchester School,* 72(6):24-35.

Fine, B. & Rustomjee, Z. 1996. The political economy of South Africa: from minerals-energy complex to industrialisation. London: C. Hurst & Co. (Publishers) Ltd.

Floyd, T.B. 1960. Town planning in South Africa: History, present and future. Pietermaritzburg: Shuter & Shooter.

Forslid, R. 1999. Agglomeration with human and physical capital: An analytically solvable case. Discussion paper no. 2102. London: Centre for Economic Policy Research.

Fujita, M. & Krugman, P.R. 1995. When is the economy monocentric? Von Thünen and Chamberlain unified. *Regional Science and Urban Economics,* 25(1):8-28.

Fujita, M. & Krugman, P. 2004. The new economic geography: Past, present and the future. *Papers in Regional Science,* 83(1):139-164.

Fujita, M., Krugman, P.R. & Mori, T. 1999a. On the evolution of hierarchical urban systems. *European Economic Review,* 43(2):209-251.

Fujita, M., Krugman, P.R. & Venables, A.J. 1999b. The spatial economy: Cities, regions, and international trade. Cambridge, MA: MIT Press.

Fujita, M. & Mori, T. 1996. The role of ports in the making of major cities: Self-agglomeration and hub-effect. *Journal of Development Economics,* 49(1):93-120.

Fujita, M. & Thisse, J.-F. 1996. Economics of agglomeration. *Journal of the Japanese and International Economics*, 10(1):339-378.

Fujita, M. & Thisse, J.-F. 2002. Economics of agglomeration: Cities, regions, and international trade. Cambridge, MA: MIT Press.

Gabaix, X., 1999. Zipf's law for cities: An explanation. *The Quarterly Journal of Economics*, 114(3):739-767.

Gabaix, X. & Ioannides, Y.M. 2004. The evolution of city size distributions. (*In* Henderson, J.V. & Thisse, J.-F. *eds*. The handbook of urban and regional economics. Amsterdam: North-Holland. pp. 2341-80).

Gallup, J.L., Sachs, J.D. & Mellinger, A.D. 1998. Geography and economic development. Working paper no. 6849. Cambridge, MA: National Bureau of Economic Research.

Gelderblom, D. & Kok, P. 1994. Urbanisation South Africa's Challenge. 1st ed. Pretoria: HSRC Publishers.

Giliomee, H. 1979. The Afrikaner economic advance. (*In* Adam, H. & Giliomee, H., *eds*. The rise and crisis of Afrikaner power. Cape Town: David Phillip. pp. 145-176).

Glaesar, E.L. 2004. Rise of the creative class by Richard Florida. Cambridge, MA: Harvard University.

Glaeser, E.L. 2007. The economic approach to cities. Cambridge, MA: Harvard University Department of Economics.

Glaeser, E.L., Scheinkman, J.A. & Shleifer, A. 1995. Economic growth in a cross-section of cities. *Journal of Monetary Economics*, 36(1):117-143.

Haaland, J.I., Kind, H.J., Midelfart Knarvik, K.H. & Torstensson, J. 1999. What determines the Economic Geography of Europe*?* Discussion paper no. 2072. London: Centre for Economic Policy Research.

Haasbroek, P.J. 1985. Ekonomiese ontwikkelingsfases: Die grondslag vir verstedeliking. (*In* Van Jaarsveld, F.A., *ed*. Verstedeliking in Suid-Afrika. Pretoria: Universiteit van Pretoria. pp. 1-14).

Hartshorne, R. 1939. The nature of Geography. A critical survey of current thought in light of the past. Lancaster, PA: AAG.

Hattingh, P.S. & Olivier, J.J. 1985. Swart verstedeliking in Suid-Afrika. (*In* Van Jaarsveld, F.A., *ed*. Verstedeliking in Suid-Afrika. Pretoria: Universiteit van Pretoria. pp. 28-44).

Helpman, E. 1984. A simple theory of international trade with multinational corporations. *Journal of Political Economy*, 92(1):451-471.

Helpman, E. & Krugman, P.R. 1985. Market structure and foreign trade. Cambridge, MA: MIT Press.

Henderson, J.V. 1974. The sizes and types of cities. *American Economic Review*, 64(1):640-656.

Henderson, J.V. 1977. Economic theory and the cities. New York: Academic Press.

Henderson, J.V. 1988. Urban development: Theory, fact and illusion. Oxford: Oxford University Press.

Henderson, J.V. 1996. Ways to think about urban concentration: neoclassical urban systems versus the new economic geography. *International Regional Science Review*, 19(1-2):31-36.

Henderson, J.V. 2003. Urbanization and economic development. *Annals of Economics and Finance*, 4(1):275-341.

Hindson, D. 1987. Urbanisation strategies in South Africa: A critical evaluation. *Third World Quarterly*, 9(2):583-600.

Hirschman, A. 1958. Strategy of economic development. New Haven, CT: Yale University Press.

Hohenberg, P. & Lees, L.H. 1985. The making of urban Europe (1000-1950). Cambridge, MA: Harvard University Press.

Hoover, E.M. 1948. The location of economic activity. New York: McGraw-Hill Book Company, Inc.

Houghton, D.H. 1956. A summary of the findings and recommendations in the Tomlinson Commission report. Commission for the Socio-economic Development of the Bantu Areas

within the Union of South Africa., & South African Institute of Race Relations. Johannesburg: South African Institute of Race Relations.

Houghton, D.H. 1973. The South African economy. 3rd ed. Wynberg: Rustica Press.

Hyslop, J. & Tomlinson, R. 1984. Industrial decentralisation and the 'new dispensation'. *South African Labour Bulletin*, 10(3):114-122.

Ioannides, Y.M. & Overman, H.G. 2003. Zipf's law for cities: an empirical examination. *Regional Science and Urban Economics*, 33(1):127-137.

Isard, W. 1949. The general theory of location and space-economy. *The Quarterly Journal of Economics*, 63(4):476-506.

Isard, W. 1953. Some emerging concepts and techniques for regional analysis. *Journal of Institutional and Theoretical Economics*, 109(2):240-250.

Isard, W. 1956. Location and space economy. Cambridge, MA: MIT Press.

Isard, W. 1960. Methods of regional analysis. Cambridge, MA: MIT Press.

Isserman, A. 1993. Lost in space? On the history, status and future of regional science. *Review of Regional Studies*, 23(1):1-50.

Jacobs, J. 1969. The economy of cities. New York: Vintage.

Johnson, D.G. 2002. Have the urban-rural disparities increased since 1978 in China? Chicago: University of Chicago.

Jones, C.F. 1935. Economic geography. New York: Henry Holt & Co.

Kind, H., Midelfart-Knarvik, K. & Schjelderup, G. 2000. Competing for capital in a 'lumpy' world. *Journal of Public Economics*, 78(1):253-274.

Kleynhans, E. & Drewes, E. 2008. The influence of location on the competitiveness of South African industries. *Journal of Economic and Financial Sciences*, 2(2):139-158.

Kotzenberg, H.R.P.A. 1973. The policy and programme for the decentralisation of industry in South Africa. *Finance and Trade Review*, 10(4):137-155.

Krugell, W.F. 2005. The geographical economy of South Africa. Potchefstroom: NWU. (Thesis - PhD).

Krugell, W.F. 2014. The spatial persistence of population and wealth during apartheid: comparing the 1911 and 2011 censuses. *Economic History of Developing Regions*, 29(2):336-352.

Krugman, P.R. 1979. Increasing returns, monopolistic competition and international trade. *Journal of International Economics*, 9(1):469-79.

Krugman, P.R. 1980. Scale economies, product differentiation, and the pattern of trade. *American Economic Review*, 9(1):950-959.

Krugman, P.R. 1991a. Increasing returns and economic geography. *Journal of Political Economy*, 99(3):483-499.

Krugman, P.R. 1991b. Geography and trade. Cambridge, MA: MIT Press.

Krugman, P.R. 1991c. History versus expectations. *Quarterly Journal of Economics*, 106(1):651-667.

Krugman, P.R. 1991d. First nature, second nature and metropolitan location. *Journal of Regional Science*, 3(1):29-44.

Krugman, P.R. 1993a. On the relationship between trade theory and location theory. *Review of International Economics*, 1(1):110-122.

Krugman, P.R. 1993b. On the number and location of cities. *European Economic Review*, 37(1):293-298.

Krugman, P.R. 1993c. First nature, second nature and metropolitan location. *Journal of Regional Science*, 3(1):129-144.

Krugman, P.R. 1994. Complex landscapes in economic geography. *American Economic Review*, 84(2):412-416.

Krugman, P.R., 1995. Development, geography and economic theory. Cambridge, MA: MIT Press.

Krugman, P.R. 1996a. The self-organising economy. Oxford: Blackwell.

Krugman, P.R. 1996b. Urban concentration: The role of increasing returns and transport costs. *International Regional Science Review*, 19(1-2):5-30.

Krugman, P.R. 1998. What's new about the New Economic Geography? *Oxford Review of Economic Policy*, 14(2):7-17.

Krugman, P.R. & Venables, A. 1996. Integration, specialization and adjustment. *European Economic Review*, 40(1):959-967.

Kuznets, S. 1973. Modern economic growth: Findings and reflections. *The American Economic Review*, 63(3):247-258.

Lewis, W.A. 1954. Economic development with unlimited supplies of labour. *Manchester School of Economic and Social Studies*, 22(1):139-191.

Lösch, A. 1940. Die räumliche ordnung der wirtschaft. Jena: Gustav Fischer.

Lucas, R.E. 1988. On the mechanisms of economic development. *Journal of Monetary Economics*, 22(1):3-42.

Mabin, A. 1991. The dynamics of urbanisation since 1960. (*In* Swilling, M., Humphries P. & Shubane, K., *eds*. Apartheid city in transition. Cape Town: Oxford University Press. pp. 33-47).

Mallick, O.B. 2005. Development theory: Rostow's five-stage model of development. Essay. Newcastle: The University of Newcastle School of Social Science Faculty of Education and Arts.

Marshall, A. 1890. Principles of economics. London: MacMillian.

Martin, R. 1999. The new 'geographical turn' in economics: Some critical reflections. *Cambridge Journal of Economics*, 23(1):65-91.

Martin, P. & Ottaviano, G.I.P. 1999. Growing locations: Industry location in a model of endogenous growth. *European Economic Review*, 43(2):281-302.

Martin, P. & Ottaviano, G.I.P. 2001. Growth and agglomeration. *International Economic Review*, 42(4):947-968.

Matsuyama, K. 1991. Increasing returns, industrialisation, and indeterminacy of equilibrium. *Quarterly Journal of Economics*, 106(1):617-650.

Mills, E.S. 1967. An aggregate model of resource allocation in a metropolitan area. *American Economic Review*, 57(1):197-210.

Molho, I. 1987. The migration decision of young men in Great Britain. *Applied Economics*, 19(1):221-243.

Moolman, P.L. 1984. Rol van desentralisasiebeleid in kleinsakeontwikkeling. Potchefstroom: PU vir CHO. (Thesis - PhD).

Muller, N.D. 1983. Rural welfare and constrained urbanisation in Southern Africa. (*In* Spies, P.H., *ed*. Urban-rural interaction in South Africa. Stellenbosch: University of Stellenbosch. pp. 45-76).

Mulligan, G.F. 1984. Agglomeration and central place theory: A review of the literature. *International Regional Science Review*, 9(1):1-42.

Murray, C. 1976. Keeping house in Lesotho. Cambridge: University of Cambridge. (Thesis - PhD)

Murray, C. 1987. Displaced urbanisation: South Africa's rural slums. *African Affairs*, 86(344):311-329.

Muth, R. 1969. Cities and housing. Chicago: University of Chicago Press.

Myrdal, G. 1957. Economic theory and underdeveloped regions. London: Duckworth.

National Planning Commission **see** South Africa. National Planning Commission.

Naudé, W.A. & Krugell, W.F. 2003. An inquiry into cities and their role in subnational economic growth in South Africa. *Journal of African Economies*, 12(4):476-499.

Naudé, W.A. & Krugell, W.F. 2005. The geographical economy of South Africa. *Journal of Development Perspectives*, 1(1):85-124.

Neary, J.P. 2001. Of hype and hyperbolas: Introducing the new economic geography. *Journal of Economic Literature*, 39(2):536-561.

Neary, J.P. 2004. Monopolistic competition and international trade theory. (*In* Brakman S. & Heijdra, B.J., *eds*. The monopolistic competition revolution in retrospect. Cambridge: Cambridge University Press. pp. 159-84).

Nijkamp, P. & Poot, J. 1997. Endogenous technological change, long run growth and spatial interdependence: A survey. (*In* Bertuglia, C.S., Lombardo S. & Nijkamp, P., *eds*. Innovative behaviour in space and time. Advances in spatial science ed. Heidelberg: Springer-Verlag Berlin. pp. 213-238).

Nunn, N. & Puga, D. 2007. Ruggedness: The blessing of bad geography in Africa. Discussion paper no. 6253. London: Centre for Economic Policy Research.

Ohlin, B. 1933. Interregional and international trade. Cambride, MA: Harvard University Press.

Oranje, M. & Merrifield, A. 2010. National spatial development planning in South Africa 1930-2010: an introductory comparative analysis. *Town and Regional Planning*, 56(1):9-45.

Ottaviano, G.I.P. 2003. Regional policy in the global economy: Insights from the New Economic Geography. *Regional Studies*, 37(1):665-673.

Ottaviano, G.I.P. & Puga, D. 1998. Agglomeration in the global economy: A survey of the 'new economic geography'. *The World Economy*, 21(6):707-731.

Ottaviano, G., Tabuchi, T. & Thisse, J.-F. 2002. Agglomeration and trade revisited. *International Economic Review*, 43(2):409-435.

Ottaviano, G.I.P. & Thisse, J.-F. 2004. Agglomeration and economic geography. (*In* Henderson J.V. & Thisse, J.-F., *eds*. The handbook of regional and urban economics. vol. IV, Cities and geography ed. Amsterdam: North-Holland. pp. 2563-2608).

Owen, D. & Green, A. 1992. Migration patterns and trends. (*In* Champion T. & Fielding, T., *eds*. Migration processes & patterns. London & New York: Belhaven Press. pp. 17-40).

Oxford Dictionary, 1964. London: Oxford University Press.

Palmer, P.N. 1980. Industrialising the national states: The Bophuthatswana example. *Developing studies Southern Africa*, 3(1):33-55.

Parnell, S. & Mabin, A. 1995. Rethinking urban South Africa. *Journal of Southern African Studies*, 21(1):39-61.

Perroux, F. 1955. Note sur la notion de pole de croissance. *Economique Appliquée*, 7(1):307-320.

Phalatse, M.R. 2001. Abandoned industrial spaces in post-apartheid South Africa: Implications for women in Mogwase. *South African Geographical Journal*, 83(2):167-172.

Porter, M. 1990. The competitive advantage of nations. London: Macmillan.

Porter, M. 1994. The role of location in competition. *Journal of the Economics of Business*, 1(1):35-39.

Porter, M. 1996. Competitive advantage, agglomeration economies and regional policy. *International Regional Science Review*, 19(1-2):85-90.

Posel, D. 1991. Curbing African urbanization in the 1950s and 1960s. (*In* Swilling, M., Humphries P. & Shubane, K., *eds*. Apartheid city in transition. Cape Town: Oxford University Press. pp. 19-30).

Posel, D. 2003. Have migration patterns in post-apartheid South Africa changed? Paper presented at the Conference on African Migration in Comparative Perspective, Johannesburg, 4-7 June. http://www.queensu.ca/samp/migrationresources/Documents/Posel_have.pdf Date of access: 20 Oct. 2014.

Quah, D.T. 1997. Regional cohesion from local isolated actions: I. Historical outcomes. Discussion paper no. 378. London School of Economics.

Redding, S.J., Sturm, D.M. & Wolf, N. 2007. History and industry location: Evidence from German airports. Discussion paper No. 6345. London: Centre for Economic Policy Research.

Ritsilä, J. 1999. Regional differences in environments for enterprises. *Entrepreneurship and Regional Development*, 11(1):187-202.

Ritsilä, J. 2000. Mobility of the highly educated labour force. Working paper 214. Jyväskylä, Finland: University of Jyväskylä.

Ritsilä, J. & Ovaskainen, M. 2001. Migration and regional centralization of human capital. *Applied Economics*, 33(1):317-325.

Robinson, E.V.D. 1909. Economic geography: An attempt to state what it is and what it is not. *American Economic Association Quarterly, 3rd series*, 10(1):247-257.

Rogerson, C.M. 1997. Local economic development and post-apartheid reconstruction in South Africa. *Singapore Journal of Tropical Geography*, 18(2):175-195.

Rogerson, C.M. 2008. Consolidating local economic development in post-apartheid South Africa. *Urban Forum*, 19(3):7-28.

Romer, P.M. 1986. Increasing returns and long-run growth. *Journal of Political Economy*, 94(1):1002-1037.

Romer, P.M. 1990. Endogenous technological change. *Journal of Political Economy*, 98(2):71-101.

Rosen, K.T. & Resnick, M. 1980. The size distribution of cities: An examination of the Pareto law and primacy. *Journal of Urban Economics*, 8(1):165-186.

Rosenstein-Rodan, P. 1943. Problems of industrialisation in eastern and south-eastern Europe. *Economic Journal*, 53(1):202-211.

Rostow, W. 1960. The stages of economic growth. Cambridge: Cambridge University Press.

Sachs, J.D. 2003. Institutions don't rule: Direct effects of geography on per capita income. Working Paper no. 9490. Cambridge, MA: National Bureau of Economic Research.

Samuelson, P.A. 1952. The transfer problem and transport costs: The terms of trade when impediments are absent. *Economic Journal*, 62(1):278-304.

Sbergami, F. 2002. Agglomeration and economic growth: Some puzzles. HEI working paper no. 02/2002. Geneva: The Graduate Institute of International Studies.

Schaefer, F.K. 1953. Exceptionalism in geography: A methodological introduction. *Annals, Association of American Geographers*, 43(1):226-249.

Schaeffer, P. 1985. Human capital accumulation and job mobility. *Journal of Regional Science*, 25(1):103-114.

Scott, A.J. 2000. Economic geography: The great half-century. *Cambridge Journal of Economics*, 24(1):483-504.

Simon, C.J. 1998. Human capital and metropolitan employment growth. *Journal of Urban Economics*, 43(1):223-243.

Sjaastad, L. 1962. The costs and returns in human migration. *Journal of Political Economy*, 70(1):80-93.

Smit, P. 1973. Die ontvolking van die blanke platteland: Onlangse tendense. Pretoria: Universiteit van Suid-Afrika.

Smit, P., Booysen, J.J. & Cornelius, I. 1983. Bevolkingsverspreiding in die RSA, Transkei, Bophuthatswana, Venda en Ciskei. 1980s ed. Pretoria: RGN-HSRC.

Smith, J.R. 1913. Industrial and commercial geography. New York: Henry Holt and Co.

Soo, K.T. 2005. Zipf's law for cities: A cross-country investigation. *Regional Science and Urban Economics*, 35(1):239-263.

South Africa. Department of Trade and Industry. 2007. A national industrial policy framework. Pretoria: Government Printer.

South Africa. Department of Cooperative Governance and Traditional Affairs. 2014. Towards an integrated urban development framework. A discussion document. Pretoria.

South Africa. National Planning Commission. 2011. National Development Plan: Vision for 2030. (RP270/2011). Pretoria: Government Printer.

South Africa. Department of Trade and Industry. 2013. Industrial policy action plan. Pretoria: Government Printer.

Starkey, O.P. 1967. Joseph Russel Smith, 1874-1966. *Annals, Association of American Geographers*, 57(1):198-202.

Statistics South Africa. 2012. Census 2011 statistical release - P0301.4. Pretoria.

Swilling, M., Humphries, P. & Shubane, K., *eds*. 1991. Apartheid city in transition. Cape Town: Oxford University Press.

Tabuchi, T. 1998. Urban agglomerationa and dispersion: A synthesis of Alonso and Krugman. *Journal of Urban Economics*, 44(1):333-351.

Tapscott, C.P.G. 1990. Banking on development: The development bank of Southern Africa. Collected seminal papers no. 38. Institute of Commonwealth Studies.

Thisse, J.-F. 1997. L'oubli de l'espace dans la pensée économique. *Région et Développement*, 6(1):13-39.

Turok, I. 2012. Urbanisation and development in South Africa: Economic imperatives, spatial distortions and strategic responses. Working paper no. 8. Cape Town: United Nations Population Fund International Institute for Environment and Development.

Van der Horst, S.T. 1971. Native labour in South Africa. London: Frank Cass.

Van Jaarsveld, F.A. 1982. Die Afrikaners se groot trek na die stede en ander opstelle. Johannesburg: Perskor.

Van Jaarsveld, F.A. *ed*. 1985. Verstedeliking in Suid-Afrika. Pretoria: Universiteit van Pretoria.

Van Marrewijk, C. 1999. Capital accumulation, learning and endogenous growth. *Oxford Economic Papers*, 51(1):453-475.

Van Marrewijk, C. 2012. International economics. 2nd ed. Oxford: Oxford University Press.

Van Zyl, J.C. 1983. South Africa in world trade. (*In* Spicer, M., Begg, S. & Vale, P.C.J., *eds*. International affairs bulletin. 73rd ed. South African Institute of International Affairs. pp. 18-37).

Vernon, R. 1960. Metropolis 1985. Cambridge, MA: Harvard University Press.

Vining, D.R., Pallone, R. & Plane, D. 1981. Recent migration pattern in the developed world: A classification of some differences between our and IIASA's findings. *Environment and Planning*, 13(1):243-250.

Von Thünen, J.H. 1826. Der isolierte staat in beziehung auf landwirtschaft und nationalökonomie. Hamburg: Perthes.

Walz, U. 1996a. Long-run effects of regional policy in an economic union. *Annals of Regional Science*, 30(1):165-183.

Walz, U. 1996b. Transport costs, intermediate goods and localised growth. *Regional Science and Urban Economics*, 26(1):671-695.

Weber, A. 1909. Über den standort der industrien. Tübingen: JCB Mohr.

Weiss, Y. 1971. Learning by doing and occupational specialization. *Journal of Economic Theory*, 4(1):189-198.

Wellings, P. & Black, A. 1984. Industrial decentralisation under apartheid: An empirical assessment of industrial decentralisation in South Africa. Working paper no. 12. Durban: University of Natal Development Studies Unit.

Wellings, P. & Black, A. 1986. Industrial decentralisation in South Africa: Tool of apartheid or spontaneous restructuring. *GeoJournal*, 12(2):137-149.

Wessels, A. & Wentzel, M.E. 1989. Die invloed van relevante kommissieverslae sedert Uniewording op regeringsbeleid te opsigte van swart verstedeliking en streekontwikkeling. Verslag IGN-T1. Pretoria: Raad vir Geesteswetenskaplike Navorsing.

Wheaton, W. & Shishido, H. 1981. Urban concentration, agglomeration economies, and the level of economic development. *Economic Development and Cultural Change*, 30(1):17-30.

Whitbeck, R.H. 1914. Industrial and commercial geography by J. Russel Smith. *Bulletin of the American Geographical Society*, 46(1):540-541.

Whitbeck, R.H. 1915-16. Economic geography: Its growth and possbilities. *Journal of Geography*, 14(1):284-296.

Whitbeck, R.H. & Finch, V.C. 1935. Economic Geography. 3rd ed. New York: McGraw-Hill.

Whitbeck, R.H. & Finch, V.C. 1924. Economic Geography. New York: McGraw-Hill.

Williamson, J. 1965. Regional inequality and the process of national development. *Economic Development and Cultural Change*, 14(1):3-45.

Williamson, J.G. 1988. Migration and urbanisation. (*In* Chenery H. & Srinivasan, T.N., *eds*. Handbook of development economics. 1st ed. Amsterdam: North-Holland. pp. 425-465).

Wilson, F. 1971. Farming 1866-1966. (*In* Wilson M. & Thompson, L., *eds*. The Oxford History of South Africa: Volume II: South Africa 1870-1970. Oxford: Oxford University Press. pp. 104-171).

Wilson, F. 1972. Labour in the South African gold mines, 1911-1969. Cambridge: Cambridge University Press.

World Bank, 2000. World Development Report 1999/2000. 21st ed. Washington, DC: The World Bank.

World Bank, 2009. Reshaping economic geography. World development report. Washington, DC: The World Bank.

Yudelman, D. 1984. The emergence of modern South Africa: State, capital, and the incorporation of organised labour on the South African gold fields, 1902-1939. Westport, Ct: Greenwood Press.

Zipf, G. 1949. Human behaviour and the principle of last effort. Cambridge, MA: Addison-Wesley.