THE GEOGRAPHICAL ECONOMY OF SOUTH AFRICA

W.F. KRUGELL M.Com. M.Sc.

Thesis submitted for the degree Philosophiae Doctor in Economics at the Potchefstroom Campus of the North-West University

Promoter: Prof. W.A. Naudé

May 2005

Potchefstroom

Acknowledgements

An enterprise such as this can never be a one-man show and this thesis owns a debt of gratitude to many.

First and foremost I would like to thank my supervisor, Professor Wim Naudé, for a working relationship that started ten years ago with an undergraduate essay. Since then, his ideas, guidance and support have proven to be the cornerstone of my research efforts.

Second, this thesis forms part of a project cluster on spatial economics in the Research Focus Unit: *WorkWell* and draws on broader outputs produced from different collaborations. These outputs include five conference papers that I have co-authored with Prof Wim Naudé and a number of articles. The published work includes a chapter in a book, a working paper, a conference review article, as well as articles in international journals:

NAUDé, W.A., KRUGELL, W.F. & SERUMAGA-ZAKE, P. 2002. Cumulative causation and decentralised industrial development in South Africa. (*In*: Higano, Y., Nijkamp, P., Poot, J. & Van Wijk, J.J. *eds*. The region in the new economy. London: Ashgate. p. 407-430.)

NAUDé, W.A. & KRUGELL, W.F. 2004. An enquiry into cities and their role in sub-national economic growth in South Africa. WIDER research paper RP2004/08.

NAUDé, W.A. & KRUGELL, W.F. 2003. Spatial inequality in Africa. (Review of the conference on spatial inequality in Africa (WIDER project on spatial disparities in human development), Centre for the study of African economies, University of Oxford.) *Development Southern Africa*, 20(1): 161-167.

NAUDé, W.A. & KRUGELL, W.F. 2003. Are South African cities too small? *Cities: The international journal of urban policy and planning*, 20(3): 175-180.

NAUDé, W.A. & KRUGELL, W.F. 2003. An enquiry into cities and their role in sub-national economic growth in South Africa. *Journal of African economies*, 12(4): 476-499.

NAUDé, W.A. & KRUGELL, W.F. 2005. Economic geography and growth in Africa: The determinants of sub-national growth in South Africa. *Papers in regional science*, forthcoming.

Thanks to *WorkWell* and again Prof. Naudé for the opportunity to be part of the spatial economics project cluster.

Also, I am pleased to acknowledge the financial support that the above work received from the National Research Foundation (Grant number 2053340) and Volkswagen Foundation.

The Volkswagen Foundation grant also involved collaboration with the University of Paderborn in Germany. Many thanks are due for the valuable inputs of Prof Thomas Gries at the University of Paderborn.

The work here would also not have been possible if it were not for the support of Global Insight Southern Africa. Martin Cameron graciously provided access to the Regional Economic Focus and expertise at Global Insight.

There are also many others that I wish to thank on a more personal note. My parents, Jan and Elsa Krugell, my grandparents, as well as brothers Johann and Cobus and my sister Marike were steadfast in their support. Hannelie always provided encouragement. Thanks are also due to my colleagues in the School of Economics, Risk Management and International Trade for their good humour and all the coffee breaks throughout my studies.

The language editing was done by Rod Taylor.

Printing was done by Platinum Press.

Binding was done by the North-West University library (Potchefstroom Campus).

Waldo Krugell

Potchefstroom, May 2005.

Abstract

This study examines the determinants of economic growth at sub-national level in South Africa, and investigates cross-locality medium-term (five-year) growth rate differentials between 354 magisterial districts. The period in question is 1998 to 2002. A dynamic panel data regression model is used that includes measures of geography (distance and natural resources) as well as recent estimates of physical and human capital. It is found that the significant determinants of local economic growth are distance from internal markets, human capital, export propensity, and the capital stock of municipalities (reflecting institutional quality and governance on local government level). Distance from international harbours, as a measure of transport costs, and urban agglomeration (or density) affects growth indirectly through its significant effect on the ability of a region to export. Overall, these results indicate that geography is important for economic growth, independent of its effects in institutions. Bearing in mind the medium-term focus of the work, no evidence of absolute convergence could be found over a five-year period, rather the tentative evidence suggests slow beta convergence.

Key words:

Geographic concentration, and agglomeration, panel data regression, growth determinants, South Africa, spatial economic development

Opsomming

Hierdie studie stel ondersoek in na wat ekonomiese groei op 'n sub-nasionale vlak in Suid-Afrika bepaal, en kyk na verskille in groeikoerse tussen 354 plaaslike owerhede oor die medium termyn (vyf jaar). Die tydperk is van 1998 tot 2002. 'n Dinamiese paneeldata regressiemodel word gebruik en maatstawwe van geografie (afstand en natuurlike hulpbronne) en van fisiese en menslike kapitaal word gebruik. Die resultate toon dat ekonomiese groei op plaaslike vlak bepaal word deur 'n plek se afstand vanaf binnelandse markte, menslike kapitaal, uitvoergeneigdheid en die kapitaalvoorraad van die plaaslike owerheid (dit reflekteer die gehaalte van die instellings en die uitvoering van die plaaslike owerheid se take). Die afstand vanaf internasionale hawens, as 'n maatstaf van vervoerkoste, en stedelike agglomerasie (of digtheid) het 'n indirekte invloed op groei deurdat dit 'n beduidende bepaler is van 'n area se vermoë om uit te voer. Saam dui die resultate daarop dat geografie 'n belangrike invloed het op ekonomiese groei, onafhanklik van die invloed wat dit het op instellings. Daar is geen bewys dat absolute konvergensie plaasgevind het oor die vyf jaar periode nie, maar daar is voorlopige bewyse wat dui op stadige beta-konvergensie.

Sleutelwoorde:

Geografiese konsentrasie, agglomerasie, paneeldata regressie, determinante van groei, Suid-Afrika, ruimtelike ekonomiese ontwikkeling

Table of contents

List of tables

List of figures

Chapter 1: Introduction

1.1 Introduction

A particular characteristic of economic activity across South Africa and across the globe is its spatial lumpiness. Geographically, economic activity tends to be unequally distributed and concentrated. In South Africa, 70 per cent of GDP is produced in only 20 per cent of places.

The South African economy is, however, better known for the challenges it faces such as low economic growth, poverty and inequality. In the ten years since democratisation much has been written about the South African economy, and the major themes include economic growth (or the lack thereof), the impact of the opening up of the economy and globalisation, fiscal adjustment, inflation targeting, exchange rate management and issues of poverty and inequality. With much of this work, though, the focus is only on the level of the national economy. The post-democratisation period was, however, also marked by a significant decentralisation of economic decision making in an economy characterised by significant spatial inequality. The transformation of the system of local government has resulted in local authorities that are constitutionally responsible for the development of their areas. Recently, the National Spatial Development Perspective set spatial priorities for all spheres of government.

Nonetheless, in academic (specifically economics) circles, the questions of the location of economic activity in South Africa have received limited attention. On this point, it is, however, prudent to emphasise the distinction between geographical economics and economic geography. Economic geography examines the role of geography in urban and regional economics and this is a well-developed field in South Africa[1]. The development of geographical economics is, however, more recent. Geographical economics aims to show that the decisions of economic agents are determined by geography and the geography itself can be derived from the behaviour of economic agents (Brakman, Garretsen and Van Marrewijk, 2001:22). Concerning the field in South Africa, Naudé (2005) stated that "*By and large, however, economists have not yet provided sufficient rigorous economic analyses of the Southern African spatial economy*". This study aims to contribute to addressing this

[1] Section 4.3 provides a brief overview of the literature.

shortcoming by following the geographical economics approach and examining the determinants of the spatial growth and development of the South African economy.

1.2 Background

The significance of the location of economic activity in space is not limited to the South African context. The issues above can be framed against the background of developments in economics' explanations of the occurrence of trade and economic growth.

The maturity of "New Growth" and "New Trade" theories has seen the explanations of the occurrence of trade and economic growth move from those of differential allocations of resources, labour and capital, to accounts that emphasise imperfect competition, institutions and geography. In the words of Warner (2002:1) "recent research on the causes of the large differences in economic development across countries has framed the issue as a competition between geography and institutions".

Concurrently a new economic geography, or geographical economics framework, has recently developed. It argues that the explanation of differential growth and trade experiences lies beyond so-called "first nature geography". Analyses of economic growth and trade should firstly be about explaining the location of production in space – and this is driven by a fundamental trade-off between increasing returns and transport costs (Fujita & Thisse, 2002). The explanations of the ways in which economic agglomerations are formed appeal to first-nature geography (for example climate, or unevenness in the distribution of resources), to non-market institutions (such as externalities that give rise to endogenous spatial inhomogeneities, i.e. second-nature geography), and to an imperfectly competitive paradigm (Fujita & Thisse, 2002:45).

In the South African circumstance, little cognisance has been taken of the implications that geographical economics holds for the challenges that face the South African economy. The relevance of the location of production in space and the trade-off between increasing returns and transport costs is, however, vividly illustrated in the following article from the *Beeld* daily newspaper (De Lange, 2004).

The heading says, "*Trailers move away from Mookgopong*" and the article proceeds to report as follows. Mookgopong (which used to be known a Naboomspruit) is probably going to turn into a ghost town when Venter trailers and Challenger trailers, the two firms that form the backbone of this town, move to the city. Venter Trailers produces its products in the

main road of the town, but will soon be moving to new premises in Roodepoort (in the Johannesburg area) and Challenger is moving to a factory building that it has bought in Watloo, Pretoria. In fact, Venter trailers has for some time threatened to move due to the cost of transporting raw materials from Gauteng. When they recently announced the move to Roodepoort, Challenger had to follow suit. Mr Jasper Venter, the owner of Challenger trailers states that the transport costs from Gauteng amount to R500 per trailer, and the relocation gives Venter trailers a comparative advantage. He states that three loads are transported from Gauteng every day and that the toll fees cost R300 per trip. Earlier it was possible to use rail transport by Spoornet, but the rates are unaffordable now – ten years ago it cost R120 to transport a trailer by rail to Cape Town, but today it is R1200. Mr Jasper Venter says that the decision to move Challenger trailers to Pretoria is about nothing more than distance. It may be a case of move to the city or sell out. The local council has tried to improve conditions for the companies: they pay special low rates for services and the council has tried to convince the provincial government to waive their toll fees, but that would have created a precedent. Considering the employment opportunities and buying power lost, this is without doubt a major setback for the town.

Box 1.1: Excerpt from Beeld

Waentjies trek weg uit Mookgopong

Jan de Lange

Mookgopong, vroeër Naboomspruit, gaan waarskynlik 'n spookdorp word omdat Venter Sleepwaens en Challenger-sleepwaens, die twee ondernemings wat die ruggraat van die Bosveld-dorpie is, padgee stad toe.

Venter Sleepwaens, wat sy produkte in 'n groot aanleg in Naboomspruit se hoofstraat bou, trek binnekort na 'n nuwe perseel in Roodepoort en Challenger het 'n fabriekgebou in Waltloo, Pretoria, gekoop.

Die twee ondernemings verskaf gesamentlik werk aan sowat 400 mense met 'n koopkrag van 'n geskatte R1,5 miljoen per maand. Die genoteerde Venter Sleepwaens, wat deesdae aan 'n Duitser, mnr. Deetleff Hamman, behoort, dreig lankal om te trek weens die vervoerkoste van grondstowwe van Gauteng.

Venter Sleepwaens huur sy fabriekperseel van mnr. Jasper Venter, wat die onderneming 40 jaar gelede gestig het, maar twee jaar gelede eienaar geword het van Challenger Sleepwaens, wat sowat 2 km buite die dorp is.

Toe Venter Sleepwaens onlangs aankondig dat hy 'n perseel in Roodepoort gekry het, het Jasper Venter besluit dat hy 'n plan sal moet maak.

"Die vervoerkoste van Gauteng hierheen werk uit op sowat R500 per sleepwaentjie. Dit sal beteken dat Venter Sleepwaens 'n mededingende voordeel bo my het," sê hy.

Drie vragmotors ry elke dag vrag van Gauteng aan en lewer Challenger-waentjies af. Net aan tolgeld kos dit R300 per rit. Vrag kon vroeër deur Spoornet vervoer word, maar deesdae is die tariewe onbekostigbaar.

"Tien jaar gelede het dit R120 gekos om 'n waentjie per spoor Kaap toe te stuur. Vandag kos dit R1 200. Die tariewe is vroeër volgens gewig bereken, maar deesdae is dit volgens volume," sê Venter.

"Dit gaan vir ons oor niks anders as die afstand nie. Dit kan baie maklik 'n geval word van stad toe trek of bankrot raak."

Mookgopong se stadsraad het talle vergaderings met alle maatskappye gehou om toestande vir hulle te probeer vergemaklik. "Hulle betaal spesiale lae tariewe vir dienste en ons het selfs met die provinsiale owerheid probeer reël vir vrystelling van tolgeld, maar dit sou 'n presedent skep," sê 'n woordvoerder van die stadsraad.

Die werknemers van die twee waentjie-vervaardigers is die welvarende gedeelte van Mookgopong se sowat 16 000 inwoners. "Dit is ongetwyfeld 'n enorme terugslag vir die dorp," sê die woordvoerder.

As this excerpt in Box 1.1 shows, questions of the location of economic activity in South Africa and of the determinants of spatial growth and development, are relevant and significant for the greater questions of growth and development and for the challenges these pose, particularly for local policymakers. Thus, issues of geography, policies and institutions are important. Yet, as stated in the introduction, very few researchers have looked strictly at the spatial distribution of economic activity or the determinants of the growth of economic activities across different localities – and even fewer explicitly take account of geography. Thus, the contribution of this study may be threefold:

Firstly, it provides provide an empirical test of some of the implications of the geographical economics framework. It has until recently been said that there exists a gap between the theory and empirics of the new economic geography. At a conference held in October 2004 in Hamburg, Germany, efforts were started to draw together researchers who strive to bridge such a gap. The latest work includes direct tests of the theories of

geographical economics in the forms of tests for the home market effect as well as tests for a spatial wage structure. The results are, however, mostly for advanced, industrialised economies and more evidence is needed from developing countries. This study of the determinants of the growth of economic activities across different localities in South Africa, will add to the empirical literature.

Secondly, it contributes to the current debate on the relative roles of institutions versus geography. Here, Warner (2002) pointed out that studies of sub-national economies may be useful to identify the effects of geography on economic activity, as the broader national institutions will be similar across sub-national regions, a feature that does not hold in the case of most country-level studies. This study examines the role of geography as a determinant of growth in 354 magisterial districts in South Africa.

Thirdly, it contributes toward the practical policy debate in South Africa. The sustainability of the decentralisation and the creation of a more equitable spatial economy, may benefit from policies that are informed of the determinants of sub-national growth rates, in particular the ways in which geography impacts on the growth performance of a region.

1.3 Problem statement

The South African economy is characterised by spatial inequality and the economic growth rates differ between cities and towns.

The question is, what determines the spatial growth and development of the South African economy: Is it geography, institutions or policy? The answer is important, for if it is geography, it could imply that current inequalities will persist or even worsen. This has significant implications for policy efforts such as the National Spatial Development Perspective.

1.4 Research question

The primary research question is, to what extent is economic growth across South Africa driven by geography, institutions and/or policies?

This question may be answered by answering two secondary research questions:

- Have per capita incomes across space been converging or diverging?

- Do cities (i.e. places characterised with a high degree of agglomeration) grow faster than places with low degrees of agglomeration, and if so, why?

1.5 Objectives of this study

The primary objective is to provide empirical tests of the extent to which economic growth across South Africa is determined by geography.

This may be achieved by reaching a number of secondary objectives:

- Determining whether per capita incomes across space have been converging or diverging.
- Determining whether cities (i.e. places characterised with a high degree of agglomeration) grow faster than places with low degrees of agglomeration.

1.6 Hypothesis

The central working hypothesis of this study is that geography has a significant effect on the economic growth rate of South African cities and towns, independent of the effects of institutions and policies.

1.7 Methodology

To achieve the above objectives requires a literature review, data, and empirical analysis.

The literature review presents the theory of economic geography, focusing specifically on the new economic geography that explains the location of economic activity in space, and thus trade and growth.

The literature overview also extends to the South Africa space economy, looking at earlier work in the field. The background to the South African space economy involves analysis of data to characterise the economic activities and growth patterns of South Africa's cities and towns. As regards the importance of cities in growth and development in South

Africa, measures such as the primacy ratio, Zipf's Law and the H-measure are used to look specifically at city size, growth and dispersion.

The data analysis is closely linked to the empirical analysis. This study undertakes panel data regression analysis of the determinants of growth and to test whether localities in South Africa have been converging or diverging (β- and σ-convergence). The methods are explained in greater detail in Chapters 3, 4 and 5.

1.8 Outline

This study can be outlined as follows: Chapter 1 presents the introduction. Chapter 2 provides an overview of the theory of economic geography. The aim is to provide an intuitive explanation of the working of the so-called core model of geographical economics (see Brakman *et al.*, 2001), as a starting point to looking for empirical evidence of the importance of geography for growth in South Africa. Chapter 3 outlines the empirical methods used to test the theories of economic geography. In this field the empirical literature is still developing and the aim of the chapter is to place empirical work in South Africa in context – as extensions of the core model that emphasise economic growth and the role of cities in the space economy. Chapter 4 reviews the South African space economy, discussing the history and institutions, and providing an overview of the literature on the South African space economy. The chapter also shows the importance of cities in the space economy and introduces the economic growth patterns of cities and towns over the period 1998 to 2002. Chapter 5 tests the determinants of growth and the convergence or divergence of local economies. The data used in the empirical analysis are outlined and the results of the regression estimates presented. Conclusions and recommendations are presented in Chapter 6.

Chapter 2: Theoretical overview of the determinants of spatial economic growth

2.1 Introduction

In Chapter 1 it was noted that economics has increasingly been concerned with the spatial distribution of economic activity and aspects of geography. Typically, the spatial distribution of economic activity is highly unequal, with considerable variation in the economic size of cities, towns and regions. Similarly, economic growth rates vary significantly between localities. In this it is possible to distinguish between concentration, specialisation and agglomeration of economic activity. Specialisation examines a country's or a region's economic structure. As Brakman *et al.*, (2001:131) put it, it is the question of whether or not a location's share in the production of cars or apparel is relatively large to the share of other locations in the production of cars or apparel. In contrast, concentration and agglomeration refer to the question of how economic activity as a whole (a specific industry or the whole manufacturing sector) is distributed across space. In this the distinction is a matter of degree. Concentration considers a few well-defined sectors and agglomeration considers the location across space of a much larger part of economic activity, for example the manufacturing sector as a whole (Brakman *et al.*, 2001:129). Thus, explaining the spatial distribution of economic activity, and the determinants of the growth of economic activities across different localities, is about explaining agglomeration. The question that arises is, what are the relative impacts of policies, institutions or geography on agglomeration?

The recent literature on economic growth, development and trade has seen a debate on the relative roles of determinants of growth (see for example Bloom & Sachs, 1998; Gallup & Sachs, 2000; Acemoglu *et al.*, 2001; Sachs, 2001; and Rodrik *et al.*, 2002). In the words of Warner (2002:1) "*recent research on the causes of the large differences in economic development across countries has framed the issue as a competition between geography and institutions*".

In the geographical economics literature (also called the new economic geography framework) such arguments over the determinants of growth are framed in terms of agglomeration. These arguments favour the relative importance of geography as a

determinant of the spatial distribution of economic activity and its growth. However, when geographical economics argues that geography matters for agglomeration and growth, it is not only so-called first-nature geography that is important. The analysis goes beyond climate, disease burden or distance. Explanations of the ways that economic agglomerations are formed appeal to nature (the unevenness in the distribution of resources), to non-market institutions (such as externalities that give rise to endogenous spatial inhomogeneities), as well as to imperfect competition (Fujita & Thisse, 2002:45).

The purpose of this chapter is to review the theories of the spatial structure of economic activity, and determinants of spatial economic growth, so as to provide a basis from which to derive empirically testable hypotheses. The focus is specifically on explanations of agglomeration and the so-called core model of geographical economics of Krugman (1991) (see, for example, Brakman *et al.*, 2001). This approach is, however, neither the only nor the first to proffer explanations of the spatial distribution of economic activity or of the determinants of spatial economic growth – long traditions exist in the fields of urban economics, regional economics, trade and growth theories and development economics. Geographical economics builds on the insights from these different fields and extends them by showing that "*the decisions of economic agents are determined by geography and that geography itself can be derived from the behaviour of economic agents*" (Brakman *et al.*, 2001:22).

The chapter is structured as follows: Section 2.2 provides background to the different explanations of the location of economic activity in space. The section reviews the contributions from urban economics and regional economics, economic growth theory, trade theory and the development economics literature. In Section 2.3 the core model of geographical economics is presented. The aim is to provide an intuitive explanation of the working of the model as a basis for the empirical work to follow. Section 2.4 outlines the testable hypotheses of the determinants of spatial economic growth that can be distilled from this theoretical overview. Section 2.5 takes note of the criticism of geographical economics' approach to explanations of the determinants of spatial economic growth. Section 2.6 concludes.

2.2 Early explanations of the location of production in space

A wide variety of economic theories may be called upon to explain the location, specifically the agglomeration, of production in space. These range from the fields of urban economics, regional economics, trade and growth theories and development economics.

Many of the explanations that follow involve economies of scale. Box 2.1 first gives a brief explanation of these phenomena.

Box 2.1: Economies of scale

Economies of scale (also sometimes labelled as increasing returns to scale) refer to the circumstances where an increase in the level of output produced leads to a decrease in the average cost per unit of output of a firm (Brakman *et al.*, 2001:26). This fall in the average cost occurs because of externalities – costs or benefits that spill over beyond the private costs of, or benefits to, the individual firm.

In this it is possible to distinguish between internal and external economies of scale. Internal economies of scale occur at firm level where increased production results in a cost advantage over smaller firms. This implies market power and a market structure of imperfect competition.

External economies of scale occur at industry level. In this case, an increase in the output of the industry as a whole leads to a decrease in average costs. Such external economies of scale can be further divided into pure (or technological) external economies, and pecuniary external economies.

In the case of pure external economies, an increase in industry-wide output causes a change in the technological relationship between inputs and output for each individual firm. There are two examples of this. The first is that of knowledge sharing, learning and innovation: As industry output rises, the stock of knowledge rises and information spills over to firms. This is a positive external benefit that is not paid for, reducing cost and causing an increase in the level of output at the firm level. Glaeser, Kallal, Scheinkman and Shleifer (1992) distinguished between three types of these externalities: (i) Marshall-Arrow-Romer externalities that are due to knowledge sharing, learning and imitation between firms in the same industry and where local monopoly fosters these spillovers;

(ii) Porter externalities that are industry-specific knowledge spillovers, but where local competition fosters the spillovers; and (iii) Jacobs externalities where knowledge spillovers occur between firms of different industries and where local competition stimulates these spillovers. The second type of spillover from non-market institutions involves public goods.

The supply of public goods and services provides benefits to members of a community – benefits that are non-rivalrous and non-excludable in consumption. Non-rivalry in consumption means that each individual's consumption does not detract from any other individual's consumption of the good or service. Non-excludability in consumption means that it is impossible to exclude anyone from consumption – even when they are not willing to pay for the benefits. Public goods or services thus have external benefits that lower costs and enhance efficiency, giving rise to increasing returns in the aggregate. In both these cases it is important to note the importance of proximity. Proximity makes it possible to capture the spillovers of knowledge, or from infrastructure, which increase productivity and lower costs.

In contrast to such pure externalities that affect the production function, pecuniary externalities affect a firm's output decisions through price effects that are transmitted via the market. Two approaches to pecuniary externalities can be distinguished: The Chamberlainian approach to the diversity of intermediate inputs and the Smith-Marshallian approach to the matching process on the labour market.

The Chamberlainian approach rests on the idea that a large market allows for a large number of intermediate commodities and final goods. Particularly, diversified and non-tradable inputs, such as legal and communication services, non-traded industrial inputs, maintenance and repair services and finance can enhance the productivity of the final sector (Fujita & Thisse, 2002:98). The economy then displays increasing returns to scale at the level of the agglomeration (city level).

The Smith-Marshallian approach holds that the size and proximity of economic activity found in agglomerations ensures a thick labour market that allows for better matching between workers and jobs. In this approach there are two models.

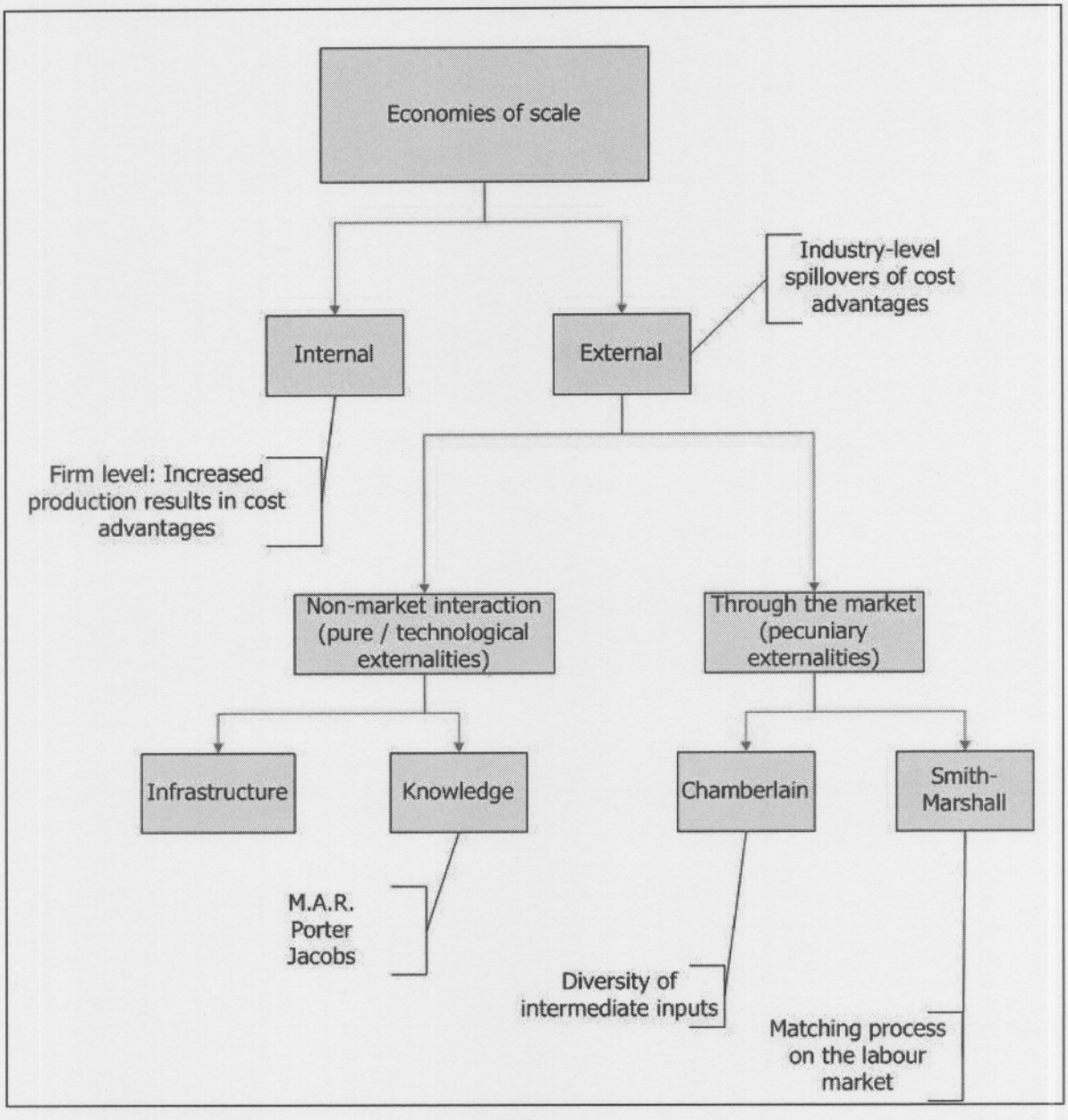
Helsley and Strange (1990) showed that a large city allows for a better average match between heterogeneous workers and firms' job requirements and this enhances efficiency. On the other hand, Duranton (1998) argued that a large market allows workers to become more specialised and, therefore, to be more efficient. Either way, the better matching gives rise to increasing returns in the aggregate.

It is important to note that the price effects of pecuniary externalities can only come about with imperfect competition.

Finally, in the case of external economies, a further distinction is possible between localisation economies and urbanisation economies. Localisation economies are industry-specific external economies. Urbanisation economies are external economies that apply to firms across industries and capture the notion of positive spillovers for a firm as a result of total economic activity at a location (Brakman *et al.*, 2001:29).

Visually, economies of scale may also be illustrated as shown below in Figure 2.1. It is shown that it is possible to distinguish between internal and external economies of scale. Internal economies of scale occur at firm level, where increased production results in cost advantages. External economies of scale occur at industry level in the form of spillovers of cost advantages. The spillovers can take place through non-market interaction or through the market. In the case of non-market interaction the cost advantages accrue to the individual firm from infrastructure or from knowledge sharing. Where the external economies of scale are pecuniary in nature, it is possible to distinguish between the cost benefits that firms receive due to a diversity of intermediate inputs (Chamberlain) and those due to an improved matching process on the labour market (Smith-Marshall).

Figure 2.1: Economies of scale



Following Brakman *et al.* (2001) the rest of this section provides a brief overview of the different contributions from the different fields of urban economics, regional economics, trade and growth theories and development economics. Section 2.3 will show what the recent advances in geographical economics have to add to these analyses.

2.2.1 Agglomeration in urban economics

The uneven distribution of economic activity across space is the starting point of urban economics – its focus is specifically on explanations of the spatial structure of urban areas.

The benchmark model of urban economics is the monocentric city model of Von Thünen. Brakman *et al.* (2001:25-26) explained the working of the model. Briefly, Von Thünen assumed a flat and homogeneous plain, in the middle of which there is a city; outside of the city farmers grow crops that they sell in the city; the farmers face positive transport costs which differ for the various crops, and the prices of the crops differ as well. In essence, the farmers' question of where to locate is determined by a trade-off between land rents and transport costs. Land rents near the city are higher for lower transport costs. For the higher transport costs from the edge of the plain, land rents are lower. In the end, competition for locations ensures that the equilibrium allocation of land among the farmers is efficient (Brakman *et al.*, 2001:24).

In the 1960s Alonso (1964) replaced the city with a central business district and the farmers with commuters. Such a model produces a similar result in that the competition for land among commuters results in the efficient allocation of land.

This stylised explanation of the spatial structure of urban areas concurs with real world observations: Firstly, population density declines with distance from central business district, and secondly, cities have decentralised along with declining transport costs.

The monocentric city model's explanatory power rests on the assumption that the location of economic activity does not involve external effects. However, to explain why the city is there to begin with, or to explain interaction between cities, requires some type of increasing returns to scale. Fujita and Thisse (2000:6-9) argued that economies of scale is one of the basic drivers of urban agglomeration.

Henderson (1988) put forward a model that focuses on the forces that determine the size of cities and the interactions between them. In this model, external economies of scale that are industry-specific make up the agglomerating forces. Thus, when a firm from a specific industry locates in a city where other firms from the same industry are located, it benefits from the positive spillovers of information sharing, a pooled labour market and the existence of specialised suppliers. The model does, however, also contain dispersion forces in the form of diseconomies of scale that depend on the overall size of the city – a large city

implies relatively high commuting costs and land rents. Together, the agglomeration and dispersion forces make it possible to explain systems of cities, where different size cities cater to the needs of different industries and trade with each other. This extension of the model is, however, not without shortcomings. The urban systems approach of Henderson (1998) does not take the existence of the city for granted like the monocentric model, but it deliberately neglects the non-city part of the location of economic activity (Brakman *et al.*, 2001:30).

To summarise, urban economics explains the location of production in space in terms of the formation of cities and their sizes. The concentration of activities in cities is clearly a determinant of spatial economic growth. In larger places, economic activity benefits from the positive spillovers of information sharing, from a pooled labour market and from the existence of specialised suppliers. Agglomeration (and thus the prospects for spatial growth) is, however, dampened by transport costs.

2.2.2 Agglomeration in regional economics

Martin (1999:61) called regional economics "*the formalised successor to the German 'location economics' tradition*". Regional economics has its roots in the tradition of Von Thünen, Christaller, Weber and Lösch, and takes economy-wide space into account to analyse where economic activity will take place (Brakman *et al.*, 2001:31). Two of the key explanations of the location of production in space proffered by regional economics are briefly recounted below: central place theory and the market potential approach.

Central place theory states that centrality determines the types of goods that a location provides. That is to say that the opposing forces of internal returns to scale and transport costs result in a hierarchy of locations that are evenly distributed across space. In the hierarchy the central place is a city that performs all functions (supplies all goods and services) and there are villages that perform only some functions. In Brakman *et al.*'s (2001) example, there are many small locations where bakers sell bread (that has limited increasing returns) and relatively few larger locations where electronics firms sell television sets (that have more scope for increasing returns to scale and people buy television sets less frequently). To minimise transport costs, both locations are rather evenly distributed across space (p.32). In this way, central place theory deals explicitly with the location of economic activity as it is determined by the interplay of increasing returns to scale and transport costs. The shortcoming of this approach is that it is largely a descriptive story and lacks a

microeconomic foundation for the behaviour of consumers and firms. It does, however, reinforce the notion that increasing returns to scale favour the agglomeration of economic activity in specific locations and in that way determines spatial economic growth – again, counterbalanced by transport costs.

Regional economics' second explanation of agglomeration is the market potential approach, but it is less of a theory than an empirical regularity. Brakman *et al.* (2001:35) discussed how Harris (1954) found that the market potential equation provides an indication of the general proximity of a location to total demand. The equation can be reproduced as follows:

$$MP_i = \sum_{j=1}^{R} \left(\frac{M_j}{D_{ij}} \right)$$

Where MP_i is the market potential of location i, M_j is the demand by location j for goods from location i, and D_{ij} is the distance between locations i and j.

Empirical studies have found that market potential is typically high in those areas where production is located. In this way, demand also plays a part in the agglomeration of economic activity: demand is high where production is located as a result of the purchasing power of the workers making production at that location possible (Brakman *et al.*, 2001:36). Thus, large local demand becomes a determinant of spatial economic growth. The shortcoming of this approach is, however, that although it is empirically convincing, it again lacks a theoretical foundation.

2.2.3 Agglomeration in economic growth theory

The relationship between explanations of the location of production in space and economic growth is tenuous. On the one hand, growth theory, allows no role or a limited role for geography as a determinant of growth. On the other hand, data show that economic growth rates vary considerably between countries, between regions and between localities, and high and low growth localities are often geographically concentrated (Brakman *et al.*, 2001:50).

Neo-classical growth theory explains economic growth in the short run by means of capital accumulation. The capital accumulation is subject to the law of diminishing returns. Then, if the capital stock (per capita) is low for initially poor countries, regions, or localities this implies a high return of investment and leads to a convergence process whereby

countries, regions, or localities end up with the same equilibrium level of output per capita. This absolute convergence assumes that all places have access to the same technology and are equal in structures, institutions, etc. It is the exogenous technological progress that determines growth in the long-run steady state. In this model spatial agglomeration of high or low economic growth rates is hardly important.

In practice there is, however, little evidence of absolute convergence and this has lead to the study of conditional convergence. This means that neo-classical growth theory is modified to allow for differences between countries, regions, or localities. Convergence is then conditioned on the characteristics of a place –countries, regions, or localities then need not to converge to the same long-run equilibrium level of output per capita. Brakman *et al.* (2001:51) argued that this extension allows for a link between neo-classical growth theory and the place where the growth occurs.

The location of production matters for conditional convergence to the extent that differences in technology or institutions may be location-specific. Here the reference is specifically to physical geography, or so-called first-nature geography, that gives rise to natural cost advantages and agglomeration of economic activity. A range of empirical studies have shown that physical geography has an impact on economic growth at different locations. Particularly, at the country level, studies have shown that countries in the tropics are relatively poor (malaria significantly lowers growth) compared to those in more temperate zones, and landlocked countries are relatively poor compared to those at the coast (see for example Gallup, Sachs and Mellinger, 1998; Gallup & Sachs, 2000; Acemoglu, Johnson and Robinson, 2001 as well as Sachs, 2001). In terms of a theoretical model this link between growth and the places where the growth occurs is, however, quite indirect and determined outside the model.

A different extension of neo-classical growth theory that also allows for a link between growth and the location of economic activity is to examine the alternative model known as the new growth theory. This approach makes economic growth endogenous and allows for increasing returns to scale. Endogenisation of the growth process focuses on the roles of human capital (Lucas, 1988), research and development (Romer, 1986), learning-by-doing (Young, 1993) and infrastructure (Aschauer, 1989) and their associated positive external economies. This makes it possible no longer to have diminishing returns to accumulable factors of production. In itself this does not mean that there is automatically a

role for agglomeration – if economies are described by a uniform global production function, location is unimportant. Location only matters if the spillovers associated with the external economies are somehow localised (Brakman *et al.*, 2001:52). Grossman and Helpmann (1991) provided an example of localised spillovers where the positive externalities associated with research and development only exist within a certain group of countries. The existence of such localised externalities may be ascribed to differences between cultures, politics and institutions in different countries, regions, or localities. From the modelling point of view this makes it possible to explain agglomeration and account for differences in growth rates. However, the role of location is still not endogenous to the model.

2.2.4 Agglomeration in development economics

When discussing the role of location and agglomeration of economic activity in development economics, development economics refers to the theories of Rosenstein-Rodan, Myrdal and Hirschman. These are what Krugman (1993) called the "high theories of development economics" – theories that today have, to some extent, been replaced with the application of neo-classical economics to developing economies. Rosenstein-Rodan, Myrdal and Hirschman nevertheless showed interesting early insights into the spatial dimension of economic development.

Rosenstein-Rodan (1943) argued that industrialisation requires a "big push", meaning that a government-led investment effort will overcome the problems of a small local market and secure the benefits of scale economies for firms. As Brakman *et al.* (2001:55) put it, without a big push in investment, the periphery cannot catch up with the core.

Myrdal (1957) similarly addressed the question of the sustainability of core-periphery patterns of development, but introduced the concept of cumulative causation. He stated that once a country or region takes the lead in economic development (is established as a core), strong localised spillovers of positive external economies will ensure that more firms will want to invest in the locality and more labour will want to work there.

Hirschman (1958), in turn, showed that backward and forward linkages between firms mean that by locating production in a particular region, a firm increases the profitability of other firms doing the same. Thus, here also is the idea of economies of scale at the firm and industry level, providing a story of the location of production in space.

The shortcoming here is that though these theories of development economics offer a story of core-periphery location, they lack the micro-foundations to explain the equilibrium location of economic activity resulting from the underlying behaviour of the economic agents. It does, however, strengthen the intuition behind a number of determinants for spatial economic growth – localities where the economies grow are those characterised by positive external economies. These stem from the size of the market, Marshallian spillovers and a threshold level of investment or infrastructure.

2.2.5 Agglomeration in trade theory

A discussion of the role of the location of economic activity in trade theory should start with the neo-classical trade theory. This refers to familiar theories, such as the Heckscher-Ohlin factor abundance model and the Ricardian comparative advantage model. In these models, trade flows are based on comparative advantage that is caused by technological differences, or differences in factor endowments. Brakman *et al.* (2001:37-41) offered a more complete description of these theories, which falls outside the scope of this section. Suffice to say that, as in the case of neo-classical growth theory, it is only first-nature geography that determines the location of economic activity and that matters for neo-classical trade theory. That is to say that the uneven distribution of endowments is a first-nature determinant of location: the stock of natural resources shapes comparative advantage, and physical geography, such as access to the sea or climate, is also an underlying determinant of comparative advantage. Thus, the relevance of location is given exogenously and such a model cannot endogenously explain agglomeration. Enter the new trade theory.

The new trade theory explains that trade can take place without differences in endowments or technology. Krugman (1980) showed that when two localities have the same endowments and technology and each one firm producing, for example cars, trade will take place because various types of cars are produced. The varieties of cars are imperfect substitutes and consumers also prefer more varieties of a car to fewer. Opening up to trade will then enlarge the size of the market for each type car, which means that production per variety can increase and firms can better exploit increasing returns. Thus, the reason for trade is a combination of increasing returns to scale at the firm level and the love-of-variety effect in consumers' preferences, which is an externality not taken into account by firms (Brakman *et al.*, 2001:42). The gains from trade are that the increased scale of production decreases prices and this result in a higher real wage for workers/consumers. Also, the

consumers are able to consume a greater variety of products than under autarky. It is important to note though, that the increasing returns to scale in the model require a market structure of imperfect competition. In this, the Dixit-Stiglitz model of monopolistic competition forms the backbone of the new trade theory (see Dixit & Stiglitz, 1977).

In this early version of the new trade theory (Krugman, 1979), the model still has no role for the location of production in trade – trade costs are absent and the market size is set exogenously, evenly distributed between the two localities. This means that the firms are indifferent about the location of production and there can be no agglomeration of economic activity. Brakman *et al.* (2001:45) argued that the model is nevertheless important for its analysis of producer and consumer behaviour.

In 1980 Krugman extended the new trade model. The gains from trade were made to be completely due to the love-of-variety effect. Secondly, transport costs were added to the model, which are important from the point of view of location. Most importantly, the so-called "home-market effect" was introduced. This allows for an uneven distribution of market size, which, along with the positive transport costs implies that firms will produce those varieties for which home demand is relatively strong. This is strongly linked to increasing returns. In this version of the model the location of production is now important and the concentration of economic activity can be an outcome.

Krugman (1980) did not, however, offer an encompassing explanation of the location of production in space. Three shortcomings remain: Firstly, the model does not allow for the mobility of firms or factors of production. This means that there is no decision about location, only about the varieties to produce; Secondly, the concentration of production of varieties does not allow for agglomeration – rather, the outcome of the model is that both localities are characterised by a geographic concentration of industry; Thirdly, the allocation of the market size for the varieties is given exogenously. Thus, location of economic activity is still determined outside the model.

Krugman and Venables (1990) further extended Krugman (1980) by allowing countries to differ in size. The interesting results lie in the way that a fall in transport costs (representing an increase in the degree of economic integration) affects the localities that start with a larger (the core) or smaller (the periphery) number of firms in the manufacturing sector. The model shows that when transport costs are high and prohibit trade, both localities have a share in the manufacturing sector that is equal to their share in world

endowments. For an intermediate range of transport costs, firms enter into the larger market, even though nominal wages are higher. If transport costs should continue to fall, the advantage of producing in the larger market becomes smaller. Also taking into account the stiffer labour market competition in the core, new firms will find it beneficial to start production in the periphery where wages are lower. In the extreme case of zero transport costs, nominal wages will be equal and each locality's share of manufactures will return to its share in world endowments (Krugman and Venables, 1990).

Although this model allows for the agglomeration of economic activity, it is still not a complete explanation of what determines economic activity and its growth across space. It assumes that market size differs and that the existence of the core and periphery is not derived from the model itself. Nevertheless, even more than the contributions of the other theories, trade theory serves as the basis for explanations of agglomeration. It provides a coherent microeconomic structure for the behaviour of producers and consumers, while showing external economies, market size and transport costs to be important determinants of location and growth.

In conclusion it is clear that there are wide-ranging explanations for the location of production in space. Recently, these well established spatial insights from urban economics, regional economics, growth theory, development economics, and trade theory have been drawn together in a general equilibrium framework of mainstream economic theory. In geographical economics, many of the shortcomings of the above theories in explaining the location of production in space are addressed. In the core model of geographical economics there is interdependence between location and economics, and the equilibrium location of economic activity is the result of the underlying behaviour of economic agents. The core model is discussed in the following section.

2.3 The core model of geographical economics

Geographical economics is specifically concerned with explanations as to what determines the location of production in space (Krugman, 1991). Also sometimes called the "New Economic Geography", it was launched by Krugman in a 1991 article in the Journal of Political Economy (see Krugman, 1991).

The new economic geography is not "new" in the sense that it provides radically different explanations of the location of production in space (see for example Martin, 1999 for

criticism), in fact it draws on the insights of all the fields discussed above. For example, the mechanics of the core model of geographical economics owes a burden of debt to Krugman's new trade theory. The new economic geography is however "new" in the sense that it extends the insights from urban economics, regional economics, development economics, growth and trade theory, by giving the behavioural underpinnings. That is, geographical economics provides a microeconomic foundation for understanding economies characterised by regional specialisation, cities, and trade by appealing to nature (the unevenness in the distribution of resources), to non-market institutions (such as externalities that give rise to endogenous spatial inhomogeneities), and to an imperfectly competitive paradigm (Fujita & Thisse, 2002:45).

Before explaining the mechanics of the core model it may be useful first to explain its inspiration. The leap from Krugman's new trade theory to his new economic geography came from outside the theories outlined above and it was in the form of Michael Porter's *Competitive Advantage of Nations*.

2.3.1 Porter's clusters and the new economic geography

Brakman *et al.* (2001:321) quoted Krugman on the origins of the core geographical economic model: "*Michael Porter had given me a manuscript copy of his book on Competitive Advantage of Nations, probably late 1989. I was much taken by the stuff on clusters, and started trying to make a model. ... after a few days I realised that my home-market stuff basically provided the necessary.*". This section outlines Porter's ideas that inspired the new economic geography.

Porter (1998) phrased the ideas of first-nature geography, external economies, distance and market size with an emphasis on competitiveness leading to growth. He contends that globalisation has changed the importance of the location of production in space. National and international markets for factors are more efficient and competition is less factor-intensive. Modern, flexible technologies are often less scale intensive and are coupled with outsourcing and close relationships with suppliers. There is earlier access to huge foreign markets. This has diminished the importance of factor endowments and the size of local markets. Porter (1998) argued that location now affects competitive advantage through its influence on productivity and productivity growth. This occurs within the context of clusters.

A cluster is a critical mass of producers in a particular field, in a particular location - the result of the process of localisation. Porter (1998) discussed a number of ways in which location affects the business environment, competition, and growth. These are, to a large extent, the reasons for localisation that Krugman later advanced (1991, 1995, 1998). The first influence is that of the factor conditions of a location. These range from basic inputs such as physical infrastructure, to information. In clusters, the proximity of producers leads to spillovers that improve factor conditions. This improves the flow of information and the success of innovation. This is also true for related and supporting industries. They provide specialised inputs and information, and facilitate complementarities amongst firms. Materials, components, machinery and services are supplied more efficiently and at lower cost when producers are concentrated in a particular locality. Location plays a role in firm strategy and rivalry. Rivalry that involves imitation and differentiation improves competition and leads to growth. When firms are clustered together, the rivalry is more intense. In the final instance, the demand conditions in a locality may influence the business environment, competition, and growth. Sophisticated and demanding customers at home press firms to improve and differentiate. A cluster may provide such a group of customers.

Thus, Porter (1998) saw the enduring competitive advantages in a global economy as localised. Advantages arise from concentrations that follow from highly specialised skills and knowledge, institutions, rivals and sophisticated customers in a particular region or locality. Proximity allows special access and relationships, better information and other advantages in productivity and productivity growth. Agglomeration is seen as the primary long-run source of economic growth and prosperity.

Again, this is a story of economies of scale, localised spillovers and transport costs. Krugman (1991) formalised this to set up the core model of geographical economics.

2.3.2 Explaining the core model

In the core model of geographical economics the spatial configuration of economic activities is explained as the outcome of a process involving two types of forces, namely, agglomeration (or centripetal) forces and dispersion (or centrifugal) forces (Fujita & Thisse, 2002:5). The propensity to agglomerate comes from economies of scale and transport costs. The mechanisms through which the agglomeration takes place are labour mobility and/or inter-industry linkages (Neary, 2001).

It is useful to explain the working of these forces in a simple two-sector model. The question is, what drives agglomeration? Or, in terms of the two-sector model, what makes each manufacturer want to serve the national market from a single location? The structure of the core model is laid out in Krugman (1991) (see also Brakman *et al.*, 2001:64-65, Neary 2001, and Ottaviano & Puga, 1998 for descriptions of the model).

There are two regions in the economy, and two sectors, manufacturing and food. In each region there are farm workers and manufacturing workers. The farm workers earn their income by working for the farmers in their region. The farmers in region one use the labour services of the farm workers from region one (they also hire themselves) to produce food under constant returns to scale and perfect competition. The food is sold to consumers in region one or two. It is assumed that there are no transport costs for food.

The manufacturing sector consists of N_1 firms in region one and N_2 firms in region two. Each manufacturing firm produces a differentiated product, using only labour under internal economies of scale. The manufacturing workers earn the manufacturing wage rate by supplying labour to the firms in the manufacturing sector of their region. The internal economies of scale mean that the firms determine the price of their product using monopolistic power. To sell a manufactured good in another region incurs transport costs – thus, firms will charge a higher price in the other region than they do at home.

The consumers spend their income on food and manufactures. The food is a homogenous good without transport costs and it fetches the same price in each region, which means that the consumers are indifferent to whether the food is produced in region one or two and the farmers earn the same wage in both regions. But the consumers allocate their spending on manufactures over the many varieties produced in regions one and two. Other things being equal, consuming imported varieties is more expensive than consuming domestic varieties, but since it is assumed that consumers have a liking for variety they will always consume at least some units of all varieties produced.

Setting the economy up this way creates a propensity for agglomeration. The internal economies of scale mean that increasing production at a plant would lower cost – and fragmenting production over more than one location is costly. Manufacturers will thus be inclined to produce more at a single location. But, producing in a single location only, has to be weighed up against transport costs. As set up above, selling in a region other than at home incurs transport costs and means having to charge a higher price. A manufacturer

would thus choose location in order to maximise the cost savings from large-scale production and to minimise transport costs. This, however, creates only a propensity to agglomerate. The mechanism through which the agglomeration takes place may firstly be labour mobility.

When manufacturing workers can relocate from region one to region two or vice versa, the following thought experiment is possible (following Neary, 2001:542 and Ottaviano & Puga, 1998:713): Start with an equilibrium where both regions are *a priori* identical in every respect (symmetric diversified equilibrium) and assume that for some reason a single new manufacturing firm enters in region one and ask how this affects the incentives for further entry or exit. The presence of one more firm will increase competition in the product and labour markets of region one. The competition in the product market lowers the industry price index, which reduces the demand facing each existing firm. If profits in region one fall relative to region two, the firm will exit and the initial equilibrium will be restored. This would be the end of the story if there were no migration. However, the rise in the number of local varieties and the rise in the labour demand and wages will attract more workers. Thus, manufacturing workers will relocate to region one. This creates a demand or backward linkage whereby local expenditure is increased. There is also a forward or cost linkage that eases competition in the labour market – the lower price index lowers the cost of living for workers and raises the real wages in region one but the equilibrating migration means that nominal wages must fall, reducing costs. Overall, local profits increase and more firms are attracted.

Brakman *et al.* (2001:66-94) showed that agglomeration or spreading (such as the initial symmetric diversified equilibrium) may be driven by three important parameters: the elasticity of substitution between manufactured varieties, the proportion of income spent on manufactures, and transport cost (or more formally, the fraction of manufactured goods that does not arrive at the destination when goods are shipped between regions). These parameters determine agglomeration as follows.

Firstly, the greater the proportion of income spent on manufactures, the greater the demand linkage whereby entry of a new firm induces migration, which raises demand and encourages further entry. Secondly, the greater the budget share of manufactures, the greater is the cost linkage whereby entry of a new firm lowers the cost of living and encourages further migration and entry. Thirdly, the lower the elasticity of substitution across varieties, the greater the importance of having a large variety of products available

locally – this reinforces the monopoly power of manufacturers over their own varieties, which weakens local competition and favours agglomeration. However, these agglomeration forces stand opposed to the dispersion force of transport costs. For sufficiently high transport costs imports are always so expensive that home-production is always possible. This means that production will rather be spread out. For sufficiently low transport costs the two regions are *ex ante* identical and again no agglomeration is possible. Thus, only somewhere in between there is a threshold level of trade costs where the incentives to agglomerate and labour mobility lead to agglomeration.

As stated above, labour mobility may not be the only mechanism through which agglomeration takes place: the other possibility is inter-industry linkages. The modelling is similar to that of migration and is explained by Neary (2001:545) as well as by Ottaviano and Puga (1998:715). The argument is that firms like to be close to each other because of direct input-output linkages amongst themselves. Such linkages give rise to pecuniary externalities if there are increasing returns to scale. That means that a simple buyer-supplier relationship between a downstream and upstream industry is not enough – it must be the case that an increase in the output of the downstream industry enlarges the market for the intermediates it uses and thus induces the upstream industry to produce at a more efficient scale. Where such linkages exist, they become a mechanism through which agglomeration takes place.

Thus, conditions can be set for agglomeration, namely sufficiently strong economies of scale; transport costs in an intermediate range; and a sufficiently large share of "footloose" production[2]. When these conditions are met, each manufacturer wants to serve the national market from a single location. To minimise transport costs, he chooses a location with large local demand. But local demand will be large precisely where the majority of manufacturers choose to locate[3]. The localisation, and in a larger scale, agglomeration leads to growth. The core model of geographical economics integrates these insights into a general equilibrium framework of the location of economic activity.

The working of the model can best be illustrated by means of a numerical example.

---

[2] Production not tied down to other locations either by the need to be very close to the consumer, or by the need to use natural resources very close to their source.

[3] This explanation may underplay the importance Krugman attaches to concentration as the result of demand externalities. His numerical explanation better highlights this, and the role that history plays in location. However, referring to empirical studies, Krugman (1998) later speculates that market size effects may be a less important source of agglomeration than other kinds of external economies.

2.3.3 A numerical example of the core model

Suppose that the two regions are now called north and south and a firm can reside in either the north or south. The aim is to determine the location decision of firms under the following assumptions. The total demand for each variety of manufactures is exogenous. It is assumed that each firm sells four units to workers in the manufacturing sector and six units to farmers. Thus, the total demand for each variety is 10. Agricultural production, and thus the demand it generates, is location-specific. It is assumed that of the six units sold to farmers, four units are sold in the north and two are sold in the south. These immobile workers ensure that there is always a positive demand in both regions. The model determines the location of manufacturing production. For this the assumption is that the transport costs between north and south are R1 (one Rand) per unit. Given that this makes the consumption of imported varieties more expensive, firms will choose locations to minimise transport costs.

Brakman *et al.* (2001:61) used tables to explain the location decision. These are reproduced below and three possibilities are given: all the firms in the north, all the firms in the south, or 25 per cent of firms in north and 75 per cent of firms in the south. Sales in each region are equal to sales to workers in manufacturing and sales to farmers.

Table 2.1: Geography of sales

	Sales in North	**Sales in South**	**Total sales**
All firms in North	4+ 4 = 8	0 + 2 = 2	10
All firms in South	0 + 4 = 4	4 + 2 = 6	10
25% of firms in North, 75% of firms in South	1 + 4 = 5		10

The next step is to calculate the transport costs as a function of a firm's location decision, given the location of the other firms. This is shown in Table 2.2.

Table 2.2: Transport costs

	If the firm locates in the north	**If the firm locates in the south**
If all other firms are in the north	0 + 2 = 2 (to farmers in the south)	4 + 4 = 8 (to workers and farmers in the north)
If all other firms are in the south	4 + 2 = 6 (to workers and farmers in the south)	0 + 4 = 4 (to farmers in the north)
If 25% of firms are in the north and 75% of firms are in the south	3 + 2 = 5 (to workers and farmers in the south)	1 + 4 = 5 (to workers and farmers in the north)

The table shows that the transport costs for a firm locating in the south will be R8 (eight Rands), namely R4 for sales to farmers in the north and R4 for sales to workers in manufacturing in the north. In the same way, if the firm were to locate in the north, its transport costs would only be R2 (two Rands) for sales to farmers in the south. Thus the firm will minimise transport costs by locating in the north if all other firms are also located in the north. In the second row, Table 2.2 shows that the firm will locate in the south of all other firms are also located in the south.

2.3.4 Characteristics of the core model

From the explanation of the working of the core model and the example it is possible to conclude that the geographical economics approach to explaining the location of production in space has a number of distinctive characteristics.

The first is that the combination of increasing returns and transport costs generate a home-market effect similar to that of Krugman's trade models: The region with the higher demand has a proportionately larger share of manufacturing. The strength of the home-market effect depends on trade costs. Lower trade costs (in the intermediate range) magnify the effect whereby an exogenous change in the location of upstream demand leads to a more than proportional change of downstream supply in the same direction (Ottaviano, 2003:668). Thus, lowering trade costs, or increasing trade freeness, tends to make firms more footloose.

Second, the model is characterised by endogenous asymmetry. That means that starting with two symmetric regions and high trade costs, a gradual increase in trade freeness eventually produces regional asymmetries. This is significant because it makes it

possible to explain spatial inequality independently of first-nature geography. Such asymmetries are evident in the home-market magnification explained above and in spatial wage structures where higher real wages are paid in agglomerations.

Third, the model may have multiple equilibria. That means that it is not possible to determine beforehand where agglomeration will occur – rather, it depends on initial conditions.

The fourth characteristic is that the model creates the possibility of cumulative causation. That is, if one location has attracted more manufacturing firms than the other, another new firm has an incentive to locate where the other firms are. This also means that, other things being equal, an increase in local income would lead to a higher increase in demand than the same income increase in the other region (Brakman *et al.*, 2001:61-62). Thus agglomeration forces are self-enforcing.

Self-fulfilling expectations follow from cumulative causation. Ottaviano (2003:669) argued that in the presence of cumulative causation the shared belief that all firms will cluster in a certain region is self-rewarding and thus becomes self-fulfilling.

Finally, the possibility of cumulative causation and self-fulfilling expectations leads to two further possible outcomes that characterise the model. The first is the so-called catastrophic agglomeration where, once the break point in trade costs is reached, a small increase in freeness of trade triggers a process whereby the only long-run outcome is agglomeration. The second is locational path dependency where a transitory shock can have permanent effects – moving the majority of firms to one region where agglomeration forces then lock them in (Ottaviano, 2003:669).

The following section draws on all of the above explanations to identify testable hypotheses of the determinants of the location of economic activity and its growth across space.

2.4 Testable hypotheses

Drawing on the above theoretical overview of the determinants of agglomeration and spatial economic growth, this section distils a number of testable hypotheses for factors that determine the location of economic activity and the its growth across space. The question, specifically, is what is the role of geography?

Table 2.3 classifies the determinants into those that are expected to be positively and negatively related to agglomeration and spatial economic growth.

Table 2.3: Classification of determinants

Agglomeration forces	Dispersion forces
Rosenstein-Rodan	
A large market offers economies of scale (emphasises the 'big push' needed to overcome the problems of the periphery's small market).	
Myrdal	
External economies of scale due to spillovers (emphasises cumulative causation)	
Hirschman	
Backward and forward linkages between firms.	
Henderson:	
External economies of scale due to spillovers associated with: Information sharing. Pooled labour market. Existence of specialised suppliers.	Transport costs. Land rents.
Central place theory:	
Internal economies of scale.	Transport costs.
Market potential approach:	
Large demand	Distance.
Neo-classical growth theory:	
Differences in the determinants of growth that are location-specific. First-nature geography that gives a cost advantage, e.g. proximity to large markets, or access to the ocean that lowers transport costs.	Differences in the determinants of growth that are location-specific. First-nature geography that gives a cost disadvantage, e.g. long distances to large markets or being landlocked, which raise transport costs.
New growth theory:	
External economies of scale due to localised spillovers associated with the endogenous determinants of growth: Human capital. Research and development. Infrastructure.	
Neo-classical trade theory	
First-nature geography: uneven distribution of endowments determines comparative advantage.	

Agglomeration forces	Dispersion forces
New trade theory	
Market size and consumers' love of variety allow manufacturers to achieve internal economies of scale (traded-off against transport costs, this creates the home-market effect).	Transport costs.
Geographical economics	
Internal economies of scale (traded-off against transport costs) create the propensity to agglomerate. The mechanisms through which agglomeration takes place are labour mobility and inter-industry linkages.	Transport costs.

In sum, it is possible to say that one would expect a positive relationship between the location of economic activity and its growth across space, and economies of scale. The economies of scale may be internal or external and result from the size of the market; spillovers from knowledge sharing, learning and imitation; spillovers from infrastructure, productivity gains from the diversity of intermediate inputs or the matching process on the labour market (Fujita & Thisse, 2002). Conversely, one would expect a negative relationship between the location of economic activity and its growth across space, and distance from agglomerations, or transport costs.

As to the first question of what determines agglomeration, policies, institutions or geography, it is clear the geography plays a role. First-nature geography can act as an agglomeration or dispersion force. The above theories have also shown that externalities can give rise to endogenous spatial inhomogeneities such as agglomeration and thus there is a role for second-nature geography.

Chapter 3 will explain the methodologies of how one tests the determinants of economic activity and its growth across space.

2.5 Criticism and caveats

Before proceeding to the empirical methodologies, it is prudent to take note of the criticism of the above explanations of the determinants of spatial economic growth, particularly the concerns concerning geographical economics. Brakman *et al.* (2001:318-320) identified

objections from three quarters: economic geography, regional and urban economics, and international economics.

From economic geography, Martin's (1999) main objection was to the neo-classical framework of geographical economics. He had doubts about what he called the straitjacket of the mathematical mainstream, and the neglect of real places.

Martin (1999) argued that the neo-classical framework of new economic geography models limits it to examining those aspects of economic development and spatial agglomeration that can be mathematically modelled. This leads to the marginalisation of social, cultural and institutional factors in the analysis (Martin & Sunly, 1996). Furthermore, the focus on mathematical modelling results in misrepresentations of the notions of "history" and "path dependence" (Martin, 1999). In new economic geography Krugman (1991, 1995, 1998) contended that "history matters"; in other words, arbitrary initial conditions and accidental events may set in motion particular patterns of industrial development that are subsequently locked-in via self-reinforcing effects. Detractors, however, feel that the models use a notion of abstract, logical or simulation time. The "history" is not real or context-specific periods of time over which actual spatial agglomerations evolved. Likewise, the "lock-in" or "path dependence" of agglomeration is only a solution sequence, determined by specified initial conditions and parameters of the model. The real, complex, localised socio-historical process of technological, institutional and social evolution is disregarded (Martin, 1999). This coincides with criticism of new economic geography's neglect of "real" places. "Regions" and "locations" are often just theoretical points, not real communities. This is felt to limit the understanding of the space economy. Martin (1999) argued that agglomeration also has a place-dependent character. Over time, different places develop different institutional forms and networks of inter-firm dependencies along with economic, cultural and political practices that influence the role that different places play in phases of economic development.

Compared to economic geography, the field of regional and urban economics lies much closer to geographical economics in that it uses the same toolkit of the neo-classical framework. Thus, Nijkamp's (2001) criticism of geographical economics is of its content. In his view, geographical economics:

- Has too narrow a view of geography through its reliance on iceberg transport costs and thus neglects broader transaction costs (e.g. mental distance).

- Pays no attention to spatial competition among firms.
- Pays little attention to the role of competition.
- Relies too much on numerical simulations.

Urban economists, in the form of Henderson (2000), also criticise geographical economics for its simple depiction of urban economic geography and the reliance on simulations.

Finally, there is the criticism from international economics. Neary (2001) argued that geographical economics is close to existing and well-established trade theories, but still highlights a number of weaknesses:

- A lack of analysis of individual firms.
- The use of iceberg transport costs and one-dimensional space that makes the analysis of geography too simple.
- The reliance of specific functional forms and simulations.
- A lack of empirical evidence.

In sum, it may be said that these criticisms have, until recently, lead to cautious application of geographical economics to policy issues (see Ottaviano, 2003).

The criticism in itself is, however, not damning. Neary (2001:553) conceded that geographical economics has made a persuasive case for extending the insights from other fields by giving them behavioural underpinnings. Brakman *et al.* (2001:322) argued that the critics focus too strongly on the shortcomings of the original core model (Krugman, 1991). Since then, great strides have been made with theoretical extensions and empirical evidence.

In conclusion, the criticisms of the geographical economics framework should act as caveats to its empirical and policy applications. In this study, for example, the complexities of the location of economic activity in space are taken into account in terms of history and institutions through a discussion of the origins of spatial inequality in South Africa.

2.6 Summary

The purpose of this chapter was to review the theories of the spatial structure of economic activity, and the determinants of spatial economic growth, so as to provide a basis from which to later derive empirically testable hypotheses. The focus was specifically on the so-

called core model of geographical economics. Geographical economics is, however, not the first effort to explain the spatial distribution of economic activity or of the determinants of spatial economic growth – Section 2.2 also reviewed the contributions from urban economics and regional economics, economic growth theory and the development economics literature. Following that, the core model of geographical economics was presented in Section 2.3.

A number of conclusions can be drawn. The first is that agglomeration is about economic growth in space. Agglomeration/growth is positively related to increasing returns and negatively related to transport cost. The increasing returns are the result of scale and proximity that impart the efficiency benefits of mass production, specialised intermediate inputs, better matching between workers and jobs, and the availability of public goods and services. Within this framework, the specialisation drives trade. Trade, also, is positively related to increasing returns and negatively related to transport cost.

Finally, critics caution that empirical analysis and policy applications should aim to take account of history and institutions when explaining the spatial structure of economic activity and determinants of spatial economic growth.

The following chapter examines the empirical methods that have been used to test the determinants of spatial economic growth.

Chapter 3: Empirical methodology

3.1 Introduction

Chapter 2 presented a theoretical overview of the determinants of spatial economic growth and concluded by identifying a number of testable hypotheses of the factors that determine the location of economic activity and its growth across space. It was predicted that the exogenous characteristics of locations – first-nature geography – influence where agglomerations occur. Second, it was argued that economies of scale can give rise to endogenous spatial inhomogeneities such as agglomeration. Theories put forward in urban economics, regional economics, growth, trade and development economics showed that agglomeration may occur because firms benefit from being near other firms in their own industry; because workers benefit from being near other workers; or because firms benefit from being near large consumer and industrial markets. This would indicate a positive relationship between the location of economic activity (as well as its growth across space) and economies of scale. It would also indicate a negative relationship between the location of economic activity (as well as its growth across space), and distance from agglomerations (or transport costs). Thus, the models explain a spatial distribution of economic activity that is lumpy and that across regions nominal wages will be positively correlated with the agglomeration of economic activity (Hanson, 2000:7).

Empirical studies of the determinants of spatial economic growth may be grouped into three strands (Brakman *et al.*, 2001:128). The first is that of regional and sub-national growth studies that test the determinants of spatial economic growth. This approach is often related to the market-potential approach. The second strand is studies that focus on cities as centres of economic activity. The main questions here relate to city size distribution and optimal city size. The third strand involves the testing of the predictions of the core model of geographical economics. There is currently only a small amount of literature available from the third strand, but review papers of the field by Hanson (2000), Neary (2001) and Overman, Redding and Venables (2001) are characteristic of an emerging field[4]. The

[4] The current benchmark research was presented in 2004 at a conference entitled "*New Economic Geography – Closing the gap between theory and empirics*". The conference was organised by Professors Brakman, Garretsen and Van Marrewijk at the Hamburg Institute of International Economics

purpose of this chapter is to give an overview of the different empirical methods used to test geographical economics and to focus on the methods that will be used later in this study.

The chapter is structured as follows: Section 3.2 provides a brief overview of the empirical literature introduced above. The focus is on ways in which the hypotheses from geographical economics can be tested. Section 3.3 then sets out the empirical approach to be used, particularly in Chapter 5. Section 3.4 concludes.

3.2 Overview of the empirical literature

This section follows reviews by Hanson (2000) and Overman *et al.* (2001). Empirical approaches may be classified by whether they focus on:

- Determinants of the spatial variation of per capita income.
- Determinants of the spatial variation of wages and productivity.
- Determinants of the spatial variation of industry employment and production.
- Transport costs.
- Determinants of the size and structure of cities.

The aim is not to present an exhaustive survey, but to focus on the work that typifies the methodological issues faced in this field.

3.2.1 Determinants of the spatial variation of per capita income

Overman *et al.* (2001:11) grouped studies of the determinants of the spatial variation of per capita income under three broad headings: First-nature geography, second-nature geography, and the effects of social infrastructure. Each of these three threads will be discussed in turn.

Firstly, studies that use first-nature geography as an explanatory variable relate climate and endowments to variations in per capita income across countries. For example, the work of Gallup, Sachs and Mellinger (1999), Gallup and Sachs (2000), Sachs and Warner (1999) and Sachs (2001) had a focus on first-nature geography. Gallup *et al.* (1999) found that countries with a large proportion of their population close to the coast, with low levels of malaria, large hydrocarbon endowments and low levels of transport costs, have higher levels of income per capita.

Related to this style of growth work, studies that attempt to quantify the significance of social infrastructure as a determinant of the spatial variation of per capita income, have used geographical variables as proxies. For example, Hall and Jones (1999) modelled social infrastructure as a function of the distance from the equator, openness to international trade, a fraction of the English-speaking population of a country and the fraction of population speaking a European language. Their finding was that these four variables explain 41 per cent of the cross-country variation in social infrastructure and 60 per cent of the cross-country variation in income.

The third thread of empirical research into the spatial variation of per capita income examines endogenous spatial inhomogeneities. Specifically, access to markets is seen as a determinant of per capita income (Overman *et al.*, 2001:12).

The market access approach draws on Harris' (1954) market potential equation that was explained in Section 2.2.2. Empirical tests of whether access to markets is a significant determinant of the spatial variation of per capita income uses distance-weighted variables to proxy for market access in growth regressions. Hummels (1995) (as quoted by Overman *et al.*, 2001), for example, calculated distance-weighted GDP for a number of countries. Another measure used by Hummels (1995) for market access is that of a country's distance from the three main centres of world economic activity (the United States, Japan and Germany). Another example is Leamer's (1997) examination of the importance of access to western European markets for post-reform income per capita growth in Eastern Europe – he used a distance weighting parameter derived from estimation of a gravity equation.

The market access approach is related to the openness of economies. Frankel and Romer (1999) used first-nature geography measures, such as distance, area, landlockedness and population, as instruments for bilateral trade flows. They found a positive relationship between cross-country variations in per capita income and openness.

There are also studies that construct more theory-consistent measures of market access. Redding and Venables (2000) constructed a market access variable of country *i* that comprises expenditure in each market *j*, together with the price index and it is adjusted according to the transport costs from *j* to *i*. Since the terms of the expression are unobservable, they are derived from gravity estimation. The market access variable is then used to explain cross-country variation in per capita income.

Chapter 2 showed that proximity to a large market is one of the key determinants of agglomeration. It also showed the home-market effect to be a significant characteristic of the location of production in space explained by geographical economics models. Consequently it is also tested for as a determinant of the spatial variation of wages, as well as of industry employment and production.

3.2.2 Determinants of the spatial variation of wages and productivity

Like the home-market effect, the spatial variation of wages, also called the spatial wage structure or spatial wage gradients, is one of the notable results of agglomeration. There is also considerable empirical literature wherein the variation of nominal wages is explained in terms of first-nature geography, education and spatial linkages.

First-nature geography may be invoked to explain regional differences in wages. The argument is that if consumers prefer warm, sunny weather to cold, cloudy weather, they will require a wage premium to live and work in cold, cloudy places. But if good weather also increases productivity, employers in the warm sunny places will pay even higher wages. Migration to the high-wage localities will capitalise the productivity differences into land rents. Empirically this is tested by using a hedonic pricing estimation. The dependent variable is the wage of individual workers and the explanatory variables include characteristics of the workers as well as the exogenous characteristics of locations. Roback (1982), found that metropolitan area labour earnings in the United States are positively correlated with local measures of the crime rate, air quality, heating degree days, snowfall and the number of cloudy days. It is also possible to explain the spatial variation of wages in terms of localised human capital spillovers – an explanation akin to Marshall-Arrow-Romer's view of knowledge sharing, or the Smith-Marshallian explanation of the benefits of a thick labour market. Eaton and Eckstein (1997) and Black and Henderson (1999) predicted that wages will be higher in localities with larger stocks of public knowledge or, in other words, with greater localised human capital spillovers. Empirical studies again estimate hedonic price equations for wages. Rauch (1993) found that in U.S. metropolitan areas workers appear to be more productive where their co-workers are more educated and more experienced[5]. Controlling for the possible endogeneity of average regional education levels, Moretti (1999) found a similar

5 Localised human capital spillovers may also be an important explanation of the variation in growth rates across localities. Glaeser, Scheinkman and Shleifer (1995) and Black and Henderson (1999) find that population growth is faster in U.S. metropolitan areas that begin with a more educated population.

result when regressing individual wages on the proportion of college graduates on the labour force of a city in which an individual works.

Finally, spatial linkages may also be a determinant of the spatial variation of wages and productivity. Such an explanation is closely linked to the external economies that determine agglomeration in geographical economics models. A number of empirical studies have examined proximity to agglomerations (i.e. market access) and spatial linkages as determinants of a spatial wage structure. Ciccone and Hall (1996) examined whether labour productivity is higher in U.S. states where economic activity is more densely concentrated. They draw on Marshall-Arrow-Romer's view of knowledge sharing and the Smith-Marshallian explanation of the benefits of a thick labour market, and estimate an equation in which state labour productivity is a function of the education level of workers and employment densities. They find that close interaction of workers raises productivity. The spatial linkages may also extend beyond the labour market to market potential. Hanson (1997, 1998), for example, tested for the presence of spatial demand linkages by looking at the proximity of markets as a determinant of wage rates across states in Mexico and counties in the U.S. He found that wages are higher in locations that are closer to large consumer markets.

To conclude, *ceteris paribus*, locations that are characterised by adverse exogenous amenities, that have small stocks of public knowledge and are remote from markets and supplies of intermediate inputs, have lower nominal wages (see Overman *et al.*, 2001:18).

3.2.3 Determinants of the spatial variation of industry employment and production

In Chapter 2, regional economics, development economics and geographical economic models proposed explicit reasons why firms would prefer to locate close to other firms. Among the factors that make industry clusters attractive are industry-specific spillovers that raise factor productivity as well as forward and backward linkages that help firms to obtain inputs or deliver goods to market at relatively low cost (Hanson, 2000:16). This section examines the empirical evidence of industry clustering as a determinant of the spatial variation of industry employment and production.

A first explanation of the patterns of industrial specialisation that determine the spatial variation of industry employment and production comes from trade theory in the form of the Heckscher-Olin model. This goes back to first-nature determinants of location

whereby localities specialise in industries that use intensively the abundant factors of those localities. An example of empirical evidence of the importance of factor endowments is Kim (1999) who found that factor supplies account for a large share of the cross-sectional variation in output levels across U.S. states.

The second line of enquiry is to examine the acquired characteristics of a location as a determinant of the spatial variation of industry employment and production. Empirical work focuses on the attractiveness of industry clusters. Carlton (1983) found that when controlling for local labour costs, energy costs and tax rates, that new plants in the United States are more likely to choose a city where its own industry employment is larger.

Related work examines factor productivity at a regional level as a function of own-industry scale and regional population. Nakamura (1985) found that in Japanese cities value added per worker is increasing in the scale of local industry output for most capital-goods industries. Henderson (1986) found that gross output per worker is increasing in local industry employment for capital goods industries in Brazilian cities, and for capital and consumer goods industries in U.S. cities.

Another way in which industry linkages can determine the spatial variation of industry employment and production is through a market-size or home-market effect. An empirical test of the importance of demand linkages found that regional industrial production across Japanese prefectures increased more than one to one with regional absorption of an industry's output (Davis & Weinstein, 1999). This indicates that there is an excess concentration of production in regions where the demand for a good is relatively high.

The effect of agglomeration economies on industry location can also be identified by focusing on localisation and urbanisation economies. For example, Glaeser *at al.* (1992) examined employment growth in a city industry in the U.S. and found that it is positively correlated with the initial diversity of industry employment in the city, but not with initial own-industry employment in the city. In a similar vein, Henderson (1999) explained the output growth in U.S. manufacturing plants of machinery and high-tech industries at five-year intervals over the period 1963 to 1992. Controlling for plant inputs, time fixed-effects and plant fixed-effects he finds that plant outputs are (a) positively correlated with the number of own-industry plants currently in the same county (with the effect stronger for high-tech output than for machinery), (b) positively correlated with the lagged number of

own-industry plants in the same county (high-tech only), and (c) uncorrelated with city industrial diversity or total city manufacturing employment.

To conclude, it appears that the spatial variation of industry employment and production is determined by a relatively abundant supply of workers that firms are likely to hire and by the proximity of firms in upstream and downstream activities.

3.2.4 Transport costs

Transport costs play a key role in the theories of the location of production in space, but in the empirical literature there are no direct tests for transport costs as a determinant of spatial economic growth. Instead, the significance of transport costs is tested indirectly as one of the causes of the home-market effect. Empirical work on transport costs has rather focused on the determinants of transport costs, also extending it to the broader concept of trade costs (Overman *et al.*, 2001:7-10).

Limao and Venables (2001) estimated the determinants of transport costs. They find that infrastructure quality is important, that that overland distance is around seven times more expensive than sea distance, that sharing a common border decreases transport costs, but that being landlocked increases transport costs by approximately 50 per cent. The study finds elasticities of transport costs with respect to distance of between 0.2 and 0.3.

Trade costs include transport costs, but also a wider package of cultural and political determinants as well as policy measures. This is more problematic to study empirically. Hummels (2000) (as quoted by Overman *et al.*, 2001), for example, estimated the cost of time in transit and found that the cost of an extra day's travel is around 0.3 per cent of the value shipped and this goes up to 0.5 per cent for manufacturing sectors. In a different approach to the cultural and political determinants of trade costs, Rauch and Trindade (1999) used gravity estimation to examine the role of ethnic Chinese networks in promoting trade. They found that colonial ties and Chinese networks have large significant effects in promoting trade.

Overman *et al.* (2001:10) argued that further empirical work is needed in the field of transport costs and the role of geography. These areas include looking at borders as trade barriers, as well as understanding the geography of information flows.

3.2.5 Determinants of the size and structure of cities

The final category of empirical literature is concerned with the determinants of the size and structure of cities. The focus is on cities themselves as centres of economic activity and the empirical work is more closely related to issues examined in urban economies: specialisation, localisation, city-size distributions and questions of optimal city size. Since urban economic dynamics may be an important determinant of spatial patterns of growth and development, an overview of the empirical work on the determinants of the size and structure of cities is given in this section.

Firstly, there are descriptive studies with some empirical measurement. Overman *et al.* (2001:20) pointed out that such studies basically ask two questions: how localised is a particular economic activity and how specialised is a particular geographical unit? Empirical work then typically consists of estimating location quotients. For the U.S., Kim (1995) found that industries became increasingly localised and states increasingly specialised up to 1930. After that, state specialisation fell substantially and is lower today than it was in 1860. Midelfart-Knarvik *et al.* (2000) examined the E.U. and found that countries are becoming increasingly specialised in the face of European integration.

Extension of this localisation and specialisation work aims to explain so-called excess localisation: when there is greater localisation than one would expect relative to activity as a whole at a location and relative to random location of an industry's plants. Ellison and Glaeser (1999) considered how much localisation can be explained by advantages from first-nature geography and found that at the U.S. state level between 50 per cent and 80 per cent of localisation is unexplained by natural advantages. A different attempt to explain the residual excess location comes from assessments of the importance of localisation versus urbanisation economies. In this, Henderson (1988) found that localisation increases productivity and Henderson, Kuncoro and Turner (1995) found that localisation also increases growth. This contrasts with the results of Glaeser *et al.* (1992) that rather favour urbanisation economies, finding that diversity raises growth.

Finally, there is a line of research into the determinants of urban population growth, as well as the relationship between population growth and neighbouring regions (Hanson, 2000:26). This falls outside the scope of this study.

In conclusion it may be said that the empirical literature that tests the theories of the location of production in space is quite wide-ranging. But from the different tests of the

determinants of the spatial variation of per capita income, wages, industry employment or production, a number of commonalities are clear:

- First-nature geography is a significant explanatory variable, working either through endowments or distance.
- Geographical economics models put externalities forward as explanations of agglomeration. Externalities are significant second-nature geography explanatory variables, but the empirical tests are typically for the market access or home-market effect.

The remainder of this chapter will focus on this study's approach to testing the determinants of spatial economic growth, specifically in South Africa. The following section outlines the estimation strategy followed to obtain the results that are set out in Chapter 5.

3.3 Estimating the determinants of spatial economic growth

Of the different empirical approaches to estimation, this study follows those of Section 3.2.1, focusing on the determinants of spatial economic growth, specifically in South Africa. This methodology is closely related to growth theory and empirics. Section 2.2.3 provided an introduction to economic growth theory, discussing how neo-classical and endogenous growth theory accounts for the location of production in space. This section outlines the empirical considerations of estimating growth regressions[6].

It is important to note that much of the empirical literature in this regard focuses on explaining the determinants of spatial economic growth from a country-level perspective, but the methodologies apply equally to regions or localities. In fact, as will be explained, sub-national studies of the determinants of growth are particularly suited for examining the role of geography (first- and second nature) in determining growth. It is also, however, prudent to take heed of Neary's (2001:554) cautionary comment that sub-national studies should not take regions or localities as units of analysis without specifically considering issues of location.

The following sub-sections explain (a) how growth theory's different concepts of convergence influence the determinants of growth to be tested, (b) the estimation strategies and (c) the estimators that are used.

[6] This section draws on Naudé and Krugell (2003c)

3.3.1 Concepts of convergence and determinants of growth

The literature on the determinants of economic growth among countries, regions or localities derives from the neoclassical growth model in the tradition of Ramsey (1928), Solow (1956), Cass (1965) and Koopmans (1965). The models describe the determinants of growth and generally predict convergence of per capita income levels. The mechanism driving convergence is diminishing returns to reproducible capital: a relatively less well-off economy will have lower stocks of capital, hence higher marginal rates of return on capital and thus faster growth (Romer, 1986:1003). Within this framework the determinants of growth are linked to particular notions of convergence.

There is so-called β-convergence when poorer countries, regions, or localities tend to catch up with richer ones, or when the relative position of each country, region, or locality within the income distribution tends to stabilise over time (Barro & Sala-i-Martin, 1992:383).

It is absolute β-convergence when economies converge to the same long-run ("steady state") income per capita. In other words, initially poor countries, regions, or localities tend to grow faster until they catch up with the richer ones and in the long run, expected per capita income is the same for all members of the group, independent of its initial value. Here the determinants of spatial economic growth are the different rates of return on capital. Locations with lower stocks of capital will have higher marginal rates of return and will thus grow faster.

There is conditional β-convergence when the long-run incomes per capita differ. In the Solow model, the long-run ("steady state") income levels of countries, regions, or localities depend on underlying differences in fundamentals – determinants of growth such as investment rates and the rate of population growth. This means that relative backwardness still carries with it the potential for rapid growth, and relative income levels will eventually stabilise. The convergence is only to each locality's own steady state income per capita. Thus, the convergence is conditional on controlling for other determinants of the growth rate in addition to initial income. A range of different determinants of spatial economic growth can be estimated, such as investment in physical assets and human capital, as well as technology.

When β-convergence does take place, however, it only indicates that the poorer countries, regions, or localities are converging on average and says little about whether they have actually caught up or are falling behind others in terms of levels of per capita income. To determine this, one should rather consider the dispersion of income per capita over time across countries, regions, and localities. A reduction in this dispersion is called σ-convergence. Here the focus is on the evolution of the cross-sectional income distribution – its shape and the movement of the distribution over time (Barro & Sala-i-Martin, 1992:383). All other things being equal, β-convergence may eventually lead to σ-convergence.

In the empirical literature, there are many studies testing for convergence and quantifying the importance of various sources of growth across countries, regions, and localities. The empirical work is characterised by the use of convergence equations. The basic estimation strategies can be outlined as follows.

3.3.2 Determinants of growth and estimation strategies

The first step to testing the determinants of spatial economic growth is to estimate the unconditional convergence equation, assuming that the evolution of income per capita in a given country, region, or locality can be described by

$$\Delta y_{it} = x_i - \beta y_{it} + \varepsilon_{it} \qquad (1)$$

Where Δy_{it} is the growth rate of per capita income in country, region, or locality i; y_{it} is income per capita in country, region, or locality i in the initial period t; x_i summarises the underlying determinants of growth in country/region, or locality i and is here assumed to be constant over time; and ε_{it} represents the error terms that vary across countries and time periods and have mean equal to zero.

Note that here the key assumption is that the intercept term is common to all countries, regions, or localities. The estimated value of β in this regression may be interpreted as a measure of the average speed of convergence in the sample. A positive value of β (i.e. a negative correlation between growth and initial income) means that poor countries, regions, or localities are growing, on average, faster than rich ones and this will eventually lead to σ-convergence.

Though the positive value of β is a necessary condition for σ-convergence (and trivially so in the case of unconditional β-convergence), it is not, however, sufficient. When

testing for convergence it is important to keep in mind that if the deviation from the norm is significant, income dispersion could increase even if the 'typical' country, region, or locality is converging towards the sample mean. Hence, one has first to acknowledge that the intensity of convergence (the reduction of income dispersion) depends on the average pattern described by β, as well as on the size of the deviations from it, summarised by the R^2 of the regression or the significance of its slope coefficient. Secondly, one has to test the hypothesis of absolute convergence i.e. that long-run income levels are the same for all the economies. Thus the model must incorporate the determinants of the steady state, or at least dummy variables that allow one to control for possible differences across countries, regions, or localities. Given a negative value of the slope coefficient (a positive β), the question is whether the conditioning variables are significant, i.e. whether steady states differ across countries. If the answer is no, it is not possible to reject the hypothesis of absolute convergence. If the answer is yes, one must conclude that convergence is only conditional (De la Fuente, 1997:37).

This second step, estimating the conditional convergence equation, means bringing in additional variables that may help to account for cross-country/-region/-locality differences in growth patterns. The convergence equation is similar to (1) and can be described by

$$\Delta y_{it} = x_{it} - \beta y_{it} + \varepsilon_{it} \qquad (2)$$

where x_{it} now includes conditioning variables that capture differences in steady states, varying across countries, regions, and localities and time periods.

An extensive list of candidates of explanatory variables is available. Neoclassical growth theory suggests the rates of investment and population growth (Grossman & Helpman, 1994:26). The recent growth literature extends the concept of investment to include the accumulation of human and technological capital, as well as physical assets. Macroeconomists have, in turn, identified mechanisms through which fiscal and monetary variables may affect growth, and a number of authors have pointed out that additional factors, such as political stability, a well-developed financial system, or a high degree of openness to international trade, may have a favourable impact on investment and growth (Hall & Jones, 1997:174). The current debate on the importance of institutions or geography (or both), for economic growth, has seen the use of indicators of bureaucratic red tape, corruption, climate and landlockedness, amongst others, as conditioning variables of growth (see Rodrik *et al.*, 2002 and Sachs, 2003).

As stated at the start of the section, much of the empirical literature in this regard focuses on explaining the determinants of spatial economic growth from a country-level perspective, but the methodologies apply equally to regions or localities. Sub-national studies of the determinants of growth are particularly suited for examining the role of geography (first- and second nature) in determining growth – greater homogeneity within country borders allows one to implicitly control for institutions, the legal framework and macroeconomic policies. Section 3.2.1 showed that studies which use first-nature geography as an explanatory variable relate climate and endowments to variations in per capita income across countries, regions, and localities. Tests of endogenous spatial inhomogeneities specifically test for the home-market effect. The specific choice of explanatory variables in the case of South Africa will be discussed in Chapter 5.

Along with taking account of the explanatory variables, estimating for conditional β convergence can also entail different estimation strategies[7].

One method is to use a general specification that expresses growth as a function of the determinants of the ultimate steady states and the initial level of income. This is what Temple (1999:123) called informal or "Barro" regressions, where the explanatory variables are included on an *ad hoc* basis, drawing on economic theory and previous results from the literature. The alternative method is to derive a convergence equation as a solution of a log-linear approximation to an explicit growth model. This makes it possible to give a structural interpretation to the coefficients of a convergence regression. The procedure generally follows the exposition of Mankiw, Romer and Weil (1992) and tends to explicitly incorporate human capital or technology.

Finally, β-convergence is a necessary condition for sustained σ-convergence. The way to test for σ-convergence is to plot the time path of some index of the dispersion of per capita income – the coefficient of variation or the standard deviation of the log of income per capita are typically used as measures. Dalgaard and Vastrup (2001: 286) have, however, shown that these two measures may produce different conclusions because of the weight assigned to the growth observations of the countries, regions, or localities in the sample. It is recommended that both measures be used.

[7] Temple's (1999) concerns about the problems associated with the empirical work are discussed in section 3.4.2.

3.3.3 Determinants of growth and estimators

Once the explanatory variables and estimating strategy is decided upon, the question of which estimators to use comes to the fore. The formulation for estimating the determinants of growth rates (y) reduces to the following:

$$\ln y_{i(t2)} - \ln y_{i(t1)} = -(1 - e^{-\beta\tau}) \ln y_{i(t1)} + \mathrm{X}_{i(t1)}\gamma + \varepsilon_i \qquad (3)$$

Where y = per capita income of a particular place; $X_{i(t1)}$ = a vector of determinants of local economic growth rates. Following Trivedi (2002:6) the more general specification of equation (3) makes it possible to control for variables that might influence the steady-state levels of income, but which are not included explicitly in Solow (1956).

The results from using OLS to estimate a single-period-averaged cross-locality regression are fragile for well known reasons: biased and inconsistent estimates, since it may not take into consideration the endogeneity of some of the regressors and it may suffer from omitted variable bias (Gyimah-Brempong & Traynor, 1999:56 and Tsangarides, 2001:358). To overcome these shortcomings, panel data techniques such as the fixed-effects and random-effects estimators may be used. To illustrate these estimators consider that from equation (3), the so-called "between" estimator[8], is OLS applied to the following equation:

$$\overline{y_i} = \alpha + \overline{x_i}\beta + c_i + u_i . \qquad (4)$$

Where $\overline{y_i} = T^{-1}\sum_{t=1}^{T} y_{it}$ and so on. It should be noted that the "between" estimator is not consistent because $E(\overline{x_i}c_i) \neq 0$.

The fixed effects (or "within") estimator[9] is obtained by using OLS to estimate:

$$(y_{it} - \overline{y_i}) = (x_{it} - \overline{x_i})\beta + (u_{it} - u_i) \qquad (5)$$

The random effects estimator is a weighted average of the estimates produced by the between estimator (4) and the within estimator (5):

$$(y_{it} - \theta\overline{y_i}) = (1-\theta)\alpha + (x_{it} - \theta x_{it})\beta + \{(1-\theta)c_i + (u_{it} - \theta\overline{u_i}) . \qquad (6)$$

---

[8] The between estimator only uses the variation between the cross section observations.

[9] The within estimator uses the time variation within each cross section of observations.

where

$$\hat{\theta}_i = 1 - \sqrt{\frac{\hat{\sigma}_u^2}{T_i \hat{\sigma}_c^2 + \hat{\sigma}_u^2}}.$$

Despite the strengths of fixed and random effects estimators based on panel data, there remain two further shortcomings that need to be dealt with.

Fixed-effects models condition out the unobservable individual effects and leave unexplained exactly the long-run cross-locality growth variation that is of interest. Furthermore, the distance variable and market-size variables, that are the focus of the growth-through-agglomeration arguments at sub-national level, would be eliminated in a fixed-effects specification. They are either time-invariant, or fairly constant over time, or affect growth only with a long lag and would fall away along with the initial values of capital stock and resource endowment. In addition to this, Nerlove (2000) raised a further argument against the use of fixed effects in a panel context. He argued that the inclusion of fixed effects in a dynamic model biases the coefficient of the initial value of the dependent variable, included as an explanatory variable, downward, towards zero and, therefore, towards support for the convergence hypothesis (p. 248). Nerlove suggested an alternative approach of using more appropriate random-effects models, such as two-stage feasible generalised least squares or maximum likelihood. In the analysis in Chapter 5 a GLS random-effects estimator is firstly used. However, a weakness of an RE-estimator is that of the potential endogeneity of the X_{it} as well as the loss of dynamic information (see Temple, 1999:126 and Nkurunziza & Bates, 2003:10). The incorporation of dynamics into the model allows equation (3) to be rewritten as an AR(1) model as follows:

$$y_{it} - y_{it-1} = \gamma_t + \alpha y_{it-1} + x_{it}\beta + c_i + u_{it} \qquad (7)$$

Where Δy_{it} = the log difference in per capita GDP over a period; y_{it} = the log of per capita GDP at the start of that period; x_{it} = a vector of explanatory variables, γ_t = period-specific intercept terms to capture productivity changes common to all countries; c_i = the unobserved country-specific and time-invariant effects (unobserved fixed effects); and u_{it} = the time variant error term.

Equivalently, equation (7) can be written as:

$$y_{it} = \gamma_t + (\alpha + 1)y_{it-1} + x_{it}\beta + c_i + u_{it} \qquad (8)$$

Writing (8) in first-differences eliminates the time-invariant components, c_i. This solves the problem of omitted variable bias:

$$\Delta y_{it} = \gamma_t + (\alpha + 1)\Delta y_{it-1} + \Delta x_{it}\beta + \Delta u_{it} \qquad (9)$$

However, it creates another problem, namely endogeneity; since it is clear that y_{it-1} is endogenous to the error terms through u_{it-1}. It will therefore be inappropriate to estimate (11) by OLS. Nickell (1981) found that this bias will often result in wrongly finding a high rate of convergence. To overcome this problem of endogeneity, an instrumental variable needs to be used for Δy_{it-1}. Two approaches, namely Anderson and Hsiao's (1982) instrumental variable (IV) and Arellano and Bond's (1991) two GMM-estimators (first-step and second-step, respectively), have been used in this regard. Anderson and Hsaio (1982) proposed using Δy_{it-2} or y_{it-2} as instruments. Arellano and Bond (1991) showed that using the lagged level, y_{it-2}, as an instrument is superior and that the list of instruments can be extended[10] to include further y_{it-3}, y_{it-4}...y_{it-k}. Moreover, the Anderson-Hsaio IV approach can be seen as a special case of two GMM-estimators proposed by Arellano and Bond (1991) to combine the list of instruments efficiently. These two GMM estimators are preferable as they gain efficiency by using additional moment restrictions[11].

Chapter 5 presents the empirical results from estimates of the determinants of spatial economic growth in South Africa. The choice of explanatory variables, estimation strategy and estimators are discussed in detail.

3.4 Summary

The purpose of this chapter was to explain the different empirical methodologies involved in testing the determinants of spatial economic growth, focusing specifically on the methods used in this study. The overview of the empirical literature showed that tests of the theories of the location for production in space are wide-ranging. From the different tests of the determinants of the spatial variation of per capita income, wages, industry employment or production, a number of commonalities are clear. Firstly, first-nature geography is a significant explanatory variable, working either through endowments or distance. Secondly, externalities are put forward by geographical economics model as explanations of

[10] It is assumed that there is no second-order autocorrelation in the differenced idiosyncratic error term.

[11] The IV approach leads to consistent but not necessarily efficient estimates of the parameters because it does not make use of all the available moment conditions (see Baltagi, 1995:126).

agglomeration and they are significant second-nature geography explanatory variables, but the empirical tests are typically for the market access or home-market effect.

The rest of the chapter focused on this study's approach to the estimation of the determinants of spatial economic growth in South Africa. The approach follows growth theory and empirics, using concepts of convergence to identify the determinants of growth at sub-national level. This sub-national focus is particularly suited to examining the role of geography (first- and second nature) in determining growth – greater homogeneity within country borders allows one to implicitly control for institutions, the legal framework and macroeconomic policies. Section 3.2.1 showed that studies use first-nature geography as an explanatory variable to relate climate and endowments to variations in per capita income across countries, regions, or localities. Tests of endogenous spatial inhomogeneities specifically test for a market-size effect. In addition, the benefits of this study's approach to the empirical methodology do not only lie in the sub-national focus, or inclusion of geography, but also in the modern estimators that are used in the regression analysis. Section 3.3.3 explained that a dynamic model should be estimated, using Arellano and Bond's (1991) two GMM estimators.

Given that Chapters 2 and 3 have set the scene for the empirical evidence from South Africa, the next chapter will discuss the South African space economy, looking at growth patterns and the role of cities in the economy. Thereafter, in Chapter 5, the results from empirical studies into spatial economic growth in South Africa are presented.

Chapter 4: The South African spatial economy

4.1 Introduction

During the past ten years there has been a significant increase in concerns about spatial economic development in South Africa (see Bloch, 1999:1). This has largely been motivated by the fact that past social and political policies had worsened spatial inequalities in South Africa. Through the notorious homeland policy of "grand" apartheid and the "Group Areas Act" the natural growth and development of South Africa's cities and towns was artificially curtailed at great economic cost. Reducing spatial inequality in an economy that has embraced globalisation[12] will not be easy, especially since *"South Africa has no navigable rivers, there are long distances inland between raw material sources, manufacturing facilities, and harbours, and there are long distances between the harbours and airports and the places of consumption overseas."* (Dehlen, 1993:x).

Most of South Africa's cities and towns produce very little. This is most apparent in the rural-urban divide and the concentration of more than 80 per cent of the economy's manufacturing in six urban metropolitan regions (Naudé, Krugell & Serumaga-Zake, 2002). Overall, only 20 per cent of places (towns and cities) produce 82 per cent of South Africa's GDP. The richest 20 per cent of places had an average per capita income in 2000 of R25277 compared to an average per capita income of R5452 of the poorest 20 per cent of places.

Whilst most places in South Africa produce very little, the transformation of South Africa's system of local government (see Naudé, 2001a; and Naudé and Jansen van Rensburg, 2002) has resulted in local authorities, which are constitutionally responsible for local economic development within their areas.

The primary objective of this study is to provide empirical tests of the extent to which economic growth across South Africa is determined by geography. So far Chapter 2 provided the theoretical overview of the determinants of spatial economic growth and

---

[12] The South African government's "GEAR" macroeconomic strategy adopted in 1996 is characterised by outward-oriented trade policy and an apparent abandonment of industrial policy (which was, in the past, synonymous with the regional policy of apartheid that largely failed). It is argued by Naudé *et al.* (2002) that this has particular implications for spatial development in South Africa; moreover, it may reinforce the current patterns of agglomeration.

Chapter 3 outlined the empirical methodologies. This chapter introduces the South African space economy. This shows an economy characterised by spatial inequality. The aim of the historical background presented in Section 4.2, is to take account of the complexity of the location of economic activity in South Africa in terms of history and institutions. Section 4.3 provides an overview of the South African literature to show the different approaches that researchers have taken to study the South African space economy. Currently, very little work has been done in South Africa on the question of the determinants of sub-national growth and the importance of geography. Only a few papers follow the geographical economics approach. Section 4.4 highlights South African cities, specifically the importance of the six largest cities in the space economy. Finally, Section 4.5 examines sub-national growth patterns over the period 1998 to 2002. From this it is not obvious that the smaller, poorer places in South Africa have been growing faster than the larger, better-off places, as the theory would predict. This sets the scene for the empirical analysis of the determinants of sub-national growth rates in Chapter 5. Section 4.6 concludes this chapter.

4.2 Historical origins of spatial inequality

History matters. When growth takes place through agglomeration it has a place-dependent character (Martin, 1999). Over time, different places develop different institutional forms and networks of inter-firm dependencies along with economic, cultural and political practices that influence the role that places play in phases of economic development. One implication of this view is that many economic decisions and structures are difficult and expensive to reverse. The other is that economic shocks or interventions may have more than a once-off effect, but might impact on long-run capacity through cumulative changes that can bring about convergence or divergence between regions and localities.

The following sub-sections briefly outline the history of South African structures, policies and shocks. The first looks at the forces that shaped the space economy up to democratisation in 1994. The second characterises the developments since then, examining broad trends specifically in manufacturing and spatial policies as well as the local government transition process.

4.2.1 The history up to democratisation

An explanation of spatial economic inequality in South Africa should be an explanation of the self-reinforcing development of cities and towns[13].

South Africa has six "large"[14] cities, namely Johannesburg, the East Rand (Ekurhuleni metropolitan government) Durban, Cape Town, Pretoria (Tshwane metropolitan government) and Port Elizabeth (Nelson Mandela metropolitan government). Their importance and the consequent spatial economic inequality can be explained by looking at trade, extraction, climate and culture, along with Apartheid's social engineering and the recent transformation of local governments.

Cape Town and Durban were first developed in the 17th and 18th centuries as trading posts on the shipping route between Western Europe and Asia. During the 19th century this role changed with the discovery of diamonds and gold in the interior. The port cities developed from being stop-over and service points providing shipping services, to being ports through which these precious commodities were handled. Today this dominance continues due to the importance of sea transport for South Africa's international trade. Approximately 98 per cent of the volumes of South Africa's exports are conveyed by sea. The mineral wealth determined the location and growth of the other two dominant cities, Johannesburg and Pretoria. The distances of the location of mining commodities, as well as the extraction technology required in mining then influenced the pattern of South Africa's inland development. Where railways and electric power were provided for mining, they also contributed to the development of the manufacturing sector. Industries such as steel and mining, which are heavy consumers of electric power are predominantly located in the historic mining areas whilst chemicals are concentrated heavily around Durban from where the majority of the country's crude oil imports are obtained. Suleman (1998) shows that today, the basic metals and fabricated metal products are concentrated in the Gauteng Province (around Johannesburg and Pretoria areas), paper and chemicals in the KwaZulu-Natal province (around Durban) and food processing and textiles in the Western Cape Province (around Cape Town).

---

[13] See Naudé and Krugell (2003b) for the complete discussion.

[14] The World Bank (2000:128) distinguishes between small cities (less than 0.5 million people), large cities (between 1 million and 5 million people) and mega-cities (over 5 million people).

The unequal development determined by trade and extraction was reinforced by climate and culture. Recently Acemoglu *et al.* (2001:1370) argued that "*Colonies where Europeans faced higher mortality rates are today substantially poorer than colonies that were healthy for Europeans....settler colonies had representative institutions which promoted what the settlers wanted and what they wanted was freedom and the ability to get rich by engaging in trade*". Following this argument, the location of South Africa's southern ports and the Mediterranean climate around Cape Town translated into better subsequent economic performance. The persistence of British institutions (Westminster-style democracy, property rights and education systems) in South Africa had a further favourable impact on settler mortality. These effects persist and provide explanation for the sizes and functions of the major cities and the associated spatial inequality.

In the 20th century Apartheid reinforced the historical regional development patterns induced by the emerging mineral-energy complex of the 19th century with its homeland policies and Group Areas Act. Support of inefficient industries in the homelands and the segregation of cities created a spatial economy characterised by inefficient land use, excessive transport costs, and under-investment in transport infrastructure, telecommunications and electric power. It also resulted in segmented labour and consumption markets and created artificial internal barriers to trade.

The cost of unequal development was paid particularly by the manufacturing sector. Nel (2002:83) showed that by 1970, South African had a relatively advanced and diversified manufacturing sector, but thereafter output stagnated and employment declined. Contributing factors included: declining gold exports and gold prices, a reduction in global commodity demand from the early 1980s, the debt crises of the 1980s, depreciation of the value of the Rand, the imposition of sanctions, foreign exchange shortages, skill and capital shortages. By the 1990s job losses occurred in places and de-industrialisation took place.

Against this background of spatial inequality, the new democratic government has, since 1994, been introducing new spatial policies and engaging in local government transformation.

4.2.2 Spatial trends and developments since 1994

In 1994 the new democratic government inherited an economy challenged with low growth, high unemployment and severe inequality. Much has been written about the growth of the

economy in aggregate, the fiscal and exchange rate challenges, but less so about the spatial character of the economy.

Spatially, the historical dominance of the six largest cities remains largely unaltered, though there are a number of factors impacting on the location of economic activity. Rogerson (1991:364) for example, states that "*the new industrial geographies of post-Apartheid South Africa appear set to be dominated by a re-focussing of manufacturing activity around the large metropolitan centres, (and) the demise of the industrial base of several favoured decentralised growth points*". But despite the strength of the traditional core areas it should also be acknowledged that all is not well within the cores. Nel (2002:86) cited the closure of textile firms in Cape Town and the Witwatersrand area of Gauteng that is experiencing a contraction of its manufacturing economy. Nel also identified a number of factors that currently influence the location of economic activity (p.87-88):

- Trade liberalisation has benefitted industries and locations that have expanded to new markets, such as the motor industry in the Eastern Cape, but it has harmed sectors and locations that have struggled to cope with international competition, such as textile manufacturers in the Western Cape.
- Related to liberalisation, there are also sectors that have benefitted from foreign direct investment, including information technology and telecommunications, energy and oil, food and vehicles.
- There has been a gradual shift towards serviced-based and knowledge-intensive activities.
- There has been a decline in employment opportunities in manufacturing due to fluctuating commodity prices, skills shortages and the mixed fortunes of gold and the global economy in general.
- Also, there is increasing informalisation of the economy and the growth of the small business sector.
- Finally, new growth areas such as Midrand, the development corridors and special Industrial Development Zones all impact on the location of economic activity.

The spatial aspects of government policies since 1994 that are mentioned last are of particular significance to future developments. These can be discussed in two parts: the spatial characteristics of industrial policy and the local government transition process.

4.2.2.1 The spatial characteristics of industrial policy

Early on, the new government started out with an 'aspatial' industrial policy. This was in contrast to the apartheid government's regional development strategies, which were aimed at supporting apartheid-created homelands. So, for example, in 1997 a tax holiday scheme was introduced to encourage industrial development throughout the country. In recent years, however, thinking has swung back to spatially focused considerations in the form of Spatial Development Initiatives (SDI) and Industrial Development Zones (IDZ) (Nel, 2002:90).

The SDI programme was conceived and launched during 1995-96 as an important component of restructuring the post-apartheid economy. With the SDIs, government seeks to encourage investment, manufacturing and other economic activities along a series of defined transport corridors. Here, bottlenecks to investment, such as inadequate infrastructure, are to be removed and strategic opportunities for private sector investment identified. Lemon and Rogerson (2002:8-9) identified several components of the SDI strategy:

- Firstly, it involves crowding-in of private investment through co-ordination of public sector investment in areas with proven potential for economic development, which is currently unexploited.
- Secondly, to ensure fast and focused planning, government provides political support, commitment, and buy-in for the SDI process.
- Thirdly, government clears obstacles to investment and contributes towards the building of vital infrastructure.
- Detailed information about the SDIs and packages of potential investment opportunities are publicised.
- Targeted industries are clustered around anchor projects to ensure the strongest local linkages and greatest multiplier effects.

With such a process, it may still be early for an evaluation of the SDI strategy, but Nel (2002:90) argued that, despite the considerable efforts that have been put into the promotion of the SDIs, it does not yet appear that the initiatives have brought about dramatic economic transformation in their areas – with the exception of the Maputo Development Corridor.

Similarly, the establishment of the IDZs has been slow. The IDZs are industrial zones with a defined export focus and are all planned to be sited at major ports and airports. Only in 2001 did government announce IDZ incentives to the value of R3 billion. Of this, R600 million is for wage incentives to boost job creation. There is also a 50-100 per cent investment allowance for investors who undertake approved projects (Nel, 2002:91).

Along with the spatial focus of industrial policy, the South African space economy has recently also been shaped by the local government transition process.

4.2.2.2 The local government transition process

After 1994 the local government transition process was given a very strong legal driving force through three pieces of legislation. The most important was probably the Local Government Transition Act (LGTA) of 1993 that provided for transitional local governments as well as for a clearly defined transition process. Other pieces of important legislation were the Development Facilitation Act which attempted to address the "mindset" of local governments and to steer them in the direction of being more participative, and the Demarcation Act that allowed for the radical reconsideration of the geographical areas of jurisdiction of local governments and to ensure that every area of South Africa falls under democratically elected local government – so-called "wall-to-wall" local government.

The Demarcation Act (1997) has lead to a complete change in the borders and size of local governments in South Africa. Before December 2000, South Africa had 843 municipalities, many of which were financially unsustainable. According to surveys in 2000, about 100 municipalities in South Africa were unable to pay or service their debts to banks, while another 100 were in some form of financial distress. In terms of the Demarcation Act, an independent, expert committee, the Demarcation Board, was mandated to demarcate South Africa's local government's areas of jurisdiction. In deciding on the boundaries for new municipalities after 5 December 2000, the Demarcation Board considered the following factors:

- The interdependence of people, communities and economics such as employment, public transport, human settlement, migration patterns and access to services and recreational facilities
- The relationship to districts, voting areas, health, police, population, existing or expected land use, type of land in the area and environmental implications.

The result was that the number of municipalities was reduced to 284[15]. The demarcation process, together with the organizational restructuring effected by the Municipal Structures Act, and the way in which municipalities function in terms of the Municipal Systems Act (2000) has revolutionised local government in South Africa.

The current generation of municipal councillors, mayors and administrations are the first to "test drive" this system. If human and financial resource constraints can be managed adequately local government can become the primary institutional vehicle for economic development in South Africa, particularly in alleviating absolute poverty and ensuring the provision of basic services to all.

The tensions that local authorities will have to manage are often contradictory in nature (Naudé, 2001b). Firstly, municipalities will have to overcome historical legacies including having inherited administrations that are often ill equipped, inexperienced and unmotivated. Secondly the system makes them responsible for basic service delivery and development. Thirdly, national fiscal policy is reducing financial resources available to fulfil these obligations. Finally, globalisation is pressuring municipalities to create internationally competitive locations for businesses and households, going beyond merely supplying basic services and infrastructure.

In conclusion, this brief historical overview shows the complexity of the location of economic activity in South Africa in terms of history and institutions. As such, different approaches have been used to study the space economy. The following section provides an overview of the South African literature in order to put in context this study's geographical economics approach to the determinants of spatial economic growth.

4.3 Overview of the South African literature

The academic literature on the South African space economy is made up of divergent contributions from urban and regional planners, economic geographers and economists. Categorising the literature according to the explanations of the location of production in space set out in Section 2.2, would make for limited reading. Very few researchers have looked strictly at the spatial distribution of economic activity or the determinants of the

[15] Spatially, 237 municipalities constitute so-called wall-to-wall local government in South Africa. These are 231 local municipalities of various sizes but with less than a million inhabitants, and the six metropolitan municipalities, all with more than a million people. The 47 District Municipality areas coincide with that of the local municipalities so that they can perhaps be seen as a higher level of local government.

growth of economic activities across different localities – and even fewer explicitly take account of geography. Rather, to show the range of work, and to indicate where this study's geographical economics approach to the determinants of spatial economic growth fits in, the South African literature can be discussed under the a number of categories.

Firstly, there are the studies that examine different topics at sub-national level. The topics include issues of agriculture, manufacturing, tourism, infrastructure, employment, poverty and inequality at a provincial or local level and the work has a strong development focus. Opposed to this broad cross-section of topics that are addressed at a sub-national level, a second category of work includes the range of demographics studies that have been undertaken in South Africa. Here labour force survey and census data have been used to examine employment, unemployment, poverty and inequality, but typically with a very limited spatial focus, or none at all. A third category includes the work that looks at rural issues and questions of the urban-rural divide. Fourthly, it is possible to group together studies that focus specifically on cities, typically from a planning perspective. The part of the literature that examines local governments from the perspective of fiscal decentralisation and local economic development initiatives is related to the cities category. Finally, the sixth category includes the studies that focus specifically on the Spatial Development Initiatives. This section concludes with the limitations of the current literature and the contribution of the work on which this study is based.

4.3.1 Topics studied at sub-national level

The biggest and most wide-ranging of the categories is the first – that of studies that examine topics at sub-national level. Here the spatial level of observation is provincial or local, but the studies do not explicitly set out to explain the location of production in space or necessarily emphasise geography. The topics examined in the literature include issues of agriculture, manufacturing, tourism, infrastructure, employment, poverty and inequality, but at a provincial or local level.

For example, McDonald and Punt (2002) examined the spatial character of agricultural policy by using a SAM-Leontief analysis of the impact of policy liberalisation on agriculture in the Western Cape. They found that liberalisation should have benefitted rural communities, specifically farm workers, but the benefits are limited by supply constraints. In a different study, analysis of agricultural growth multipliers found similar constraints. Hendriks and Lyne (2003) conducted expenditure analyses from data collected from two rural areas of KwaZulu-

Natal and found weak growth linkages. They concluded that agriculture-led growth in South Africa requires public investment in physical and institutional infrastructure to reduce transactions costs and risks in all markets.

When examining issues of manufacturing at sub-national level Suleman and Naudé (2003) used export specialisation and output specialisation ratios to identify potentially competitive manufacturing clusters at a spatial level in South Africa.

In the area of tourism, a number of authors have made contributions to the sub-national literature. Rogerson (2002a) wrote of changes in the spatial organisation of tourism in South Africa. The existing tourism space economy is very uneven with investment concentrated on opportunities in the Western Cape and Gauteng. He identified two factors that may contribute to a greater spread of the benefits of the growing tourism economy: investment in essential infrastructure that "debottlenecks" untapped tourism potential, and maximisation of the impacts of local-level initiatives for tourism-led economic development. In a case study of the Highlands Meander, Rogerson (2002b) cited the example of five towns that are collaborating in their LED initiatives in order to promote the area's tourism products. He found, however, that this tourism initiative is currently not benefitting local black communities. Another example is the work of Mahoney and Van Zyl (2002) who analysed the extent to which the Makuleke tourism initiative, Umngazi River Bungalows and the commercialisation of the Manyeleti Game Reserve have improved the livelihoods of rural communities. Finally, there is also Naudé, Saayman and Saayman (2000) who examined the spatial implications of the impact of tourist spending by using an input-output model to distinguish between the relative impacts of domestic and international tourist spending. The authors concluded that there should be an emphasis on the development of the domestic tourism market.

South African literature that examines issues of infrastructure at sub-national level includes a range of issues. Le Roux Booysen (2003a) examined provincial disparities in progress on reconstruction and development. He found that provinces that are more urbanised have advantages in the delivery of infrastructure, the facilitation of demographic transition and the improvement of standards in secondary education, over the more rural provinces. At a local level Smith and Hanson (2003) examined the related development issue of access to water for the urban poor, specifically in Cape Town. They found that commercialisation and the current "basic needs" approach are creating territorial variation in

service delivery and widespread water cut-offs. In a final example of work linking "delivery" and spatial issues, a study of housing delivery in the Free State by Marais and Krige (2000) showed that cities in the province have been neglected in terms of housing investment. They do, however, find evidence that the majority of beneficiaries of housing subsidies fall in the lowest income category.

The final part of the literature that examines topics at sub-national level looks at employment, poverty and inequality. This category has close links with the following category of demographics studies and typically draws on household and labour force surveys, but here there is a specific provincial focus. For example, Leibbrandt and Woolard (1999) estimated relative poverty shares. The concept of a critical poverty range, as well as distribution-sensitive decomposable poverty measures are used to examine provincial poverty. In a study of the Western Cape, Oosthuizen and Nieuwoudt (2003) found that the poor are most often located in urban areas, they have low levels of education and live in relatively large households that are often headed by women. There are other approaches to poverty in the provinces: Pushkar and Ray (2003) looked at household resources and household composition in KwaZulu-Natal, Tsheola (2002) examined basic needs in the Northern Province and in a number of papers Naudé and Serumaga-Zake (2001), Serumaga-Zake and Naudé (2001, 2002) examined the determinants of rural and urban household poverty, the determinants of labour force participation and the determinants of labour market discrimination in the North West Province.

4.3.2 Studies of demographics

Opposed to the broad cross-section of topics that are addressed at a sub-national level, a second category of work includes the range of demographics studies that have been undertaken in South Africa. Here labour force survey and census data have been used to examine employment, unemployment, poverty and inequality, but typically with a very limited spatial focus or none at all.

A primary issue examined in this literature is that of migration. Posel (2003) examined the collection of national household survey data in South Africa and concluded that questions of labour migration have received too little attention in recent revisions of the surveys. In related work examining migration patterns in post-apartheid South Africa, Posel (2004) found that temporary internal labour migration appears to have increased, particularly

because of the rise in female migration. Cornwell and Inder (2004) focused on rural-urban migrants and found that they do well at finding formal employment.

The bulk of the studies of demographics are household studies that focus on inequality and poverty within and between income and racial groups and with little cognisance of spatial aspects. For example, Leibbrandt, Woolard and Bhorat (2000) examined household inequality and found that overall income inequality in South Africa is largely with-in group inequality and wage income contributes to total inequality. In related work Leibrandt, Bhorat and Woolard (2001) explored the relationship between wage income and household inequality and show that many of the unemployed depend on wage earners within their households, but a significant percentage of the unemployed, especially in rural areas, have no direct link to labour market earners. There is a large body of work associated with these approaches. McDonald, Piesse and Van Zyl (2000) examined the distribution of household income and measured the effect of transfers into the household. Then, Le Roux Booysen (2002) used demographic and health surveys to measure differences in the socio-economic status of households in the absence of income and expenditure data. Also, Khosa (2002) used a Human Sciences Research Council survey to examine how recent economic growth influenced the financial positions of households. And Puhskar (2002) examined the effect of household characteristics such as the gender of the household's head, his/her educational attainment and ethnicity, and the way these affect poverty and living standards.

In conclusion it should again be emphasised that though the data from the surveys used in the above studies of demographics involve sub-national units of observation, the focus of the work has typically not been on spatial issues beyond simple urban-rural divides. An exception to this is Dinkelman and Perouz's (2002) study of the determinants of labour force attachment where they found that an individual's location is a significant determinant of whether or not they choose to search for work.

4.3.3 Rural questions and the urban-rural divide

The body of literature that examines rural questions and issues of the urban-rural divide in South Africa is somewhat removed from the approach to the space economy followed in this study and consequently only a few recent papers, that are typical of the literature, are mentioned.

For example, there is a range of articles discussing rural economies from the perspective of land reform (see Bernstein, 2003). Robinson (2003) examined rural settlement patterns in the Eastern Cape and found that the land tenure system and risk-spreading strategies of households are keeping people on the land, although their existence is not subsistence-based but linked to the urban economy. Le Roux Booysen (2003b) found that urban-rural inequalities in access to health care services persist and discriminate against the poor.

4.3.4 Cities and urban management and planning

The body of work on rural questions has an urban counterpart in the literature – there are studies of urban vulnerability (Nomdo & Coetzee, 2002), social justice (Visser, 2001), urban empowerment (Lotter, 2002), quality of life (Moller, 2001) and poverty (Rogerson, 2001a). Here, too, the focus is on issues of households, poverty and inequality, but specifically within the urban context of cities and towns.

The related part of this literature addresses mostly planning and management issues. For example Buthelezi and Dollery (2004) provided an exploratory analysis of local government failure, and Cameron and Sewell (2003) looked at performance management in the Cape Town municipality. This work also fits in with the efforts of the South African Cities Network. The SACN is a network of South African cities and partners that encourages the exchange of information, experience and best practices on urban development and city management (see www.sacities.net).

These approaches are, however, somewhat removed from the approach to the space economy followed in this study and further examples may be omitted.

4.3.5 Fiscal decentralisation and local economic development issues

The literature on fiscal decentralisation and local economic development initiatives lies closer to the questions studied in this thesis – specifically from a policy perspective.

The fiscal decentralisation literature provides the public economics perspective on the devolution of decision making and the challenges facing provinces and local government in South Africa. Prime examples of work in this field include Grote (2000) who provided an overview of fiscal decentralisation in South Africa, Rubenfeld (2001) who examined the system of intergovernmental grants and Rodden *et al.* (2003) who studied the incentives for

fiscal discipline in the decentralised system. In recent papers Niksic (2004) re-examined the decentralisation strategy specifically from the perspective of local governments, and Ntsebeza (2004) looked at the dilemmas of traditional authorities and land administration. Also, related literature studies the financial aspects of decentralisation from the point of view of sub-national taxation, specifically local property taxes. Here a volume of articles entitled "*Property taxes in South Africa: Challenges in the post-apartheid era*" (Bell & Bowman, 2002), provides a useful overview of the literature.

Together with the above public economics approach to fiscal and financial questions of provinces and local governments, there is further literature on their economic development challenges. The focus is specifically on Local Economic Development (LED) initiatives and the Integrated Development Planning (IDP) processes currently being undertaken in South Africa. Here Nel has contributed a number of key papers. He provided a review and assessment of the current status of LED initiatives (Nel, 2001), reviewed the policy and legal developments (Nel & Binns, 2001) and provided three case studies of the LED and IDP processes (Binns & Nel, 2002).

4.3.6 Spatial development initiatives

Section 4.2.2.1 has already introduced the SDI programmes and along with the work by Lemon and Rogerson (2002) and Nel (2002) mentioned in that section, there are a number of articles that examine issues in this field. For example, Hartzenberg (2001) reviewed South Africa's regional industrial policies and the development of the SDI programmes and questioned the role of incentives in influencing firm-level decisions, specifically their location decisions. Harrison and Todes (2001) looked at the use of spatial frameworks in regional development in South Africa examining the problems confronting attempts at spatial planning. There are also a number of articles looking at specific SDI programmes:

- Kepe (2001) looked at the management challenges that confronted the implementation of the Wild Coast SDI project.
- Ehlers and Antonites (2000) evaluated the opportunities and obstacles for investment in the Maputo Development Corridor.
- Rogerson (2001b) also evaluated the Maputo Development Corridor.
- Soderbaum and Taylor (2003) are the editors of a collection of papers on the Maputo Development corridor. The papers discuss the objectives, strategies and design

principles of the Maputo Development Corridor (De Beer & Arkwright), the Maputo Development Corridor and the broader regional context (Tevera & Chimhowu); the role of the state in the Maputo Development Corridor (Soderbaum & Taylor), Governance in the Maputo Development Corridor (Soderbaum); Public-private partnerships (Ngwenya & Taylor); the Moçambique aluminium smelter (Hentz); and the view from Maputo (Baptista-Lundin & Taylor).

- Kleynhans, Naudé and Van der Merwe (2003) provided an overview and evaluation of the Platinum Spatial Development Initiative, putting forward the history of regional industrial policy and the current provincial policy context[16].

Finally, Du Pisanie (2002) examined the SADC's regional development corridors and suggests inter-corridor competition to attain efficiency, but intra-corridor co-operation, for example in transport.

In conclusion there is a vast body of South African literature that examines economic activity and space. But, as this brief review shows, very few researchers have looked strictly at the spatial distribution of economic activity or the determinants of the growth of economic activities across different localities – and even fewer explicitly take account of geography. In fact, Naudé (2005), states that "*By and large, economists have not yet provided sufficient rigorous economic analyses of the Southern African spatial economy and the relationship between geography and economic development*". He writes of the recently published book *Geography and Economy in South Africa and its Neighbours* (Lemon & Rogerson, 2002) and states that there is limited analysis of economic geography and most of the recent work does not take account of the new economic geography approach of Krugman and no attempt is made to approach spatial questions in South Africa from an explicit underlying framework (Naude, 2005). As such, this study's geographical economics approach to the determinants of spatial economic growth builds on recent theoretical advances and contributes to filling a void in the South African literature.

The following sections now move away from the historical perspectives and overviews to a description of the South African space economy in accordance with the data that is used in the empirical estimation. Section 4.4 highlights South African cities,

[16] Related to work on the North West Province, Dippenaar (2001) examines the IDZs as spatial policy instruments, focusing specifically on the Mafkeng IDZ.

specifically the importance of the six largest cities in the space economy. Section 4.5 examines sub-national growth patterns over the period 1998 to 2002. Section 4.6 concludes.

4.4 South African cities

The level of urbanisation in South Africa is currently around 58 per cent according to the United Nations (2001). Moreover, the rate of urbanisation in South Africa is high. Between 1996 and 2001 the population of the largest 21 urban centres rose from 18.4 million to 21.1 million – that is growth of 14.23 per cent over the period, or on average 2.7 per cent per annum (SACN, 2004:37). It has been predicted to be around 2.09 per cent per annum over the period 2001 to 2005. Some of the fastest growth occurs in so-called second-tier cities. Migration accounts for the city growth trends with the key drivers being household splitting and permanent migration from rural villages to urban areas – particularly the migration of secondary school children (South African Cities Network, 2004:9).

Cities play an important role in economic growth through providing economies of scale through the advantages of agglomeration[17]. South African cities have been growing faster than the national average growth rate and they have less unemployment. They are however also characterised by skills shortages and significant inequality. Given the huge backlogs in basic service provision, agglomeration and its resulting higher population density might provide economies of scale in the provision of basic goods and public goods. However, in the light of the possibility of diseconomies of scale as a result of congestion, the extent to which further agglomeration will be good for economic growth may depend on the current levels of agglomeration in South Africa. In other words, the economic growth impact of future city growth may depend on the current distribution and sizes of South Africa's cities. The following sub-section examines absolute and relative city size and Section 4.2.2 thereafter, the growth and dispersion of South African cities.

4.4.1 Absolute and relative city size in South Africa

As said in section 4.2.1, South Africa has six "large" cities, namely Johannesburg, The East Rand Metropole, Durban, Cape Town, Pretoria (Tshwane metropole) and Port Elizabeth (Nelson Mandela metropole). Three of these (Durban, Cape Town and Port Elizabeth) are

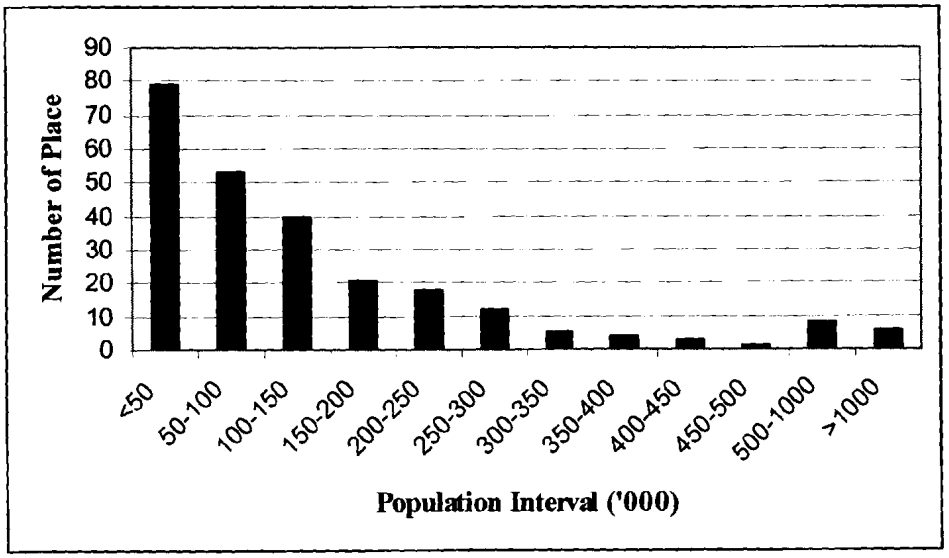
[17] This section draws on Naudé and Krugell (2003b)

port cities through which much of South Africa's international trade moves. Durban is also close to the port of Richards Bay, and Cape Town close to the port of Saldanha. These two ports are important for South Africa's exports of coal and steel, respectively. Johannesburg and the East Rand Metropole, although for administrative (legal) purposes two entities, form one large urban agglomeration. According to Beaverstock, Smith and Taylor (1999:453) Johannesburg can be seen as the primate city of South Africa, and one of the leading African cities. They identified Johannesburg as the only African city that serves as a major global accountancy and legal services centre. Whilst Johannesburg is classified as a "gamma" world city (comparable to Boston, Amsterdam, Melbourne), they found only minimal evidence that Cape Town is developing into a world city (see Beaverstock, Smith & Taylor, 1999:457).

The dominance of three port cities amongst South Africa's largest cities is due to the importance of sea transport for South Africa's international trade. Approximately 98 per cent of the volume of South Africa's exports is conveyed by sea.[18]

Figure 4.1 shows the distribution of city and town sizes in South Africa.

Figure 4.1: Distribution of Municipal Population Sizes in South Africa, 2001



[18] In this, South Africa is very similar to most developing countries. Sachs and Warner (1997:339) note that "... only certain goods can be economically shipped by air, and most countries still import and export the majority of goods by sea".

Figure 4.1 shows that the distribution of the sizes of South African municipalities. By tabulating the cumulative distribution of municipal sizes it can be seen that about 80 per cent of South African municipalities have less than 250 000 inhabitants. Indeed the median[19] municipality in South Africa in 2001 had a population of 87 000 people.

The 40 largest cities in South Africa all have populations of 250 000 or more, but apart from Greater Johannesburg (including the "East Rand area"), less than 5 million. Thus South Africa is a country with a peculiar spatial distribution of economic activity and population: the cities tend overall to be "small", with six "large" cities and no "mega" city. The cities and their populations in 2001 are depicted in Table 4.1.

Table 4.1: Contribution of South Africa's Six Metropolitan Areas to Total GDP, 1990, 1996 and 2000 (%)

CITY	Population 2001	% Contribution to SA GDP in 2000	% Contribution to SA GDP in 1996	% Contribution to SA in 1990	Coastal City with Port	Average Annual Population growth Rate, 1996-2001
Johannesburg	2962759	14.98%	14.16%	14.22%	No	1.02%
Cape Town	2858743	14.01%	12.82%	11.90%	Yes	1.01%
Durban	2981237	7.77%	8.06%	8.18%	Yes	0.96%
Pretoria	1454290	8.55%	7.88%	9.06%	No	1.03%
East Rand	2054133	7.78%	8.00%	8.67%	No	1.01%
Port Elizabeth	1015334	2.46%	2.31%	2.58%	Yes	1.02%
Total %	13326496	55.6%	53.2%	54.5%		1.5%

(Source of data: STATSSA, 2001; PIMSS, 2001)

Table 4.1 shows that South Africa's six major cities, with 31 per cent of South Africa's total population, contribute, in total, 55 per cent of South Africa's GDP. Since 1990 the share has increased slightly from 54 per cent to 55 per cent. These six cities also contain 6 million of South Africa's approximately 17 million voters. The ranking of Johannesburg and Cape Town as the two largest cities in economic terms, has remained unchanged since 1990. Indeed, both these cities have enlarged their share of total GDP, in particular Cape Town. In

[19] The average municipal size in 2001 was 178365 and is twice the size of the median due to the 6 "outlier" cities with more than 1 million people each.

contrast, the share of all other cities declined since 1990, most notably that of Durban (the largest single city) and the East Rand.

The spatial distribution of South Africa's six largest cities is implicit in Table 4.1, as suggested by the nature of the city as a port city or not. Figure 4.2 depicts the geographical location of the six cities.

Figure 4.2: Geographical Location of South Africa's Six Largest Cities

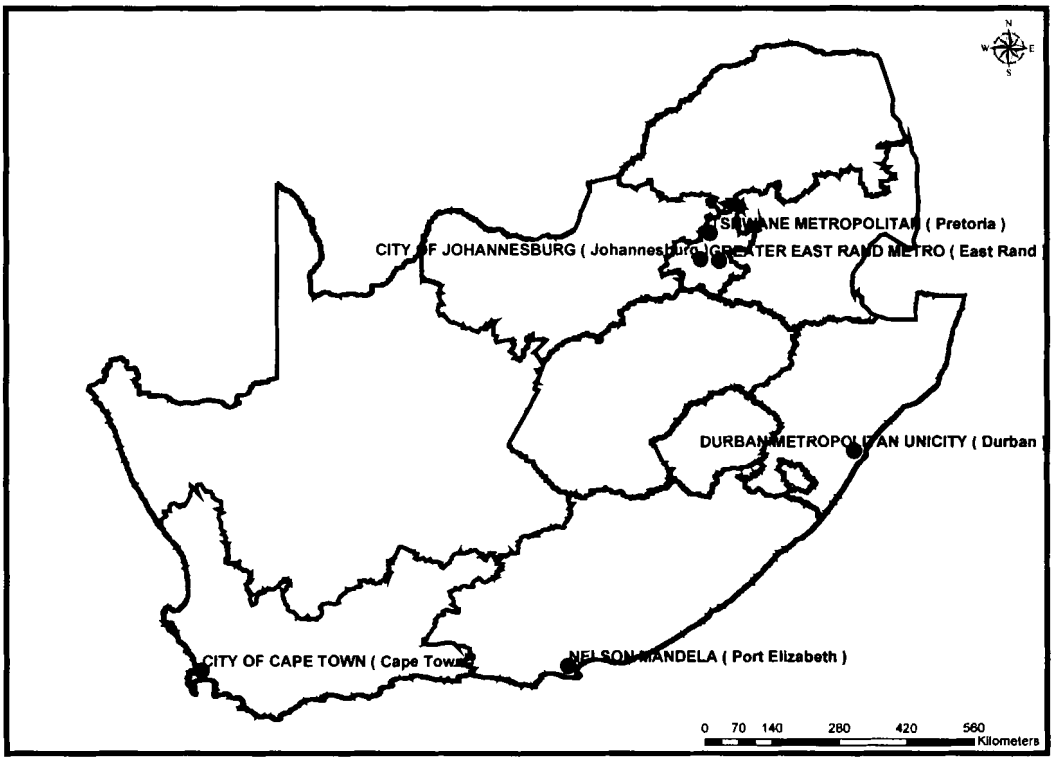


Figure 4.2 shows that three of South Africa's largest cities are port cities. In particular, Durban and Cape Town's harbours are the largest harbours in Southern Africa. Whilst the other three large cities in South Africa are all inland cities without a harbour or port, Johannesburg, East Rand and Pretoria are located in close proximity within Gauteng Province, exhibiting the nature of one "cluster" city rather than three geographically separated cities. All three are relatively close to the largest international airport in Africa, and all are connected via the N3 Highway (South Africa's busiest) to the harbour ports of Durban, Richards Bay, and Maputo (Moçambique), and via the N1 and N12 Highways to Cape Town and Walvis Bay (Namibia).

Table 4.1 shows that in 2000, the city with the largest single contribution to South Africa's GDP was Johannesburg, with 14.98 per cent (although in terms of population Johannesburg has been overtaken by Durban). If one includes East Rand as part of the City of Johannesburg, then the dominance of Johannesburg as a "primate" city in South Africa becomes clear: it then produces 22.8 per cent of South Africa's GDP. In this respect then, South Africa does have a primate city, unlike other former British Colonies such as Australia and Canada.

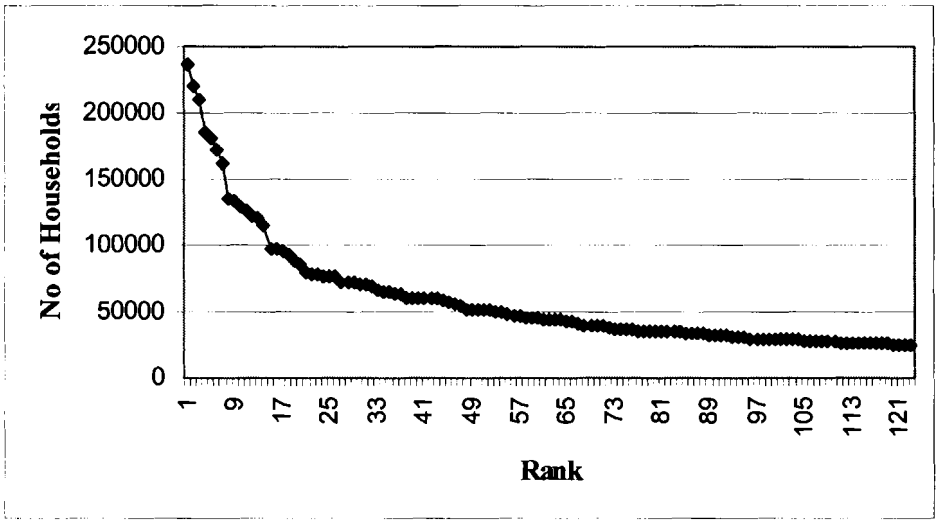
The Primacy Ratio gives an indication of the dominance of the primate city (the city ranked number 1 in terms of population) in the urban system of a country (see Rosen & Resnick, 1980). The so-called Primacy Ratio[20] for South Africa, calculated on the basis of the data in Table 4.1, is 24 per cent if Johannesburg and the East Rand are treated as two separate urban agglomerations. However, with Johannesburg and East Rand as one urban agglomeration, the Primacy Ratio increases to 38 per cent. According to Henderson, Shalizi and Venables (2000) (cited in Brakman, Garretsen & Van Marrewijk, 2001) the "optimal" degree of primacy that maximises economic growth for middle-income countries such as South Africa is 25 per cent. This could suggest that South Africa's primate city (Johannesburg-East Rand combined) is larger than optimal.

4.4.2 City growth and dispersion in South Africa

The Rank-Size distribution of cities throughout the world follows a law that states that the number of cities with a population larger than S is approximately proportional to S-q (Gabaix, 1999; Knudsen, 2001). In other words, rank times population size is approximately the same constant for all cities. If q is equal to or close to unity, it is also known as "Zipf's Law" (Zipf, 1949). Thus Zipf's Law is a special case of the rank-size distribution (Knudsen, 2001:125). In Figure 4.3 the rank-size distribution of South Africa's 123 largest places is shown.

[20] Following Rosen and Resnick (1980) the Primacy Ratio is calculated as the size of the largest city as a percentage of the sum of the sizes of the five largest cities.

Figure 4.3: Rank-Size Distribution for South Africa (123 Largest Places)



A log-linear regression of the 123 largest cities and towns in South Africa (a population of 100 000 was taken as cut-off) (S) and its rank (N) based finds that

$$Ln(S) = 6.6 - 0.75Ln(N)$$

(s.e. = 0.009)

With $R^2 = 0.98$

Thus, in the South African case the coefficient q = 0.75 . Fujita *et al.* (1999:216) reported a coefficient of 1.004 for 130 large USA cities. Thus, whilst the rank-size rule applies to South Africa (its explains 98 per cent of the variance in city sizes) Zipf's Law does not hold for South Africa. What are the implications of q = 0.75 for evaluating the sizes of South Africa's cities?

According to Brakman *et al.* (2001) if $q < 1$, then a more even distribution of city sizes results than if Zipf's Law holds, and in the limiting case where q = 0 all cities would be of the same size. Therefore it may be concluded that South Africa's cities tend to be more evenly spread. Gabaix (1999) showed that Zipf's Law would hold if cities were characterised by either constant returns to scale or by external economies of scale where positive and negative externalities cancel out. This could be interpreted as the point where the agglomeration forces and congestion (dispersion) forces cancel out (see also Brakman *et al.*, 2001). Brakman, Garretsen, van Marrewijk and van den Berg (1999) supported this notion

and calculated q for the Netherlands for the periods 1600, 1900 and 1990, and noted that "*industrialisation lead to an increase in q*". By 1600 the value for q was much lower than 1 (0.55) and it subsequently rose as city sizes increased.

Bearing in mind the challenge that "*nobody has come up with a plausible story about the process that generates the rank-size rule*" (Fujita *et al.*, 1999:225) the implications of the Gabaix (1999), Brakman *et al.* (1999) and Knudsen (2001) explanations would suggest that South Africa's cities (urban agglomerations) are too small and that these cities are still predominantly offering urbanisation economies rather than localisation economies (see also Henderson *et al.*, 1995).

The rejection of Zipf's Law suggests that South Africa's cities are of relatively even size (as inspection of Table 4.1 confirms for the six largest cities) but not yet in a steady state distribution (Knudsen, 2001:132). It may therefore be instructive to consider the geographic distribution of the large cities (e.g. those with more than 100 000 people, of which South Africa has 123). The so-called "inverse H-measure" has often been used in the literature as a rough measure of the degree of concentration in the city spread. Wheaton and Shushido (1981:22) reported the inverse "H-measure" for concentration for 38 countries, including South Africa. This "H-measure" can take a value from 1 (extreme concentration) to n, where n = the number of agglomerations in the sample. For South Africa they took n =19, i.e. 19 metropolitan areas (containing 70 per cent of the urban population). The inverse H = 11.3.

Due to the closeness of the measure to n=19 this suggests a reasonable degree of dispersal (in terms of number of large locations) in South Africa. From theory (see e.g. the model of Krugman, 1991) this could be the case for a geographically large country such as South Africa only if transport costs were high and the need for economies of scale were low (e.g. when production tends to be labour rather than capital intensive). Given the nature of the dominant sector in the South African economy such as mining and metals-based manufacturing that depend on economies of scale due to large fixed costs, and the large distance between points of extraction or manufacturing and the markets, the spread of urban agglomerations could depend on the relative strengths of high transport costs (favouring dispersion) and the need for economies of scale (favouring agglomeration in one

place)[21]. Also, political and social policies constrained urban development in the past and would have forced the dispersion of agglomerations. Lower transport costs and the removal of political barriers to urbanisation over the past decade, could suggest that the current extent of dispersal of agglomerations may be inefficient.

4.5 Economic growth patterns, 1998-2002

The characteristics of the 25 fastest growing magisterial districts in South Africa over the period 1998 to 2002 are summarised in Table 4.2. The table shows the average annual real per capita growth in Gross Value Added (GVA), as well as the population growth rate, level of urbanisation, literacy rate and share of exports in the local economy.

Table 4.2: The fastest growing cities and towns in South Africa, 1998-2002

Place	Growth in GVA pc	Population	Population Growth	Level of urbanisation	Literacy rate	Export share
Warmbad	13.14%	55254.9	1.40%	0.62	67%	6.83%
Lower Umfolozi	9.78%	271925.5	1.48%	0.28	65%	115.89%
Randburg	8.02%	417401.2	1.70%	0.94	87%	37.51%
Waterberg	7.46%	63473.7	1.43%	0.46	60%	0.51%
Potgietersrus	6.76%	55213.0	1.41%	0.49	66%	0.57%
Laingsburg	5.98%	6570.2	0.79%	0.60	57%	0
Mossel Bay	5.21%	64442.6	1.32%	0.88	79%	19.78%
Tulbagh	5.07%	32436.9	1.03%	0.57	65%	4.66%
Pretoria	4.74%	783566.8	1.10%	0.93	91%	22.84%
Thabazimbi	4.50%	63867.2	1.64%	0.46	64%	0.09%
Wellington	4.45%	49019.1	1.09%	0.81	78%	32.27%
Rustenburg	4.38%	417167.8	1.68%	0.44	68%	94.26%
Belfast	4.31%	26488.1	1.26%	0.66	57%	1.14%
Knysna	4.21%	66807.3	1.74%	0.87	78%	6.06%
Waterval Boven	3.90%	10860.4	1.40%	0.72	69%	0
Witrivier	3.89%	22309.6	1.63%	0.43	71%	5.34%
Phalaborwa	3.85%	34171.0	1.25%	0.50	72%	33.59%
Chatsworth	3.80%	200570.1	1.08%	1.00	84%	20.14%
Humansdorp	3.61%	54074.9	1.06%	0.67	71%	19.34%
George	3.52%	125516.2	1.60%	0.88	76%	2.42%
Wonderboom	3.43%	369416.5	1.90%	0.87	85%	36.51%
Sterkstroom	3.28%	8445.6	0.98%	0.74	51%	0

[21] Puga (1998) shows that when transport costs are very low and increasing returns to scale significant, an urban system with a large primate city will develop.

Place	Growth in GVA pc	Population	Population Growth	Level of urbanisation	Literacy rate	Export share
Montagu	3.06%	25741.7	1.36%	0.74	67%	21.57%
Volksrust	2.99%	34400.9	1.71%	0.79	59%	0.45%
Schweizer-Renecke	2.98%	49566.5	0.89%	0.73	39%	3.06%
Average	5.05%	132348.3	1.36	0.68	69%	19.40%

Source: Regional Economic Focus, 2005.

Table 4.2 shows that the 25 fastest growing cities and towns in South Africa averaged an annual growth rate of 5.05 per cent per annum over the period 1998 to 2002. The high-growth cities and towns are a mix of large, urbanised centres with a literate population and large share of exports in the economy, as well as smaller towns. Randburg, Pretoria, Wonderboom and Chatsworth are examples of large cities that have grown fast over the period. There, the population is urbanised and educated and exports make up a substantial share of economic activity. There are also smaller places like Mossel Bay, Humansdorp, Wellington and Montagu that fall in this league. Rustenburg and Phalaborwa were also fast growers with significant export shares, but are less urbanised, with resource extraction making up a significant part of the economy. Smaller places like Waterberg, Potgietersrus, Laingsburg, Thabazimbi, Belfast, Knysna, Witrivier, Belfast, Sterkstroom, Volksrust and Schweizer-Renecke also managed high average growth rates.

In contrast to the fastest growing places in South Africa, Table 4.3 summarises the slowest growing places over the period 1998 to 2002. In all there were 201 magisterial districts that registered negative growth rates on average over the period.

Table 4.3: The slowest growing cities and towns in South Africa, 1998-2002

Place	Growth in GVA pc	Population	Population growth	Level of urbanisation	Literacy rate	Export share
Dannhauser	-13.32%	82457.7	0.97%	0.05	62%	0.26%
Theunissen	-7.78%	40747.9	0.38%	0.61	57%	0.16%
Odendaalsrus	-7.51%	104405.1	0.99%	0.91	68%	0.07%
Welkom	-7.16%	282937.3	1.16%	0.85	71%	1.43%
Bronkhorstspruit	-6.53%	40054.2	1.67%	0.61	64%	30.69%
Hlabisa	-6.34%	222338.1	1.63%	0.07	49%	0.42%
Mtunzini	-6.26%	218703.8	1.43%	0.24	66%	15.12%
Hennenman	-6.15%	29812.6	1.15%	0.84	66%	4.56%
Virginia	-5.97%	86478.4	0.43%	0.76	69%	1.62%
Mitchells Plain	-5.84%	843232.6	2.68%	1.00	79%	0.21%
Kuilsrivier	-5.63%	271300.8	1.71%	0.98	85%	21.91%
Klerksdorp	-5.30%	369766.8	0.98%	0.86	69%	0.81%
Vryheid	-5.24%	98120.5	1.15%	0.40	61%	0.15%
Mahlabathini	-5.14%	161450.5	1.48%	0.12	51%	0
Maclear	-5.03%	21026.1	0.88%	0.74	51%	0
Estcourt	-4.63%	176508	1.50%	0.17	63%	3.31%
Somerset West	-4.60%	72399.9	1.34%	0.89	87%	25.34%
Underberg	-4.55%	18663.2	1.51%	0.12	56%	0
Aliwal North	-4.51%	30809.3	1.03%	0.89	65%	0
Nkomazi	-4.20%	276045.9	2.09%	0.06	48%	0
Standerton	-4.11%	97385.4	1.33%	0.69	62%	0.30%
Glencoe	-4.03%	34348.7	1.12%	0.76	68%	2.57%
Stutterheim	-3.93%	35370.4	1.05%	0.41	55%	5.34%
Umvoti	-3.87%	97098.3	1.64%	0.13	49%	0.68%
Heilbron	-3.84%	41977.9	0.60%	0.70	56%	0.33%
Average	-5.66%	150137.6	1.28%	0.56	63%	4.61%

Source: Regional Economic Focus, 2005.

Table 4.3 shows that the 25 slowest growing cities and towns in South Africa contracted at an average rate of -5.66 per cent per annum over the period 1998 to 2002. The low-growth cities and towns again present a varied profile. Amongst the slow growers there are former centres of resource extraction (especially areas dependent on gold mining) that have contracted significantly, for example Welkom, Klerksdorp and Standerton. There are also localities close to the major metropolitan centres that have struggled, for example, Mitchells Plain and Kuilsrivier near Cape Town. The greater share of the towns and cities are however small, more rural than urban, with low levels of human capital and insignificant

export shares. In conclusion it is clear that, on average, the slowest growing cities and towns tend to have smaller populations, are less urbanised and have lower literacy rates. It also appears from Table 4.3 as if the slower growing places were less export-oriented than the faster growing places.

From the tables, however, it is not clear that the smaller, poorer places in South Africa have been growing faster than the larger, better-off places, as the theory would predict.

4.6 Summary

This chapter has shown a South African space economy characterised by significant spatial inequality that is the product of a complex history and different institutions.

The chapter also reviewed the different approaches that researchers have taken to study the South African space economy. The literature was grouped into broad categories of studies that examine development topics at sub-national level, demographics, rural questions, the rural-urban divide, cities and urban management and planning, fiscal decentralisation and local economic development issues, as well as spatial development initiatives. In conclusion it was said that there is a body of South African literature that examines economic activity and space. But as this brief review shows, very few researchers have looked strictly at the spatial distribution of economic activity or the determinants of the growth of economic activities across different localities – and even fewer explicitly take account of geography. As such, this study's geographical economics approach to the determinants of spatial economic growth builds on recent theoretical advances and contributes to filling a void in the South African literature.

Section 4.4 highlighted South African cities, specifically the importance of the six largest cities in the space economy. It was argued that the economic growth impact of future city growth may depend on the current distribution and sizes of South Africa's cities. The analysis showed the dominance of the six large cities with their importance having remained unchallenged. Though there is no clear mega city, Johannesburg and the East Rand combined has a primacy ratio of 38 per cent. The rank-size rule is however rejected in the South African case, suggesting that South Africa's cities are of relatively even size, but not yet at a steady state distribution. Finally, Section 4.5 examined sub-national growth patterns over the period 1998 to 2002 and found that it is not clear that the smaller, poorer

places have been growing faster than the larger, better-off places, as the theory would predict.

Chapter 5 sets out to test the growth theories explicitly by using a dynamic panel data regression model to examine the determinants of economic growth at sub-national level in South Africa.

Chapter 5: Convergence and divergence

5.1 Introduction

Chapter 4 showed that South Africa is characterised by significant inequality in the spatial distribution of economic activity. Most places produce little and sub-national growth patterns show that average economic growth rates ranged between 13.4 per cent and -13.3 per cent per annum during the period 1998 through 2002. This means that different places in the country have shared very unequally in the positive economic growth that accompanied the country's growing internationalisation since 1994. From the growth patterns examined in Section 4.5, it is not clear that the smaller, poorer places in South Africa have been growing faster than the larger, better-off places, as the theory would predict.

This chapter examines the determinants of economic growth at sub-national level in South Africa. The aim is to contribute to the understanding of why growth rates within a single country might differ amongst regions, and particularly the role that geography may play, and to explore the policy measures that may be effective in raising growth. This contributes to the growing empirical literature on sub-national growth and convergence.

Overviews of the convergence literature are found in De la Fuente (1997), Durlauf and Quah (1999) and Temple (1999). Barro (1991) and Barro and Sala-i-Martin (1992) are among the first authors who report evidence of convergence across countries. The early literature finds a stable rate of convergence at around 2 per cent per year. Temple (1999: 134), however, noted that recent estimates have varied between zero and 30 per cent a year, and that the consensus that is emerging is one of uncertainty. Studies of convergence at regional or local level also present mixed evidence. Recent work by Ferreira (2000) and Nagaraj, Varoudakis and Venganzones (2000) found evidence of regional (within country) convergence in Brazil and India, respectively. There is however also evidence of divergence, from Bandyopadhyay (2001) in the case of India and Cikurel (2002) in the case of Mexico.

As for the case of South Africa, Section 4.3 concluded that few researchers have looked strictly at the spatial distribution of economic activity or the determinants of the growth of economic activities across different localities – and even fewer explicitly take

account of geography. This study makes a clear contribution in a country where economic decision-making is increasingly decentralised.

The empirical evidence presented here makes use of several modern panel data estimators, such as GLS and GMM (see Section 3.3.3).

The chapter is structured as follows[22]. Section 5.2 describes the data used for the empirical work of this study. Section 5.3 presents the evidence on absolute convergence. Section 5.4 outlines the estimating equation and estimators used in the conditional convergence exercise and presents the GLS and GMM regression results. The results of tests for sigma convergence are presented in Section 5.5. Section 5.6 concludes.

5.2 The Regional Economic Focus data

Before 5 December 2000, local government in South Africa consisted of 354 magisterial districts. Following a process of demarcation, described in Section 4.2.2.2, 284 municipalities replaced the magisterial districts after that date. Currently, however, the only comprehensive database that is available corresponds to the magisterial district boundaries. It is a system of integrated databases known as the Regional Economic Focus (REF) that provides information for each magisterial district over this period. The REF is compiled by Global Insight Southern Africa and draws together many different sources of sub-national economic information from Statistics South Africa, government departments, development agencies and Regional Services Councils. The data components are internally consistent and add up to national totals. A number of indicators of the economies and people of the magisterial districts are used in this paper.

For economic activity, Gross Value Added (GVA) by locality is used. It is a proxy for income and is used for the variables income per capita and growth of income per capita. The GVA statistics differ from Gross Domestic Product (or Gross Geographic Product) in that it excludes "Other taxes on products" and "Other subsidies on products", which is not available on a regional basis. The compensation of employees on a regional level and GVA by sector is, however, available and therefore it is possible to derive estimates of GVA at magisterial district level.

[22] This chapter draws on Naudé and Krugell (2003c).

The growth of income per capita (or more accurately, GVA per capita) is to be explained by a number of conditioning variables that determine the steady state of the magisterial district economies. These variables include an education measure of human capital, the initial capital stock, the initial mineral endowment, the share of exports in local output as well as a number of variables put forward in the economic geography literature, such as distances, population density and rainfall.

The human capital measure is an education measure – the number of people per locality with matric and a Ph.D. The number of people in this education category in each magisterial district was first estimated using the 1996 census data. The trend in educational attainment between 1996 and 1999 was estimated from the 1999 October Household Survey and used to interpolate national level educational attainment for the years 1997 and 1998. From 2000 onwards, Labour Force Survey raw data was used to calculate the national trend in educational attainment. The final annual estimates for the magisterial districts were re-checked to ensure that they are consistent with the population numbers.

The value of the initial capital stock in 1990 comes from outside Global Insight's Regional Economic Focus and was obtained from a different consultancy, called Econometrics.

The measure of initial resources is an estimate of the mineral endowment of each magisterial district. It is calculated as the number of workers in the mining sector relative to the total workforce. The series are from the REF, compiled from numerous sources including the 1996 census, the Standardised Employment Series, October Household Surveys, Surveys of Total Employment and Earnings and the Labour Force Survey. The number of people employed in the mining sector was benchmarked on magisterial district level estimates of employment in the mining sector obtained from the Minerals Bureau.

Population density is used as a measure of market size and agglomeration. This is defined as the number of people per square kilometre.

The distance variables used in the analysis were obtained from outside the REF. The distance from each magisterial district to the nearest port and to Johannesburg was calculated using SA Explorer software provided by the Municipal Demarcation Board.

The rainfall data is sourced from the South African Weather Service. The data is obtained on weather station level. The average rainfall is calculated from the weather stations that are geographically within each magisterial district.

The following section explains the empirical work in which the data are used.

5.3 Absolute convergence

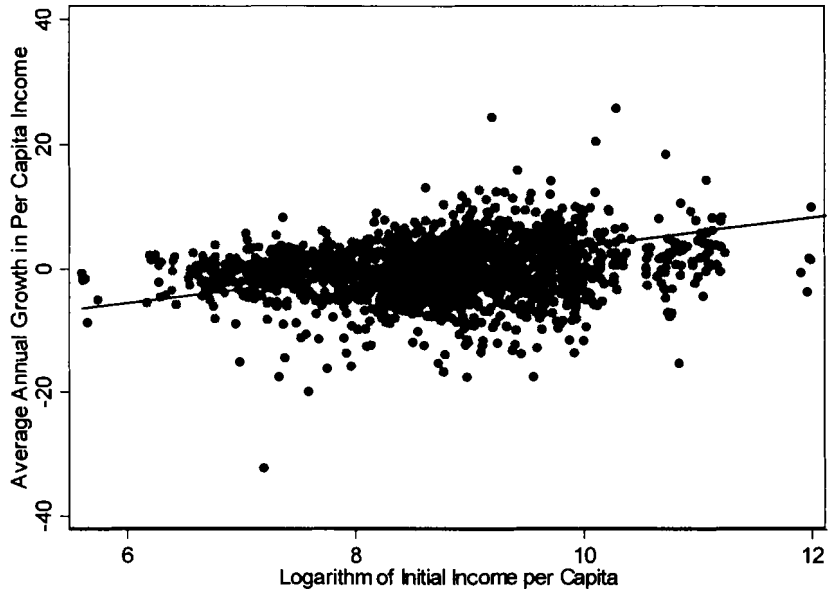
Section 3.3.1 explained that absolute convergence occurs when economies converge to the same long-run income per capita. The empirical work is characterised by the use of convergence equations. Thus, the first step is to estimate the unconditional convergence equation (1) to determine whether the poor magisterial districts in South Africa, on average, grew faster than the rich ones. Equation (1) is reproduced here as equation (10), assuming that the evolution of income per capita in a given magisterial district can be described by

$$\Delta y_{it} = x_i - \beta y_{it} + \varepsilon_{it} \qquad (10)$$

Where Δy_{it} is the growth rate of per capita income in magisterial district i; y_{it} is income per capita in magisterial district i in the initial period t; x_i summarises the underlying determinants of growth in locality i and is here assumed to be constant over time; and ε_{it} represents the error terms that vary across magisterial districts and time periods and have mean equal to zero. As explained in Section 3.3.2, the estimated value of β in this regression may be interpreted as a measure of the average speed of convergence in the sample. A positive value of β (i.e. a negative correlation between growth and initial income) means that poor magisterial districts are growing, on average, faster than rich ones and this will eventually lead to σ-convergence.

Figure 5.1 displays the results of an unconditional convergence equation and the corresponding scatter diagram.

Figure 5.1: Unconditional convergence among South African Magisterial Districts, 1998-2002



$(y02 - y98)/5 = -5.513 + 0.593y98$

(t =) (0.00) (0.00) $R^2 = 0.018$

The positive slope of the regression line is a clear sign of divergence. The R^2 of 0.018, however, indicates a poor fit and suggests that other factors have played a role as determinants of the growth rate.

The following section presents the conditional convergence estimates.

5.4 Conditional convergence and growth regressions

The approach taken here is based on the general specification described in Section 3.2.2. Growth in the various magisterial districts is expressed as a function of the determinants of the steady states and the initial levels of income.

Since the aim is to evaluate the evidence of β-convergence of magisterial districts in South Africa, rather than to explicitly test a particular growth model that predicts convergence, the *ad hoc* Barro-style specification is used.

5.4.1 Estimating equation

As shown in Section 3.3.3, the formulation for estimating the determinants of growth rates (y) reduces to the following. Equation (3) is reproduced here as equation (11):

$$\ln y_{i(t2)} - \ln y_{i(t1)} = -(1 - e^{-\beta\tau}) \ln y_{i(t1)} + \mathrm{X}_{i(t1)}\gamma + \varepsilon_i \qquad (11)$$

Where y = per capita income of a particular place and initial per capita income enters on the right hand side. The parameter β can then be interpreted as the rate of convergence to the steady state, with τ = t2 – t1 > 0. $X_{i(t1)} =$ a vector of determinants of local economic growth rates. Following Trivedi (2002:6) the more general specification of equation (11) makes it possible to control for variables that might influence the steady-state levels of income, but which are not included explicitly in Solow (1956). Also, magisterial district level data on investment – a key variable in the Solow model – is not available in the data set. Instead, proxies for physical and human capital are used to control for the different steady state levels in most of the regressions.

The choice of the conditioning variables draws specifically on the growth theories described in Section 2.2.3 and the empirical literature outlined in Sections 3.2.1 and 3.3. The vector of determinants of local economic growth rates includes measures of physical capital, human capital, and resource endowment, the openness of the local economy and agglomeration and geography.

The "new" or "endogenous" growth theory emphasises the importance of economies of scale, agglomeration effects and knowledge spillovers as determinants of economic growth and convergence (Button, 1998). It suggests that economic growth tends to be faster in countries (but also localities) that have a relatively large stock of capital, a highly educated population and an economic environment favourable to the accumulation of knowledge. The latter is especially relevant when economic activity is information-intensive, highly differentiated and dependent on reciprocal specialisation.

In this case, only the initial capital stock in 1990 is available for all 354 magisterial districts and is used accordingly in the regressions. A positive coefficient is expected. The importance of human capital for growth builds on the spillover models in the spirit of Lucas (1988) and its significance has been shown by international and intra-national studies. Here the analysis follows Fedderke (2001) who found that only the proportion of matriculation students sitting mathematics, and the proportion of degrees issued in natural, engineering

and mathematical sciences, made a significant contribution to productivity growth in South African manufacturing. However, since such indicators of high-quality screening by the education system is not available at magisterial district level, the proportion of Ph.Ds in a magisterial district is used as a measure of human capital and a positive coefficient is expected.

A locality's resource endowment may also be a determinant of its steady-state level of income and, in this case, mineral endowments are used. It is proxied by the number of workers in the mining sector relative to the total workforce. A negative coefficient is expected.

As for the openness of the local economy, the export share is measured by the value of exports as a proportion of GVA. A vast literature exists on the benefits of openness for growth and a positive sign is expected here also.

The inclusion of measures of agglomeration and geography draws on explanations of growth that emphasise market size, spillovers, and transport costs, as outlined in Chapter 2. Henderson (2000:2) pointed out that agglomeration is essential for a country to kick-start industrial development. He states (p.2) that, "*By spatially concentrating industrialisation, often in coastal cities, the economy conserves on economic infrastructure – physical infrastructure capital (transport and telecommunications) and managerial resources*". In the analysis here, measures of population density and distance are used to capture the effects of the forces of agglomeration and geography in growth. This method follows the growth regression-based work described by Overman *et al.* (2001) (see Section 3.2.1). Measures of first- and second-nature geography are used to explain the spatial variation of growth[23]. Population density is used as a proxy for market access and a positive coefficient is expected. The distance measures include the distance from the nearest port and the distance from Johannesburg. Greater distance from markets and harbours is considered to be detrimental to growth through agglomeration and a negative coefficient is expected. Finally, rainfall serves as a further measure of geography. A positive coefficient is expected.

The error term (ε) will here include unmeasured attributes affecting local growth rates such as global shocks, climate, etc.

[23] This is different from spatial econometric approaches that use a measure such as Moran's I to capture spatial dependence (see Lim, 2003) and falls outside the scope of this study.

The data that is used was obtained form the Regional Economic Focus, described in Section 5.2. Exceptions are the capital stock that was provided by Econometrics and the distances, which were calculated using SA Explorer software provided by the Municipal Demarcation Board (see www.demarcation.org.za).

Where possible, the natural logs of variables were taken. Regressions were estimated using Stata 8.0.

Before setting out the regression results, the estimation strategy and a number of caveats regarding the estimators, as explained in Section 3.3.3, should briefly be reiterated. The first point is that using OLS to estimate a single-period-averaged cross-locality regression would produce biased and inconsistent estimates, since it may not take into consideration the endogeneity of some of the regressors and it may suffer from omitted variable bias. To overcome these shortcomings, panel data techniques such as the fixed-effects and random-effects estimators may be used. The fixed-effects models, however, have shortcomings of their own – they condition out the unobservable individual effects and leave unexplained exactly the long-run cross-locality growth variation that is of interest. In this study, the distance variable and market-size variables would be eliminated in a fixed-effects specification. In addition to this, Nerlove (2000: 248) raised a further argument against the use of fixed effects in a panel context. He argued that the inclusion of fixed effects in a dynamic model biases the coefficient of the initial value of the dependent variable, included as an explanatory variable, downward, towards zero and, therefore, towards support for the convergence hypothesis. Consequently, in this study, a GLS random-effects estimator is firstly used. However, a weakness of a RE-estimator is that of the potential endogeneity of the X_{it} as well as the loss of dynamic information. Thus, a dynamic model is also estimated, using Arellano and Bond's (1991) two GMM-estimators.

The following sections report firstly the results from the GLS estimates, and secondly the results of the dynamic (Arellano-Bond) model.

5.4.2 Estimation results

Table 5.1 shows the GLS estimates.

Table 5.1: GLS Random Effects Regression Results

Variable	Coefficient	z
Intercept	1.639	0.22
Initial income per capita in 1997	0.595	1.16
Initial education	0.616	1.78*
Initial capital stock	-0.145	-1.61
Initial resources	-0.189	-1.04
Export share	0.233	1.79**
Density	-0.828	-2.78**
Distance from nearest port	-0.139	-0.35
Distance from JHB	-0.550	-1.88*
Rainfall	-0.103	-0.16
R^2 within: 0.0018	R2 between: 0.1871	R2 overall: 0.07
Number of observations: 485	Number of groups: 109	

***Significance at the 1%, **at the 5%,* and at the 10% level.

Table 5.1 shows that in this case only density is significant at the 5 per cent level, but with an unexpected negative coefficient. The capital stock is insignificant, and enters with a negative sign. This may be explained with reference to Glaeser *et al.* (1992) who argue that localities follow the fortunes of the industries that they were initially exposed to. This view is linked to a vintage capital model where areas that invest in older types of capital do not replace the capital as it becomes obsolete. This may be because (a) existing capital presents a sunk investment and (b) the pre-existing capital "crowds-out" newer capital (scarce resources such as land are used with the older capital). As the capital becomes more out of date, the marginal product of labour, wage rate and income fall, and thus a negative coefficient is possible. In Table 5.1 there is, however, no evidence of convergence and the overall R^2 shows that the regression equation explains only 7 per cent of the variation of per capita growth rates.

The table above also indicates that initial education, initial resources (or mining intensity in employment), export share, distance from the nearest port and distance from Johannesburg all have the expected sign. The initial education, export share and distance from Johannesburg are significant at the 10 per cent level. This suggests that human capital,

openness and export orientation as well as access to domestic markets are important determinants of local economic growth.

The lack of significance of the distance to the nearest port was further explored, due to the possible interaction between this term (which is a proxy for domestic transport costs) and the export share from a particular area. An OLS regression of export shares on distances, education, capital, density, rainfall and resources, was performed and it was found that the coefficient on the distance from a port was -0.36 and highly significant whilst the distance to Johannesburg was insignificant. It was also found that population density was significantly and positively correlated with export shares, and entered with a coefficient of 0.20. This suggests that more urbanised areas with higher population density are relatively more export oriented, which may help explain the significance of distance in the table above.

Table 5.2 presents the results of the dynamic (Arrelano-Bond) model. It corrects for White[24] in the procedure.

Table 5.2: One-step GMM Regression Results

Variable	**Coefficient**	**Heteroskedasticity-consistent z-value**
Intercept	9.114	2.29**
Initial income per capita in 1997	-0.772	-2.46***
Education (lagged)	7.485	4.09***
Growth per capita (lagged)	0.154	2.15**
Initial capital stock	0.244	2.87***
Initial resources	-0.159	-0.98
Export share (lagged)	0.761	2.08**
Density (lagged)	88.020	1.04
Distance from nearest port	0.287	1.03
Distance from JHB	-0.847	-2.77***
Rainfall	1.218	1.07
Arellano-Bond Test for second order autocorrelation	p-value =0.043	
Wald $\chi 2$ (10)	70.32	
Number of observations: 270	Number of groups: 101	

***Significance at the 1%, **at the 5%,* and at the 10% level.

[24] White (1980) proposed the heteroskedasticity-robust variance matrix estimator to adjust the standard errors of a regression in the presence of heteroskedasticity.

In the dynamic model, initial income per capita, initial capital stock, initial resources and the distance variables are time-invariant and they are included as exogenous variables. The model then uses one-period lags of education, the export share and density and the changes in rainfall. The lags are used to avoid endogeneity between growth, education, exports and population movements.

The dynamic model clearly provides better results than the static RE-model. All the variables, save the distance from the nearest port, have the expected sign. The coefficient of the initial income per capita shows evidence of conditional convergence, albeit at a very low rate. Education, the initial capital stock, export share, and distance from Johannesburg (access to markets / transport costs) are shown to be significant determinants of the growth rate. Although the distance to the nearest port is not significant as a determinant of local economic growth rates, the distance to port variable was found highly significant as a determinant of the share of exports from a particular area. It may therefore be expected that transport costs have a significant impact on local growth rates through their effect on exports, which is found to be a highly significant determinant of local growth. Because a dynamic model is used there are lagged export shares, to avoid influences from growth to exports. Similarly, lagged education (here measured as the proportion of the labour force with matriculation and higher qualifications) is highly significant. This result emphasises the importance of human capital in economic growth.

5.5 Sigma convergence

Tentative results on σ-convergence may be obtained from calculating the coefficient of variation of income per capita, or the standard deviation of the log of income per capita, across the magisterial districts over the period 1998 to 2002. These results are shown in Table 5.3.

Table 5.3: Sigma-convergence among South African Magisterial Districts

Year	Coefficient of variation	Standard deviation of log of income per capita
1998	0.04076	0.55262
1999	0.04151	0.55178
2000	0.04154	0.55267
2001	0.04187	0.55214
2002	0.04201	0.55257

From Table 5.3 it can be seen that there is little evidence of σ-convergence among magisterial districts. When using the coefficient of variation as a measure it indicates that the variation of income per capita among magisterial districts increased slightly over the period – in other words, divergence occurred. The standard deviation of the log of income per capita decreased somewhat between 1998 and 1999, but then increased again.

5.6 Summary

This chapter examined the determinants of economic growth at sub-national level in South Africa and tested for tentative convergence in income per capita between 354 magisterial districts over the period 1998 to 2002. A dynamic panel data model was used to test for the determinants of economic growth.

The results support other findings that geography has an independent and significant effect on economic growth. In particular, it was found that internal geography, through the location of various places at differing distances from internal agglomerations, has a highly significant negative impact on the local economic growth rate of a magisterial district. Distance to a harbour (measuring external geography, or access to international markets) was not found to be directly significant, but indirectly through its highly significant effect on the export propensity of a particular magisterial district. Urban density (indicating agglomeration forces) also affects growth significantly through its positive effects on exports. Being located in or near a city in South Africa therefore helps economic growth through an internal market effect (reflecting urbanisation economies) as well as through a greater ability to penetrate export markets.

It was also found that human capital, export propensity and the capital stock (proxying for investment) of a magisterial district are significant determinants of sub-national

economic growth. To the extent that municipal investment reflects the institutional capacity and governance quality at a magisterial district level, the significance of the coefficient on capital stock can be taken to suggest that institutional quality on the local level matters.

As far as the question of convergence is concerned, no evidence could be found of absolute convergence and only tentative evidence of (very) slow beta-convergence in South Africa. In short, the analysis cannot find unambiguous results in favour of either strong convergence or divergence, perhaps due to the relatively short timeframe (five years) used. For instance, there is evidence of slow β-convergence, but possible σ-divergence and absolute divergence among magisterial districts. When using the coefficient of variation (σ) as a measure, it indicates that the variation of income per capita among magisterial districts has increased slightly over the period. This might suggest that if one does not control for other determinants such as human capital, institutional capital and geography, the tendency would be for per capita incomes to diverge.

Next, Chapter 6 concludes this study and offers a number of recommendations for policy and future research.

Chapter 6: Conclusions and recommendations

This study began with the observation that economic activity tends to be unequally distributed across space. The study set out to answer the question of what determines the spatial growth and development of the South African economy. For this the hypothesis was that geography has a significant effect on the economic growth rate of South African cities and towns, independent of that of institutions and policies.

Chapter 2 reviewed the theories of the spatial structure of economic activity and the determinants of spatial economic growth. This showed that economic growth in space is about agglomeration. Agglomeration is the result of increasing returns that follow from scale and impart the efficiency benefits of mass production, specialised intermediate inputs, better matching between workers and jobs, and the availability of public goods and services. Thus, economic growth in space is determined by so-called second-nature geography: large markets and localised external benefits. This introduced the importance of proximity, and first-nature geography, upon which agglomeration also depends. That is, that firms and workers benefit more from the spillovers from external economies when they are closer to one another. Within this framework, specialisation drives trade and growth. Chapter 2 concluded with a basic testable hypothesis of the determinants of spatial economic growth: That one would expect a positive relationship between the growth of economic activity across space and economies of scale; and a negative relationship between the growth of economic activity across space and distance from agglomerations.

Chapter 3 examined the empirical methodologies used to econometrically test for the determinants of spatial economic growth. Approaches to testing the theories of the location of production in space were found to be wide-ranging, but it was concluded that a number of commonalities emerge from the empirical literature: Firstly, first-nature geography is a significant explanatory variable, working either through endowments or distance. Secondly, externalities are put forward by geographical economics model as explanations of agglomeration and they are significant second-nature geography explanatory variables, but the empirical tests are typically for the market access or home-market effect. Subsequently, Chapter 3 focused on this study's approach to the estimation of the determinants of spatial economic growth in South Africa. The approach follows growth theory, using concepts of convergence to identify the determinants of growth at sub-national level. This sub-national focus is particularly suited to examining the role of geography (first- and second nature) in

determining growth – greater homogeneity within country borders allows one to implicitly control for institutions, the legal framework, and macroeconomic policies. First-nature geography variables such as climate and endowments are then used to explain variations in per capita income across localities. Tests of endogenous spatial inhomogeneities specifically test for a market-size effect. In addition to the sub-national focus and inclusion of geography, the benefits of this study's approach to the empirical methodology also lie in the modern estimators that are used in the regression analysis.

Where Chapters 2 and 3 set the scene for empirical estimates of the determinants of the spatial growth and development of the South African economy, Chapter 4 introduced the South African space economy with an overview of the historical origins of its spatial inequality, the South African literature that has studied it and a contemporary view of the importance of cities in the spatial growth and development of the economy. Examination of sub-national growth patterns over the period 1998 to 2002 showed that it is, however, not clear that the smaller, poorer places have been growing faster than the larger, better-off places as the theory would predict. This lead on to the empirical investigation in Chapter 5 of the determinants of economic growth at sub-national level in South Africa.

Chapter 5 examined the determinants of economic growth at sub-national level in South Africa and tested for tentative convergence in income per capita between 354 magisterial districts over the period 1998 to 2002. The results from the dynamic panel data regression model show that the significant determinants of local economic growth are distance from internal markets, human capital, export propensity and the capital stock of municipalities. Distance from international harbours, as a measure of transport costs, and urban agglomeration (or density) affects growth indirectly through its significant effect on the ability of a region to export. Overall, the results indicate that geography is important, independent of its effects on institutions, for economic growth. However, bearing in mind the medium-term focus of the work, no evidence of absolute convergence could be found over the five-year period, rather the tentative evidence suggests slow beta convergence.

Thus, this study has made three important contributions:

- Firstly, it has provided an empirical test of some of the implications of geographical economics. The results show that first- and second-nature geography makes for significant drivers of economic growth at sub-national level. This was achieved using

advanced econometric techniques, producing evidence in a developing country context.

- Secondly, it added to the current debate on the relative roles of institutions versus geography in growth. This study at sub-national level has found that geography is important for growth, independent of its effects on institutions.
- Thirdly, it has contributed towards the policy debate in South Africa. The results inform policymakers of the determinants of sub-national growth and may thus aid the sustainability of decentralisation and the creation of a more equitable space economy. Specifically, for policy making at local government level, the findings support policies to lower domestic transport costs (e.g. through provision of public infrastructure), promote exports, attract and retain high-skilled labour and raise local municipal investment through improved local governance. Moreover, local economic development policies as suggested may be important since otherwise the results suggest that poorer magisterial districts may not catch up with richer ones and that the challenges of low growth and spatial inequality may remain

Scope for further research on geographical economics will depend greatly on the availability of reliable sub-national data. Where industry-level data are available, future work should attempt further testing of the geographical economics framework, testing for instance, for the home-market effect and spatial wage curves. Work could include more in-depth analysis of city size and economic specialisation in South Africa, analysis of the role of transport costs in the economy, as well as further work related to sub-national growth. In particular, work on sub-national growth may be extended through using spatial econometrics approaches and testing for convergence clubs.

List of references

ACEMOGLU, D. JOHNSON, S. & ROBINSON, J.A. 2001. The colonial origins of comparative development: An empirical investigation. *American economic review*, 91(5):1369-1401, December.

ALONSO, W. 1964. Location and land use. Cambridge Mass.: Harvard University Press.

ANDERSON, T.W. & HSIAO, C. 1982. Formulation and estimation of dynamic models using panel data. *Journal of econometrics,* 18(1): 47-82, January.

ASCHAUER, D.A. 1989. Is public expenditure productive? *Journal of monetary economics*, 23(2):177-200, March.

ARELLANO, M. & BOND, S. 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of economic studies*, 58(2):277-297, April.

BANDYOPADHYAY, S. 2001. Twin peaks: Convergence empirics of economic growth across Indian states. [Web:] http://www.wider.unu.edu/publications/dps/dp2001-142.pdf [date of access: 14 April 2005]

BALTAGI, B.H. 1995. Econometric analysis of panel data. New York: John Wiley.

BARRO, R. 1991. Economic growth in a cross-section of countries. *Quarterly journal of economics*, 106(2):407-443, May.

BARRO, R. & SALA-i-MARTIN, X. 1992. Convergence. *Journal of political economy*, 100(2):223-251, April.

BEAVERSTOCK. J.V., SMITH, R.G., & TAYLOR, P.J. 1999. A roster of world cities. *Cities*, 16(6):445-458.

BELL, M.E. & BOWMAN, J.H. 2002. Property taxes in South Africa: Challenges in the post-apartheid era. Cambridge Mass.: Lincoln Institute of Land Policy.

BERNSTEIN, H. 2003. Land reform in Southern Africa in a world-historical perspective. *Review of African political economy*, 30(96):203-226, June.

BINNS, T. & NEL, E. 2002. Devolving development: Integrated Development Planning and developmental local government in post-apartheid South Africa. *Regional studies*, 36(8):921-932, November.

BLACK, D. & HENDERSON, J.V. 1999. A theory of urban growth. *Journal of political economy*, 107(2):252-284, April.

BLOCH, R. 1999. Sub-national economic development in South Africa: Towards a strategy for the Department of Trade and Industry. Paper presented at the Annual TIPS Forum, 19-22 September, Glenburg Lodge, Muldersdrift.

BLOOM, D.E. & SACHS, J.D. 1998. Geography, demography and economic growth in Africa. Brookings papers on economic activity, 2.

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, February.

BRAKMAN, S., GARRETSEN, H. & VAN MARREWIJK, C. 2001. An introduction to geographical economics: Trade, location and growth. Cambridge: Cambridge University Press.

BUTHELEZI, A. & DOLLERY, B. 2004. An exploratory analysis of local government failure in South Africa. *Journal for studies in economics and econometrics*, 28(2):91-105, August.

BUTTON, K. 1998. Infrastructure investment, endogenous growth and economic convergence. *The annals of regional science*, 32(1):145-162, February.

CAMERON, R. & SEWELL, B. 2003. The implementation of performance management in the Cape Town municipality: 1996-2000. *South African Journal of economic and management sciences*, 6(2):243-261, June.

CARLTON, D.W. 1983. The location and employment choices of new firms. *Review of economics and statistics*, 65(3):440-449, August.

CASS, D. 1965. Optimum growth in an aggregative model of capital accumulation. *Review of economic studies*, 32:233-240, July.

CICCONE, A. & HALL, R. 1996. Productivity and the density of economic activity. *American economic review*, 86(1):198-210, March.

CHIKUREL, D.C. 2002. Why Mexico's regional income convergence broke down" Paper delivered at the conference on spatial inequality in Latin America (WIDER project on spatial disparities in human development). Universidad de las Americas-Puebla, Cholula, Mexico, 1-3 November 2002.

CORNWALL, K. & INDER, B. 2004. Migration and unemployment in South Africa: When motivation surpasses the theory. Monash econometrics and business statistics working papers, 2/04.

DALGAARD, C-J. & VASTRUP, J. 2001. On the measurement of σ-convergence. *Economic letters*, 70(2):283-287, February.

DAVIS, D.R. & WEINSTEIN, D.E. 1999. Economic geography and regional production structure: An empirical investigation. *European economic review*, 43(2):379-407, February.

DE LA FUENTE, A. 1997. The empirics of growth and convergence: A selective review. *Journal of economic dynamics and control*, 21(1):23-73, January.

DEHLEN, G.L. 1993. Efficiency of transport to enhance South Africa's international competitiveness. Pretoria: CSIR Research Report R92/078.

DE LANGE, J. 2004. Waentjies trek weg uit Mookgopong. *Beeld*: May. 19.

DINKELMAN, T. & PIROUZ, F. 2002. Individual, household and regional determinants of labour force attachment in South Africa: Evidence for the 1997 October Household Survey. *South African journal of economics*, 70(5):865-891, June.

DIPPENAAR, A.A. 2001. Industrial development zones as a policy instrument in South Africa. Potchefstroom: P.U. for C.H.E. (Dissertation – M.Com.).

DIXIT, A. & STIGLITZ, J. 1977. Monopolistic competition and optimal product diversity. American economic review, 67:297-308.

DU PISANIE, J.A. 2002. Competition versus co-operation: The case of SADC's regional development corridors. *South African journal of economics*, 70(2):310-333, March.

DURANTON, G. 1998. Labour specialisation, transport costs and city size. *Journal of regional science*, 38(4): 533-573, November.

DURLAUF, S.N. & QUAH, D.T. 1999. The new empirics of economic growth. (*In:* Taylor, J.B. & Woodford, M. *eds.* Handbook of Macroeconomics, Volume 1A. New York: North-Holland. p.235-308.)

EATON, J. & ECKSTEIN, Z. 1997. Cities and growth: Theory and evidence from France and Japan. *Regional science and urban economics*, 27:443-373.

EHLERS, M.B. & ANTONITES, A.J. 2000. An evaluation of the opportunities and obstacles for investments in the Maputo Development Corridor. *Journal of African business*, 1(1):41-63.

ELLISON, G. & GLAESER, E.L. 1999. The geographic concentration of industry – Does natural advantage explain agglomeration? *Journal of political economy*, 105(5):889-927, October.

FEDDERKE, J.W. 2001. Explaining the growth absence: Reviewing the evidence that can account for the poor growth performance of the South African economy. Paper presented at the TIPS Annual Forum. Misty Hills, Muldersdrif, South Africa.

FERREIRA, A. 2000. Convergence in Brazil: Recent trends and long-run prospects. *Applied economics*, 32(4):79-90, March.

FRANKEL, J.A. & ROMER, D. 1999. Does trade cause growth? *American economic review*, 89(3):379-399, June.

FUJITA, M., KRUGMAN, P. & VENABLES, A. 1999. The spatial economy: Cities, regions, and international trade. Cambridge MA: MIT Press.

FUJITA, M. & THISSE, J-F. 2000. Economics of agglomeration. (*In:* Huriot, J-M. and Thisse, J-F., *eds.* The economics of cities: Theoretical perspectives. Cambridge: Cambridge University Press.)

FUJITA, M. & THISSE, J-F. 2002. Economics of agglomeration. Cambridge: Cambridge University Press.

GABAIX, X. 1999. Zipf's law for cities: An explanation. *Quarterly journal of economics*, 114(3):739-767, August.

GALLUP, J.L., SACHS, J.D. & MELLINGER, A.D. 1998. Geography and economic development. (*In:* Pleskovic, B. & Stiglitz, J.E., *eds*, Annual conference on development economics. Washington D.C.: World Bank. p.127-178.)

GALLUP, J.L., SACHS, J.D. & MELLINGER, A.D. 1999. Geography and economic development. *International regional science review*, 22(2):179-232, August.

GALLUP, J.L. & SACHS, J.D. 2000. The economic burden of malaria. CID Working Paper no. 52, Harvard University.

GLEASER, E., KALLAL, H. SCHEINKMAN, J. & SHLEIFER, A. 1992. Growth in cities. *Journal of political economy*, 100(6):1126-1152, December.

GLAESER, E., SCHEINKMAN, J. & SHLEIFER, A. 1995. Economic growth in a cross-section of cities. *Journal of monetary economics*, 36(1):117-143, December.

GROSSMAN, G. M. & HELPMAN, E. 1991. Trade, knowledge spillovers and growth. *European economic review*, 35(2-3): 517-526, April.

GROSSMAN, G. M. & HELPMAN, E. 1994. Endogenous innovation in the theory of growth. *Journal of economic perspectives*, 8(1):23-44, Winter.

GROTE, M. 2000. Aspects of fiscal devolution in South Africa. *South African journal of economic and management sciences*, 3(1):59-74, March.

GYIMAH-BREMPONG & TRAYNOR, T.L. 1999. Political instability, investment and economic growth in sub-Saharan Africa. *Journal of African economies*, 8(1):52-86, March.

HALL, R.E. & JONES, C.I. 1997. Levels of economic activity across countries. *American economic review*, AEA papers and proceedings, 87(2):173-176, May.

HALL, R.E. & JONES, C.I. 1999. Why do some countries produce so much more output per worker than others? *Quarterly journal of economics*, 114(1):83-116, February.

HANSON, G. 1997. Increasing returns, trade and the regional structure of wages. *Economic journal*, 107(440):113-133, January.

HANSON, G. 1998. Market potential, increasing returns and geographic concentration. NBER Working Paper no. 6429, Cambridge MA: The NBER.

HANSON, G.H. 2000. Scale economies and the geographic concentration of industry. NBER Working Paper no. 8013, Cambridge MA: The NBER.

HARRIS, C.D. 1954. The market as a factor in the localisation of industry in the United States. *Annals of the association of American geographers*, 44:315-348.

HARRISON, P. & TODES, A. 2001. The use of spatial frameworks in regional development in South Africa. *Regional studies*, 35(1):65-72, February.

HARTZENBERG, T. 2001. South African regional industrial policy: From border industries to Spatial Development Initiatives. *Journal of international development*, 13(6):767-777, August.

HELSLEY, R.W. & STRANGE, W.C. 1990. Matching and agglomeration economies in a system of cities. *Regional science and urban economics*, 20(2):189-212, September.

HENDERSON, J.V. 1986. Efficiency of resource usage and city size. *Journal of urban economics*, 19(1):47-70, January.

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

HENDERSON, J.V. 1999. Marshall's scale economies. NBER Working Paper no. 7358, Cambridge MA: The NBER.

HENDERSON, J.V. 2000 How urban concentration affects economic growth. Working paper, infrastructure and environment, Development research group, World Bank. Washington DC.

HENDERSON, J.V., KUNCORO, A. & TURNER, M. 1995. Industrial development in cities. *Journal of political economy*, 103(5):1067-1090, October.

HENDRIKS, S. L. & LYNE, M.C. 2003. Agricultural growth multipliers for two communal areas of KwaZulu-Natal. *Development Southern Africa*, 20(3):423-444, September.

HIRSCHMAN, A.O. 1958. The strategy of economic development. New Haven: Yale University Press.

KEPE, T. 2001. Clearing the ground in the Spatial Development Initiatives: Analysing 'process' on South Africa's Wild Coast. Development Southern Africa, 18(3):279-293, September.

KHOSA, M.M. 2002. Perceptions of economic and financial conditions of households in 2000. *Development Southern Africa*, 19(5):681-697, December.

KIM, S. 1995. Expansion of markets and the geographic distribution of economic activities. *Quarterly journal of economics*, 110(4):881-908, September.

KIM, S. 1999. Regions, resources and economic geography: Sources of U.S. regional comparative advantage, 1880-1987. *Regional science and urban economics*, 29(1):1-32, January.

KLEYNHANS, E.P.J., NAUDé, W.A. & VAN DER MERWE, S.J. 2003. Spatial economic development in South Africa: An overview and evaluation of the Platinum Spatial Development Initiative. *Development Southern Africa*, 20(5):617-631, December.

KNUDSEN, T. 2001. Zipf's law for cities and beyond: The case of Denmark. *American journal of economics and sociology*, 60(1):123-146, January.

KOOPMANS, T. 1965. On the concept of optimal economic growth. The econometric approach to development planning. Amsterdam: North-Holland.

KRUGMAN, P.R. 1979. Increasing returns, monopolistic competition and international trade. *Journal of international economics*, 9(4):469-479, November.

KRUGMAN, P.R. 1980. Scale economies, product differentiation and the pattern of trade. *American economic review*, 70(5):950-959, December.

KRUGMAN, P.R. 1991. Geography and trade. Cambridge MA, MIT Press.

KRUGMAN, P.R. 1993. The rise and fall of development economics. [Web:] http://www.wws.princeton.edu/~pkrugman/dishpan.html [Date of access: 14 Apr. 2005]

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

KRUGMAN, P.R. 1998. What's new about new economic geography? *Oxford review of economic policy*, 14(2):7-17, Summer.

KRUGMAN, P.R. & VENABLES, A.J. 1990. Integration and the competitiveness of peripheral industry. (*In* Bliss, C. & Braga de Macedo, J., *eds*, Unity and diversity in the European economy. Cambridge: Cambridge University Press. p.56-75.)

LEAMER, E.E. 1997. Access to western markets and eastern effort. (*In* Zecchini, S., *ed.*, Lessons from the economic transition, central and eastern Europe in the 1990s. Dordrecht: Kluwer Academic Publishers. p.503-526.)

LEIBBRANDT, M. & WOOLARD, I. 1999. A comparison of poverty in South Africa's nine provinces. *Development Southern Africa*, 16(1):37-54, Autumn.

LEIBBRANDT, M., WOOLARD, I. & BHORAT, H. 2000. Understanding contemporary household inequality in South Africa. *Journal for studies in economics and econometrics*, 24(3):31-51, November.

LEIBBRANDT, M., BHORAT, H. & WOOLARD, I. 2001. Household inequality and the labour market in South Africa. *Contemporary economic policy*, 19(1):73-86, January.

LEMON, A. & ROGERSON, C.M. 2002. Geography and economy in South Africa and its neighbours: An introduction. (*In* Lemon, A. & Rogerson, C.M. *eds*. Geography and economy in South Africa and its neighbours. Aldershot: Ashgate. p. 1-24.)

LE ROUX BOOYSEN, F. 2002. Using demographic and health surveys to measure poverty – An application to South Africa. *Journal for studies in economics and econometrics*, 26(3):53-70, November.

LE ROUX BOOYSEN, F. 2003a. The extent of, and possible explanations for, provincial disparities in progress on reconstruction and development in South Africa. *Development Southern Africa*, 20(1):21-48, March.

LE ROUX BOOYSEN, F. 2003b. Urban-rural inequalities in health care delivery in South Africa. *Development Southern Africa*, 20(5):659-674, December.

LIM, U. 2003. A spatial analysis of regional income convergence. *Planning forum*, 9:66-80.

LIMAO, N. & VENABLES, A.J. 2001. Infrastructure, geographical disadvantage and transport costs. *World Bank economic review*, 15(3): 451-479.

LOTTER, J.C. 2002. Urban economic empowerment in the greater Pretoria area: The continuation of history? *South African journal of economics*, 70(2):347-368, March.

LUCAS, R.E. 1988. On the Mechanics of Economic Development. *Journal of Monetary Economics*, 22:3-42.

MAHONY, K. & VAN ZYL, J. 2002. The impacts of tourism investment on rural communities: Three case studies of South Africa. *Development Southern Africa*, 19(1):83-1003, March.

MANKIW, N.G., ROMER, D. & WEIL, D.N. 1992. A contribution to the empirics of economic growth. *Quarterly journal of economics*, 107(2):407-437, May.

MARAIS, L. & KRIGE, S. 2000. Who received what, where in the Free State? An assessment of post-apartheid housing delivery and policy (1994-98). *Development Southern Africa*, 17(4):603-619, October.

MARTIN, R. 1999. The new 'geographical turn' in economics: Some reflections. *Cambridge journal of economics*, 23(1):65-91, January.

MARTIN, R. & SUNLEY, P. 1996. Paul Krugman's "Geographical Economics" and its implications for regional development theory: A critical assessment. *Economic geography*, 72(3):259-292, July.

McDONALD, S. & PUNT, C. 2002. Supply constraints, export opportunities and agriculture in the Western Cape of South Africa. *Journal for studies in economics and econometrics*, 26(1):1-15, April.

McDONALD, S., PIESSE, J. & VAN ZYL, J. 2000. Exploring the distribution of household income in South Africa. *South African journal of economics*, 68(3):423-454, September.

MIDELFART-KNARVIK, K.H., OVERMAN, H.G., REDDING, S.J. & VENABLES, A.J. 2000. The location of European industry. *Economic papers*.

MOLLER, V. 2001. Monitoring quality of life in cities: The Durban case. *Development Southern Africa*, 18(2):217-238, June.

MORETTI, E. 1999. Estimating the social return to education: Evidence from longitudinal and repeated cross-sectional data. *Journal of econometrics*, 121(1-2):175-212, July/August.

MYRDAL, G. 1957. Economic theory and under-developed regions. London: Duckworth.

NAGARAJ, R., VAROUDAKIS, A. & VENGANZONES, M.A. 2000. Long-run growth trends and convergence across Indian states. *Journal of international development*, 12(1):45-70, January.

NAKAMURA, R. 1985. Agglomeration economies in urban manufacturing industries: A case of Japanese cities. *Journal of urban economics*, 17(1):108-124, January.

NAUDé, W.A. 2001a. Local government transformation in South Africa: Implications for local and regional economic development. Development ISSues, Institute for Social Studies, The Hague, 3(2).

NAUDé, W.A. 2001b. The elasticity of geography and local economic development in the information age. Paper presented at a workshop on Local Economic Development, Klerksdorp Local Municipality, 17 April.

NAUDé, W.A. 2005. Book review: Geography and economy in South Africa and its neighbours. *Journal of African economies*, forthcoming.

NAUDé, W.A. & JANSEN VAN RENSBURG, L. 2002. Understanding poverty as a human rights violation: The case of legal reform and local economic development in South Africa. Paper presented at the 6th Annual conference of the Centre for the Study of African Economies on "Understanding poverty and growth in sub-Saharan Africa", St. Catherine's College, University of Oxford, 18 March 2002.

NAUDé, W.A. & KRUGELL, W.F. 2003a. Are South Africa's cities too small? *Cities: The international journal of urban policy and planning*, 20(3):175-180.

NAUDé, W.A. & KRUGELL, W.F. 2003b. Spatial inequality in Africa. (Review of the conference on spatial inequality in Africa (WIDER project on spatial disparities in human development), Centre for the study of African economies, University of Oxford.) *Development Southern Africa*, 20(1): 161-167.

NAUDé, W.A. & KRUGELL, W.F. 2003c. An enquiry into cities and their role in sub-national economic growth in South Africa. *Journal of African economies*, 12(4): 476-499.

NAUDé, W.A. & KRUGELL, W.F. 2004. An enquiry into cities and their role in sub-national economic growth in South Africa. WIDER research paper RP2004/08.

NAUDé, W.A. & KRUGELL, W.F. 2005. Economic geography and growth in Africa: The determinants of sub-national growth in South Africa. *Papers in regional science*, forthcoming.

NAUDé, W.A., KRUGELL, W.F. & SERUMAGA-ZAKE, P. 2002. Cumulative causation and decentralised industrial development in South Africa. (*In*: Higano, Y., Nijkamp, P., Poot, J. & Van Wijk, J.J. *eds*. The Region in the New Economy. London: Ashgate. p. 407-430).

NAUDé, W.A., SAAYMAN, A. & SAAYMAN, M. 2000. The Impact of Tourism Spending in South Africa: Spatial Implications. *South African journal for economic and management sciences*, 3:369-386, September.

NAUDé, W.A. & SERUMAGA-ZAKE, P. 2001. An analysis of the determinants of labour force participation and unemployment in South Africa's North West Province. *Development Southern Africa*, 18(3):261-278, September.

NEARY, J.P. 2001. Of hype and hyperbolas: Introducing the New Economic Geography. *Journal of economic literature*, 39(2):536-561, June.

NEL, E. 2001. Local economic development: A review and assessment of its current status in South Africa. *Urban studies*, 38(7):1003-1024, June.

NEL, E. 2002. South Africa's manufacturing economy: Problems and performance. (*In:* Lemon, A. & Rogerson, C.M. *eds.* Geography and economy in South Africa and its neighbours. Aldershot: Ashgate. p. 81-94.)

NEL, E. & BINNS, T. 2001. Initiating 'Developmental Local Government' in South Africa: Evolving local economic development policy. *Regional studies*, 35(4):355-362, June.

NERLOVE, M. 2000. Growth rate convergence, fact or artefact? Essays in panel data econometrics. Cambridge: Cambridge University Press.

NICKELL, S. 1981. Biases in dynamic models using fixed effects. *Econometrica*, 49:1417-1426.

NIJKAMP, P. 2001. Review of Fujita, Krugman and Venables. Spatial economy: Cities, regions and international trade. *The economic journal*, 111(469):F166-F168, February.

NIKSIC, G. 2004. Difficult but not impossible: The ANC's decentralisation strategy in South Africa. *Development and change*, 35(2):353-374, April.

NKURUNZIZA, J.D. & BATES, R.H. 2003. Political institutions and economic growth in Africa. CID Working Paper no. 98, Harvard University.

NOMDO, C. & COETZEE, E. 2002. Urban vulnerability: Perspectives from Southern Africa. Sterling, Va.: Stylus.

NTSEBEZA, L. 2004. Democratic decentralization and traditional authority: Dilemmas of land administration in rural South Africa. *European journal of development research*, 16(1):71-89, Spring.

OOSTHUIZEN, M.J. & NIEUWOUDT, L. 2003. A poverty profile of the Western Cape Province of South Africa. *Journal of studies in economics and econometrics*, 27(3):69-90, November.

OTTAVIANO, G.I.P. 2003. Regional policy in the global economy: Insights from New Economic Geography. *Regional studies*, 37(6&7):665-673, August/October.

OTTAVIANO, G.I.P. & PUGA, D. 1998. Agglomeration in the global economy: A survey of the 'New Economic Geography'. *World economy*, 21(6):707-731, August.

OVERMAN, H.G., REDDING, S. & VENABLES, A.J. 2001. The economic geography of trade, production, and income: A survey of empirics. CEPR Paper no. 465, London.

PORTER, M.E. 1998. Location, clusters and the 'new' microeconomics of competition, *Business economics*, 33(1):7-13, January.

POSEL, D. 2003. The collection of National Household Survey data in South Africa (1993-2001): Rendering labour migration invisible. *Development Southern Africa*, 20(3):361-368, September.

POSEL, D. 2004. Have migration patterns in post-apartheid South Africa changed? *Journal of interdisciplinary economics*, 15(3-4):277-292, Special issue.

PUGA, D. 1998. Urbanisation patterns: European versus less developed countries. *Journal of regional science*, 38(2):231-252, May.

PUSHKAR, M. & RAY, R. 2003. Resource inflows and household composition: Evidence from South African panel data. *Journal of international development*, 15(8):1037-1047, November.

PUSHKAR, M. 2002. The effect of household characteristics on poverty and living standards in South Africa. *Journal of economic development*, 27(1):75-96, June.

RAMSEY, F.P. 1928. A mathematical theory of saving. *Economic journal*, 38:543-559.

RAUCH, J.E. 1993. Productivity gains from geographic concentration of human capital: Evidence from cities. *Journal of urban economics*, 34(3):380-400, November.

RAUCH, J.E. & TRINDADE, V. 1999. Ethnic Chinese networks in international trade. NBER Working Paper no. 7189, Cambridge MA: The NBER.

REGIONAL ECONOMIC FOCUS. 2005. Database supplied by Global Insight Southern Africa.

REDDING, S. & VENABLES, A.J. 2000. Economic geography and international inequality. CEPR Discussion Paper no. 2568, London.

ROBACK, J. 1982. Wages, rents and the quality of life. *Journal of political economy*, 90(6):1257-1278, December.

ROBINSON, P.S. 2003. Implications of rural settlement patterns for development: A historical case study of Qaukeni, Eastern Cape, South Africa. *Development Southern Africa*, 20(3):405-421, September.

RODDEN, J., ESKELAND, G.S. & LITVACK, J. 2000. Fiscal decentralisation and the challenge of hard budget constraints. London: MIT Press.

RODRIK, D., SUBRAMANIAN, A. & TREBBI, F. 2002. Institutions rule: The primacy of institutions over geography and integration in economic development. CID Working Paper no. 97, Harvard University.

ROGERSON, C.M. 1991. Beyond racial Fordism: Restructuring industry in the "new" South Africa. *Tijdschrift voor economische en sociale geografie*, 82:355-366.

ROGERSON, C.M. 2001a. Redressing urban poverty in post-apartheid South Africa. (In, CROP international studies in poverty research, Poverty reduction: What role for the state in today's globalised economy? New York: Palgrave. p.341-362.)

ROGERSON, C.M. 2001b. Spatial Development Initiatives in Southern Africa: The Maputo Development Corridor. *Tijdschrift voor economische en sociale geografie*, 92:324-346.

ROGERSON, C.M. 2002a. Tourism – A new economic driver for South Africa. (*In:* Lemon, A. & Rogerson, C.M. *eds.* Geography and economy in South Africa and its neighbours. Aldershot: Ashgate. p. 81-94.)

ROGERSON, C.M. 2002b. Tourism and local economic development: The case of the Highlands meander. *Development Southern Africa*, 19(1):143-167, March.

ROMER, P.M. 1986. Increasing returns and long-run growth. *Journal of political economy*, 94(5):1002-1037, October.

ROSEN, K.T. & RESNICK, M. 1980. The size distribution of cities: An examination of the Pareto law and primacy. *Journal of urban economics*, 8(2):165-186, September.

ROSENSTEIN-RODAN, P. 1943. Problems of industrialisation in eastern and south-eastern Europe. Economic journal, 53:202-211.

RUBINFELD, D.L. 2001. Structuring intergovernmental grants to local government: Lessons from South Africa. *Constitutional political economy*, 12(2):173-187, June.

SACHS, J. 2001. Tropical underdevelopment. NBER Working Paper no. 8119, Cambridge MA: The NBER.

SACHS, J.D. 2003. Institutions matter, but not for everything. *Finance and development*, 40(2):38-41, June.

SACHS, J.D. & WARNER, A.M. 1997. Sources of slow growth in African economies. *Journal of African economies*, 6(3):335-376, October.

SACHS, J.D. & WARNER, A.M. 1999. The big push, natural resource booms and growth. *Journal of development economics*, 59(1):43-76, June.

SERUMAGA-ZAKE, P. & NAUDé, W.A. 2001. Labour market discrimination in the North West Province of South Africa. *Development Southern Africa*, 18(5):671-682, December.

SERUMAGA-ZAKE, P. & NAUDé, W.A. 2002. The determinants of rural and urban household poverty in the North West Province of South Africa. *Development Southern Africa*, 19(4):561-571, October.

SMITH, L. & HANSON, S. 2003. Access to water for the urban poor in Cape Town: Where equity meets cost recovery. *Urban studies*, 40(8):1517-1548, July.

SODERBAUM, F. & TAYLOR, I. 2003. Regionalism and uneven development in South Africa: The case of the Maputo Development Corridor. Aldershot: Ashgate.

SOLOW, R. 1956. A contribution to the theory of economic growth. *Quarterly journal of economics*, 70:65-94.

SOUTH AFRICAN CITIES NETWORK. 2004. State of the cities report 2004. [Web:] http://www.sacities.net [Date accessed: 14 April 2005]

SULEMAN, A. 1998. The competitiveness and comparative advantages of the manufacturing sector in South Africa's nine provinces. Unpublished PhD thesis, School of Economics, Risk Management and International Trade, Potchefstroom University, South Africa.

SULEMAN, A. & NAUDé, W.A. 2003. The competitiveness of South African manufacturing: A spatial view. Journal for studies in economics and econometrics, 27(2):29-52, August.

TEMPLE, J. 1999. The new growth evidence. *Journal of economic literature*, 37(1):112-156, March.

TRIVEDI, K. 2002. Regional convergence and catch-up in India between 1960 and 1992. [Web:] http://www.nuff.ox.ac.uk/economics/papers/2003/W1/convergencewp1.pdf [Date accessed: 14 April 2005]

TSANGARIDES, C.G. 2001. On cross-country growth and convergence: Evidence from African and OECD countries. *Journal of African economies*, 10(4): 355-389, December.

THSEOLA, J. 2002. Basic needs in the Northern Province and South Africa's globalisation agenda. *African development review*, 14(1):48-74, June.

UNITED NATIONS. 2001. World Urbanisation Prospects.

VISSER, V. 2001. Social justice, Integrated Development Planning and post-apartheid South Africa. *Urban studies*, 38(10):1673-1699, September.

WARNER, A.M. 2002. Institutions, geography, regions, countries and the mobility bias. CID Working Paper no. 91, Harvard University.

WHEATON, W.C. & SHISHIDO, H. 1981. Urban concentration, agglomeration economies and the level of economic development. *Economic development and cultural change*, 30(1):17-30, October.

WHITE, H. 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4): 817-838, May.

WORLD BANK. 2000. World Development Report 1999/2000. Entering the 21st Century. Washington DC : The World Bank.

YOUNG, A. 1993. Invention and bounded learning by doing. *Journal of political economy*, 101(3): 443-472, June.

ZIPF, G.K. 1949. Human behaviour and the principle of least effort. New York: Addison-Wesley.