Regional Wage Disparities in Post-Apartheid South Africa: Spatial Patterns, Convergence Dynamics, and Causes

By

Gibson Mudiriza

Thesis Presented for the Degree of

DOCTOR OF PHILOSOPHY

in the

School of Economics

UNIVERSITY OF CAPE TOWN

October 2017

Supervisor:

Professor Lawrence Edwards

Abstract

This thesis examines the spatial patterns, convergence dynamics, and causes of regional wage disparities in post-apartheid South Africa. The analysis is conducted using a unique dataset for 354 regions constructed from the 1996, 2001 and 2011 population censuses. The thesis comprises of six chapters including the introduction and conclusion. The general introduction is followed by Chapter two that presents a theoretical and empirical review of the causes of regional wage disparities. Emphasis is placed on theoretical insights derived from the new economic geography theory. The subsequent 3 chapters comprise the empirical analysis.

Chapter three applies exploratory spatial data analysis techniques to the 1996, 2001 and 2011 population censuses data and examines the spatial patterns that characterise the distribution of wages across regions in South Africa. An objective is to assess the consistency of these patterns with predictions from alternative theories. The results point to significant disparities in the distribution of wages across regions in South Africa that are greater than regional wage disparities in other countries. These disparities are characterised by positive and negative spatial autocorrelation between regions suggesting the coexistence of new economic geography forces (positive autocorrelation) and alternative economic theory (negative autocorrelation) features. These findings provide important context and input into the subsequent empirical chapters of the thesis.

Chapter four examines the convergence dynamics of wages across regions in post-apartheid South Africa. The aim of the chapter is to assess the extent to which wages have converged or diverged across regions in South Africa over the period 1996 – 2011. The convergence dynamics are analysed using a range of complementary measures, namely, the kernel density estimator and indicators of σ - convergence and β - convergence. These measures consistently reveal evidence of regional wage divergence between 1996 and 2001, but strong and robust evidence of regional wage convergence from 2001 to 2011. During the period 2001 – 2011, the unconditional β- convergence is estimated at 3.7% per year, suggesting that it could take about 19 years to reduce the gap in wages between rich and poor regions by half. However, conditional β- convergence, after controlling for initial human capital, local unemployment, market potential, industrial stucture and homeland status, is much higher, at 13% per year. This implies that it could have taken about 5 years to reduce the wage gap between rich and poor regions by half, if differences in regional specific factors had been eliminated between 2001

and 2011. The results suggest that regional specific factors, some of which can be influenced by policy, constrain regional wage convergence in South Africa.

The fifth Chapter empirically tests whether the prediction of the Helpman-Hanson wage equation derived from the new economic geography theory is consistent with observed regional wage disparities in South Africa. The chapter extends the Helpman-Hanson model to include other potential explanatory factors concerning regional wages. The results suggest that regional wage disparities in South Africa are well explained by new economic geography forces such as access to markets, but only after controlling for regional specific factors such as human capital, mineral resource endowments, local climatic conditions, local unemployment, and homeland status. The findings of this study imply that new economic geography theory alone is not sufficient to explain regional wage disparities in South Africa. Its proper application hinges on the incorporation of other regional specific factors. This finding is consistent with an emerging economy that is characterised by moderate industrial and transport sectors, on the one hand, and a strong primary sector driven by natural resource exploitation, on the other hand.

Overall, the thesis shows that, while regional wage disparities are converging over time in South Africa, they remain high compared to other countries. Convergence is very slow and is influenced by regional specific factors such as human capital, access to markets, mineral resource endowments, local climatic conditions, local unemployment and homeland status. Accordingly, complementary policies promoting human capital accumulation, access to markets, and development of former homeland areas, and improving labour market outcomes will substantially enhance wage convergence. Nevertheless, even with these policies, regional wage disparities will remain a feature of the South African economy as these are in part driven by economic forces associated with new economic geography. The thesis thus highlights how differences in access to markets and regional factor endowments exacerbate and constrain regional wage disparities in South Africa.

Declaration

I, *Gibson Mudiriza*, declare that this thesis is my own work and that the material included in this dissertation is the result of new research, and other sources have been acknowledged through referencing. I also declare that this thesis has not been submitted for a PhD degree in any another university.

Acknowledgement

I would like to take this opportunity to express my sincere gratitude to the people who have played a fundamental role in the successful completion of my PhD Study. First, my greatest appreciation goes to my supervisor Professor Lawrence Edwards for his constructive support, comments, guidance, and academic mentorship throughout this research. His assistance in solving technical problems on various chapters of the thesis has made an enormous contribution to the research and played a significant role in the eventual completion of the thesis. He also went the extra mile by financially assisting me in purchasing the necessary software that was needed to effectively complete a crucial section of the study. It was an honour and privilege to have worked with him.

Financially, I acknowledge the African Economic Research Consortium, the South African National Research Foundation (CSUR 14080989186, Grant No: 93648), UCT research scholarship and Carnegie for financially supporting my research. Without the financial support of these organisations, I would not have been able to pursue doctoral studies on a full-time basis. I am once again grateful to my supervisor, Professor Lawrence Edwards in this regard. Moreover, I want to thank DataFirst for making the South African population censuses data readily available.

I am most grateful to my fellow PhD students in the School of Economics, those who served in the Trade Reading Group and the PhD Brownbag seminar for lending me great support and encouragement along the way and for their useful comments and suggestions. I am extremely indebted to my friend, Alfred Mukong for imparting to me his expertise and providing constructive insights that helped me with some of the technical problems I had, and in shaping my thesis.

Special mention goes to my friends, Godfrey Mahofa, Refilwe Lepelle, Herbert Ntuli and Pinky Kebakile with whom I have studied, worked and laughed with. I have learnt a lot from our intellectual discussions on life, our academic studies and current affairs and I am also grateful for the love and support you gave me when my best friend, Adonia Chiminya passed away. To my brother Adonia Chiminya, we came together to pursue our PhD studies and hoped to finish this journey together but life happened. Thank you for your support and companionship. You made my PhD years bearable and fun. May your soul rest in peace.

In conclusion, my heartfelt gratitude goes to my family. I thank my parents, Charles and Masimbo Mudiriza for sacrificing so much for me to become the person I am today, and for being devoted to all their children, regardless of the circumstances. To my brothers and sisters thank you for your love and support. To my aunt and late uncle Masaraure and their children, thank you for being my role models of excellence. You nurtured and encouraged me to excel in life and in my academic career. Finally, to my wife Kudzai and son Mukundi, you have been constant in your love, prayers and support, and believed in me more than I did in myself.

Glory be to God Almighty.

Dedication

To my universe: my wife Kudzai, our son Mukundi and my parents

and

To my late uncle, Mr Masaraure and friend, Adonia Chiminya.

Table of Contents

List of Tables

List of Figures

List of Acronyms

AERC	African Economic Research Consortium
ESDA	Exploratory Spatial Data Analysis
GDP	Gross Domestic Product
GIS	Geographic Information Systems
LFS	Labour Force Surveys
LISA	Local Indicators of Spatial Association
LMDSA	Labour Market Dynamics in South Africa
MAUP	Modifiable Areal Unit Problem
NEG	The New Economic Geography
NIDS	National Income Dynamics Study
OHS	October Household Survey
PALMS	Post-Apartheid Labour Market Series
QLFS	Quarterly Labour Force Survey
SALDRU	Southern Africa Labour and Development Research Unit
Stats SA	Statistics South Africa
UCT	University of Cape Town

Chapter 1

1. Background and Motivation of the Study

1.1. Introduction

Regional wage disparities are known to be large in both developed and developing countries and are often a source of public concern. Globally, growing regional wage disparities are known to reduce well-being, social cohesion, and economic growth (Oshchepkov, 2015; International Labour Office, 2015). This is especially true in many developing countries, where regional disparities in wages and income coincide with the spatial distribution of ethnic groups or natural resources (Breinlich, Ottaviano, & Temple, 2014)[1]. Moreover, empirical evidence reveals that, in many developing countries, low productivity, social dissatisfaction, negative externalities and high levels of regional inflation, among others, are partly a consequence of regional disparities in income and wages (Henderson, 2003; Cherodian & Thirlwall, 2015). Considering these negative consequences, most governments implement policies to address regional wage disparities and promote economic growth. But the new economic geography (NEG) theory has shown that regional wage disparities are an outcome of economic forces that can conflict with policy measures aimed at reducing wage differences[2]. Properly targeted policies, therefore, require a thorough understanding of the nature, convergence dynamics, and sources of regional wage disparities.

This thesis aims to undertake a comprehensive examination of regional wage disparities in South Africa. It focuses on three issues, namely: (1) an in-depth analysis of the spatial distribution of wages across regions, (2) an analysis of the convergence dynamics of regional wages and (3) a test of the applicability of NEG in explaining regional wage disparities in South Africa, a middle-income country, with a large resource base. Each analysis is covered by a distinct chapter in this thesis. They contribute towards answering the following main question:

[1] In most developing countries, especially in Africa, political, racial, and tribal differences play a significant role in the development not only of people, but regions. Regions for elite politicians, dominate tribal groupings, and historically advantaged races are usually more affluent than other regions and the influence of these groups leads to resources being diverted to their areas.
[2] An example of this is the wage bargaining and minimum wage legalisations in South Africa where wage differences between rural and urban areas have been argued to be too low as the rural wage is set higher than economic forces in rural areas can accommodate.

how and why do wages differ across regions in South Africa? Addressing this question in the case of South Africa is important for several reasons.

Firstly, South Africa is well-known for having inequalities that are among the highest in the world. A Gini index of 0.66 for income and 0.47 for wages between 2008 and 2012, according to Finn & Leibbrandt (2013), gives an idea of overall inequalities in South Africa. The high and persistent income inequality in post-Apartheid South Africa, which is a fundamental problem in the country, has been attributed to large and increasing dispersion in labour market earnings (Ntuli & Kwenda, 2014; Wittenberg, 2016, 2017). Furthermore, the high-income inequality in the country seems to also be driven by regional wage disparities. Established international evidence suggests that regional wage disparities contribute significantly to total wage and income inequality (Oshchepkov, 2007).

However, to the best of our knowledge and at the inception of this thesis, the analysis of regional wage disparities has rarely been the main subject of research in South Africa. The bulk of existing research has focused on overall income (Leibbrandt, Poswell, Naidoo, Welch, & Woolard, 2005; Murray Leibbrandt, Finn, & Woolard, 2012) and wage (Burger & Yu, 2007; Burger, 2015; Ntuli & Kwenda, 2014; Wittenberg, 2014, 2016, 2017) inequality. A few studies focus on regional wage disparities. These include Kingdon & Knight (2006), Magruder (2012), and von Fintel (2017)[3]. The limited research on regional wage disparities in the case of South Africa provides substantial scope for further research in this area.

Secondly, South Africa has a high degree of spatial income inequality that seems to be worsening over time. For example, using regional GDP per capita data drawn from the Regional Economic Explorer (REX) database compiled by Global Insight Southern Africa, Krugell, Koekemoer, & Allison (2005), and Bosker & Krugell (2008) find evidence of increasing regional income inequalities over the period 1996 and 2004. However, little is known about what has happened to spatial income inequalities beyond 2004, or whether the same trends can still be observed using a different dataset or economic indicator such as regional wages. To fully understand the South African spatial economy, there is need to update

[3] Magruder (2012) used magisterial districts as the unit of analysis to examine the effects of bargaining council on various labour market outcomes (employment, employment by firm size and wages by industry), while Kingdon and Knight (2006), as well as von Fintel (2017) used individual workers identified by their location (360 clusters, magisterial districts, districts councils and provinces) as the unit of analysis to examine the effects of local unemployment rate on individual wage.

and verify these trends using more recent data (or alternative data) or different economic indicator.

Thirdly, theoretically, there are solid economic reasons that determine regional wage disparities. Three causes are mainly identified in the literature: Spatial differences in the skills composition of the workforce; differences in nonhuman endowments; and differences in spatial interactions between workers or firms (Combes, Duranton, & Gobillon, 2008). Given these causes, we do not know whether the regional wage disparities we see in South Africa are well-explained by existing economic theory. Are wage differences across regions in South Africa consistent with theoretical predictions? Addressing this question is critical as it provides important insights into the sources of regional wage disparities.

These insights can provide valuable information that can assist in the design of effective policies aimed at addressing regional wage disparities and promoting equalisation of standards of living. Furthermore, the insights can also provide useful information that potentially challenges the premise of some policy initiatives promoting equal distribution of wages across regions[4]. This is because some areas might have strong economic forces that attract and hold economic activities leading to higher income and wage levels in these regions than in other areas. Regional policy initiatives that are inconsistent with these economic forces may, therefore, conflict with economic realities. There is, therefore, a need for spatial development policies that are consistent with the economic forces driving regional wage disparities. The design of such policies requires a thorough understanding of the precise causes of regional wage disparities.

1.2. Theory and evidence on regional wage disparities

The literature has addressed the problem of how and why wages differ across regions from different perspectives. On the one hand, neoclassical economic theory (growth and trade models), suggests that, while regional wage disparities initially arise in the process of reallocation of resources, these disparities decrease over time. This leads to regional wage convergence, driven by diminishing returns to capital, interregional trade and factor mobility (Hofer & Wörgötter, 1997). However, this view has generated considerable controversy in the theoretical and empirical literature.

[4] An example of such initiatives in South Africa includes regional industrial policies such as the Industrial Development Zones (IDZ) and labour market legalisation such as wage bargaining and minimum wage legalisation.

Several alternative theories, among them human capital theory, endogenous growth theory, wage curve theory and amenity theory, suggest that regional wage disparities can persist or even increase, leading to regional wage divergence over time. To explain these disparities, human capital theory and endogenous growth theory highlight the importance of differences in human capital (Becker, 1962; Romer, 1986). The wage curve theory points to the significance of differences in local unemployment (Blanchflower & Oswald, 1995). The amenity theory suggests the importance of variation in local (dis)amenities (Roback, 1982). Another school of thought, the NEG theory, predicts neither an increase nor a decrease in regional wage disparities. However, it argues that access to markets driven by the interplay of increasing returns to scale, transport costs and consumer's love of variety play a significant role in the determination of wages across regions (Krugman 1991).

Numerous studies demonstrate empirically that wages vary significantly across regions in many countries (among them Lindley & Machin, 2013, 2014; Candelaria, Daly, & Hale, 2015; Lee, Sissons, & Jones, 2016). Substantial research literature has also tested the neoclassical economic theory prediction of regional wage convergence and found mixed evidence (Rosés & Sánchez-Alonso, 2000; Maza & Villaverde 2006; Zaman & Goschin, 2014; Ferens, 2015). Evidence has been affected by the underlying measures of convergence employed, statistical methods used and the country and time-period of the analysis. Furthermore, several studies have shown evidence of the importance of differences in human capital (Combes, Duranton, & Gobillon, 2008; Fally, Paillacar, & Terra, 2010; Acemoglu & Dell, 2010), local unemployment (Ramos, Nicodemo, & Sanromá, 2015; von Fintel, 2017), local amenities (Graves, Arthur, & Sexton, 1999; Deller, 2009; Partridge, Rickman, Ali, & Olfert, 2010) and access to markets (Brakman, Garretsen, & Schramm, 2004; Mion, 2004; Hanson, 2005; Kosfeld & Eckey, 2010) in explaining regional wage disparities in different countries.

While the studies cited above provide important insights into regional wage disparities, the bulk of these studies have so far focused attention on developed economies. Very little research has been conducted in developing countries, particularly in Africa. This is mainly because of the scarcity of spatially disaggregated and comparable regional wage data. This has constrained researchers and policymakers from understanding the nature, extent and potential causes of regional wage disparities in many African countries. This thesis resolves the data constraints in the case of South Africa by constructing a spatially disaggregated regional database using the full population censuses for South Africa for the years 1996, 2001, and 2011. This data, which is geographically consistent over time, is used to provide a detailed analysis of regional wage disparities in South Africa.

1.3. Objectives

The primary objective of this thesis is to provide new empirical evidence on how and why wages differ across regions in South Africa. To achieve this objective, the study starts by providing a theoretical and empirical review of the causes of regional wage disparities, paying attention to insights from the NEG theory. The thesis then constructs a unique regional level dataset for 354 magisterial districts (hereafter - regions) using 1996, 2001 and 2011 full population census data. The new dataset, which is geographically consistent over time, contributes toward resolving the problem of incomparable regional data. This data is used to carry out three specific objectives related to regional wage disparities that form the empirical chapters around which this thesis is structured.

The first specific objective of this thesis is to examine the spatial patterns that characterise the distribution of wages across regions in South Africa and assesses their consistency with predictions from alternative economic theories. To achieve this, the study employs exploratory spatial data analysis (ESDA) techniques (Anselin & Bao, 1997; Anselin, 1997; Messner & Anselin, 2004). These techniques enable visualisation and exploration of spatial data to gain insights into the nature and extent of spatial patterns that characterise the distribution of wages across regions. Specifically, the study utilises box plots, choropleth maps and Moran's I statistic (Moran, 1948) to document the facts on how wage values in one location are spatially related to those in neighbouring regions. This relationship can be positive or negative, or there can be no relationship. The revealed facts enable us to evaluate if regional wage values exhibit random spatial patterns, which suggests even distribution of wages. Or whether they show spatial concentration patterns (that can either be positive or negative), which suggests the uneven distribution of wages. These spatial patterns allow us to compare the consistency of the empirical features found in the data (spatial patterns) with predictions from alternative theories, where positive spatial concentration patterns are consistent with NEG theory predictions. In contrast, negative spatial concentration patterns are in line with the predictions of alternative economic theories. Finally, these facts enable us to compare findings on South Africa with evidence from other countries.

The second specific objective of this research is to examine the convergence dynamics of wages across regions in post-apartheid South Africa. The aim is to assess the extent to which wages have converged or diverged across regions in South Africa from 1996 to 2011.The study uses the neoclassical economic theory framework that hypothesises that regional wages will converge over time to guide the analysis. This analysis is motived by the following research questions: Are wages converging or diverging across regions in South Africa over time? What are the potential drivers of the observed patterns of convergence or divergence? To address these questions, the study draws on three different but complementary measures of convergence, namely: kernel density estimator, σ-converegnce and β-convergence.

Firstly, the study employs the kernel density estimator to describe and analyse the extent of regional wage dispersion. To check for evidence of wage convergence or divergence, the analysis concentrates on the changes in the external shape dynamics of the entire cross-section wage distribution, focusing on its spread and skewedness. Secondly, the chapter employs the notion of σ- convergence, which provides a summary measure of the extent of the dispersion of wages across regions at a given point in time. Measuring dispersion with the standard deviation and coefficient of variation of wages, convergence is confirmed by a decrease in the estimate of σ over time, while divergence is revealed by an increase in the estimate. Finally, the chapter applies the notion of β- convergence, which allows for the estimation of the relationship between regional growth rate and initial levels of wages. Convergence is implied if the coefficient of initial wages is negative and statistically significant, while divergence is shown by a positive and statistically significant coefficient. The literature acknowledges the possibility of varying results from these different measures. By using both these measures, the study ensures that the results are not contingent on a specific measure.

The final specific objective of this thesis is to empirically test the validity of the NEG wage equation in the case of South Africa. The aim is to find out whether the mechanisms emphasised by the NEG theory explains the observed regional wage disparities in South Africa. This chapter is central to the thesis, as it focuses on explaining why wages differ across regions. Most South African research in this area estimates reduced-form models incorporating various variables inspired by the NEG theory. The novelty of this chapter of our study is the estimation of a structural wage equation based on the Helpman-Hanson model derived directly from the NEG theory. While this model has already been estimated (see Mion, 2004; Brakman et al. 2004; Kiso, 2005; Paredes, 2015), to the best of our knowledge, this study is the first to apply the model to the case of South Africa.

The central question addressed in this chapter is: Can the NEG theory explain the observed regional wage disparities in a developing country like South Africa? The NEG theory explains regional wage disparities based on very narrow economic forces derived from manufacturing and transport sectors. However, these sectors are less developed in South Africa than in developed countries where the theory has largely been tested. In contrast, like many emerging economies, South Africa has a robust primary sector, highly dependent on natural resources (minerals, agricultural land, access to waterways, and favourable climate). This factor is neglected by the NEG theory. However, natural resources, together with historical institutional settings and peculiar labour market conditions might be key determinants of regional wage levels in many emerging countries. The possible tension between these factors on the one hand, and the forces emphasised by the NEG theory, on the other hand, provides a unique testing ground for the validity of the NEG theory in explaining regional wage disparities in emerging economies.

In achieving these objectives, the thesis contributes to the literature in main ways. The first contribution is the construction of a spatially disaggregated regional database that is geographically consistent over time. Second, the study contributes to the regional science literature by providing empirical evidence on regional wage disparities, in the context of Africa, a region where there are few studies of this nature. Third, the research introduces the study of regional wage convergence as an added dimension to the existing regional convergence literature in South Africa. This literature has focused mainly on the convergence dynamics of GDP per capita using data covering the period 1990-2004.

Fourth, as a contribution to the NEG empirical literature, the thesis provides an empirical validation of the Helpman-Hanson model in the case of South Africa. It further contributes to the literature by extending the Helpman-Hanson NEG model to include other potential explanations for regional wage disparities. By so doing, the thesis provides a deeper understanding of the causes of regional wage disparities in South Africa. Finally, the thesis contributes to the practical regional policy debate in South Africa, where mitigating regional economic disparities and ensuring regional equalisation of living standards is regarded as a fundamental objective of the government. This objective can be aided by regional policy initiatives that are well-informed on the convergence dynamics and underlying causes of regional wage disparities.

1.4. Structure of the thesis

The remainder of the thesis is structured as follows. Chapter 2 provides an overview of theoretical and empirical insights on why wages differ across regions. Chapter 3 presents an exploratory analysis of the spatial distribution of wages across regions in South Africa. Chapter 4 examines the extent to which wages have converged or diverged across regions in South Africa over the period 1996 - 2011. In chapter 5, a key theoretical insight derived from the NEG theory on the causes of regional wage disparities is empirically tested for South Africa. Chapter 6 presents the conclusion and policy implications of the thesis.

Chapter 2

2. Explaining regional wage disparities: Theoretical and empirical insights

2.1. Introduction

Several theories have been developed to determine how and why wages differ across space. These include classical economic theory, human capital theory, endogenous growth theory, amenity theory, wage curve theory, and the new economic geography theory. The theories have been accompanied by a significant number of empirical studies exploring the sources and determinants of regional wage disparities. The purpose of this chapter is to provide a theoretical and empirical review of these sources.

The remainder of the chapter is structured as follows: Section 2.2 provides an overview of the theoretical explanations of regional wage disparities. Section 2.3 discusses the key theoretical insights from the new economic geography theory and their implications for understanding regional wage disparities. The aim is to provide an explanation of the workings of the theory as a basis for the empirical work that follows in chapter 5. Section 2.4 presents an overview of the empirical evidence on regional wage disparities. Section 2.5 concludes the chapter.

2.2. Theoretical insights on regional wage disparities

From a theoretical view, the literature has explained regional wage disparities from different perspectives. A starting point is the neoclassical economic theory, which consists of dynamic growth models (Solow, 1956; Swan, 1956) and static trade models - comparative advantage models (Heckscher, 1919; Ohlin, 1933; Samuelson, 1949), and factor movement models (Rybczynski, 1955). Assuming perfect competition, homogeneous products and non-increasing returns to scale, the neoclassical economic theory states that, while regional wage disparities arise in the process of reallocation of resources, diminishing returns to capital, interregional trade and factor mobility promote a decrease in these disparities (Hofer & Wörgötter, 1997). This leads to convergence and subsequently, even distribution of wages across regions[5].

[5] It is the case that, if there is a high wage region and a low wage region, the neoclassical theory suggests that workers will move from the low wage to the high wage region. On one hand, this would increase the supply of labour in the rich region, creating a downward pressure on wages, while on the other hand, it would decrease the supply of labour in the poor region, generating an upward pressure on wages. The opposite dynamic applies to the movement of firms, as firms would move from a high wage region to the low wage region, prompting a decrease in demand for labour, and in turn wages, in the rich region, while increasing labour demand and wages

This prediction, which has grown to be known as the *"convergence hypothesis"* (Quah, 1996; Monastiriotis, 2014), has generated considerable controversy among researchers, particularly in light of the apparent tendency toward divergence between the income and wages of industrialised and less developed nations (Dawkins, 2003). The observed divergence across countries could be attributed to migration regulations, national borders and distance between nations. However, one may not expect the same restrictions to apply within countries and wages are more likely to converge across sub-national regions. Surprisingly, we do not see perfect convergence within countries either. In fact, existing evidence points to persistent differences in wages across regions in many countries (Estanislau, Staduto, & Parré ,2013; Zaman & Goschin, 2014; Huang & Chand, 2015; Chen, Chang, & Su, 2016; Vakulenko, 2016), even those with high factor mobility (Ferens, 2015). As a result, wage differentials are witnessed across, as well as within countries. However, the extent of these differentials differs as wage differences are generally more pronounced across countries than regions within a country. To provide an explanation of these differentials, the neoclassical economic theory was extended to incorporate several perspectives affecting the location decisions of firms who demand labour and workers who supply labour.

The first perspective, the human capital theory, postulates that regional wage disparities are an outcome of demand and supply of human capital (Becker, 1962; Willis, 1986). According to this theory, firms take into account the quality of human capital when choosing a location for production, while workers consider the returns for their skills in different locations. Accordingly, firms choose to migrate to regions with highly skilled workers who are more productive than less skilled workers, to maximise profits. In return for their high productivity, highly skilled workers in these regions receive higher wages, while highly skilled workers in low-wage regions migrate to regions offering higher wages. This leads to regional wage disparities as less skilled, therefore less mobile workers, remain in low-wage regions. The endogenous growth theory reinforces the importance of human capital in explaining regional wage disparities (Romer, 1986; Lucas, 1988).

Another perspective, the wage curve theory, identifies differences in local unemployment as critical in driving regional wage disparities (Blanchflower & Oswald, 1995). This approach postulates that high unemployment leads to lower wages. Thus, the theory predicts a negative

---

in the poor region. These dynamics promotes regional convergence of wages, GDP per capita and consequently living standards (Alexiadis, 2010; Ferens, 2015).

relationship between regional wage and local unemployment rate. The importance of local unemployment in explaining regional wage differentials also finds support from Harris & Todaro (1970). They noted that migration by workers from the rural traditional sector with low wages to the urban modern sector with higher wages promotes regional wage convergence. However, persistent unemployment in the urban sector discourages migration from the rural sector and this tends to limit convergence as wages remain high in urban areas. Although important, the wage curve theory, the human capital theory and the endogenous growth theory have been criticised in the literature for failing to provide insights into why some regions are more prosperous than others, represented by low levels of unemployment and more human capital (de Sousa, 2010).

A further perspective, the amenity theory, highlights the importance of differences in local amenities (Roback, 1982). These amenities can either be natural such as access to waterways, favourable climate, and valuable mineral resources or non-natural such as the cost of living, quality institutions, availability and quality of public services. It is the case that, if workers derive more utility from working in regions with favourable amenities, they will tend to migrate to high-amenity areas, prompting the supply of labour in those regions to increase, which in turn depresses wages. On the other hand, the supply of labour in the regions with unfavourable amenities will decrease and this, in turn, increases wages. Similarly, if firms place a high value on certain regional amenities, they will move to high-amenity areas, leading to an increase in the demand for labour in those areas. This, in turn, increases local wages in these areas. On the other hand, the demand for labour in regions with few amenities will decrease leading to lower wages. Thus, differences in local amenities can sustain regional wage disparities[6].

Taken together, we refer to the theories discussed above as "standard economic theory". These theories suggest that differences in region-specific factors such as human capital, local unemployment, and local amenities can sustain regional wage disparities[7]. This implies that removal of regional factor differences would promote a decrease in regional wage disparities,

[6] It is important to note that we reach this conclusion simply because we look at the supply and demand effects separately. However, wage equalisation could still be possible, despite the contrary predictions for each side of the market.

[7] Apart from these theories, the theories of structural change are also central in explaining regional economic development, which in turn has important implications on regional wage levels. These theories suggest that reallocation of labour from the agricultural sector to the industrial sector is the key for economic growth (Lewis (1954; Chenery, 1960). The resulting structural change extend to other economic functions, such as the change in consumer demand, as well as changes in socioeconomic factors such as urbanisation.

leading to wage convergence. However, as already discussed, evidence suggests that considerable wage differentials exist, even between regions with similar underlying factors. It is also observed that some places continue to thrive after having lost their initial advantages, such as natural resource endowments, that led to their economic prosperity, while others fall behind (Bosker, 2008). Although the standard economic theory is important, it seems to be an inadequate explanation for the spatial concentration of wages and economic activities that one observes across the globe. Therefore, a theory that goes beyond standard regional specific factors is needed to better explain wage differentials across regions.

The New Economic Geography (hereafter, NEG) theory pioneered by Krugman (1991), is a possible candidate, as it suggests that there are relevant economic forces missing from the standard economic theory that can affect regional wage disparities. The theory stresses the importance of the spatial distribution of demand driven by spatial interactions among economic agents in explaining regional wage disparities. The main insight from this theory is that spatial concentration of firms and consumers creates positive externalities (such as low transport costs and greater access to markets), as well as negative externalities (such as higher housing costs and competition) that raise local wages in those regions that offer greater access to markets. Thus, NEG through - access to markets establishes itself as an important theory for explaining differences in wages even across regions with underlying characteristics that are otherwise similar. Its importance in explaining regional economic development has also been recognised in policy circles. For instance, the World Bank 2009 World Development Report with the theme *"Reshaping Economic Geography"* focused on three key elements linked to the NEG theory, namely density, distance and division (World Bank, 2009)[8].

In summary, based on a number of alternative assumptions to those of the standard economic theory, the NEG theory seems to offer solid theoretical explanations for the substantial and persistent spatial concentration of economic activities and wage disparities observed at different geographical levels across the globe. It also provides a reasonable explanation for why wages differ across regions that are otherwise similar, which the standard economic theory

[8] While density looks at the concentration of economic activities in space and relates to the core-periphery economic structure predicted by the NEG theory, distance considers proximity between economic agents where factor mobility and transport costs matter and division points to territorial differences and the need for economic integration to eliminate regional imbalances. Given these three dimensions, the policy challenge is thus getting optimal density by reducing the distance between firms and workers. This can be achieved through economic integration which reduces divisions among economies through initiatives such as infrastructure development which lessen transport costs.

fails to explain. The next section provides a detailed discussion of the key features of this theory.

2.3. Key features of the NEG theory

The idea that access to markets is important for economic development, and therefore factor prices dates back at least to Harris (1954) who argues that a region's attractiveness as a site of production depends on its accessibility to markets. Accessibility is captured by a market potential index, defined as a distance-weighted sum of purchasing power of all other regions:

$$MP_i = \sum_{r=1}^{R} Y_r \, (d_{ir})^{-\tau} \qquad (1)$$

Y_r is the purchasing power (or market size), measured by total GDP, household or personal income for each region, and d_{ir} is the bilateral distance between two points (region i and r) that measures transport costs. The higher the market potential index of a given location, the greater the accessibility to markets, and this, in turn, promotes increased production, and higher wages, and incomes in that location. Yet, Harris (1954) did not provide a theoretical foundation for his measure. NEG theory, pioneered by Krugman (1991) and further developed by Krugman & Venables (1995), Venables (1996), Helpman (1998) and Fujita, Krugman, & Venables (1999) among others, offered a much needed theoretical foundations[9]. It provides micro-foundations to the market potential concept using a general equilibrium model setting characterised by imperfect competition, increasing returns to scale, transport costs and consumers' love for variety (Fujita & Mori, 2005).

The central idea of the NEG theory is that the interaction of transport costs, increasing returns to scale and consumers' love of variety with either labour mobility (Krugman, 1991; Helpman, 1998) or intermediate inputs (Krugman & Venables, 1995) generates demand linkages that determine the spatial distribution of economic activity, income and wages in space. These demand linkages depend on the tension between agglomeration and dispersion forces. Agglomeration forces promote the spatial concentration of economic activities within a given economic space, while dispersion forces promote spreading out.

Acting as an agglomeration force is the "home market effect". According to this, firms producing under increasing returns to scale are attracted to locations with good access to

[9] For a review of the main features of NEG theory and the empirics with regard to agglomeration and trade see Overman et al. (2001), as well as Head & Mayer (2004).

markets where they can sell their goods (backward linkages), as well as get intermediate inputs (forward linkages) at low transport costs (Redding, 2010; Fallah, Partridge, & Olfert, 2011). This leads to higher net revenue for firms, which in turn enables them to pay their workers higher wages. These market access advantages provide an incentive for other firms to relocate to areas with greater access to markets. This provides another agglomeration force, the "price index effect", according to which the new firms lead to an increase in locally manufactured goods, which combine with lower transport costs to drive the local price index down (Baldwin, Forslid, & Martin, 2005). The resulting low-price index leads to an increase in real wages, which in turn stimulates workers to migrate to locations with good access to markets. Thus, the interaction of these two effects is mutually strengthening, encouraging firms and workers to co-locate in locations with greater access to markets, and causing wages to diverge across regions.

However, the resulting co-location of firms and workers has an associated cost that act as dispersion forces, promoting spreading out of economic activity and even distribution of income and wages. These dispersion forces derive from the "crowding out effect", according to which firms and workers are discouraged from locating closer to regions with greater access to markets where competition and costs of non-tradeable goods (such as land and housing services) are higher (Kosfeld & Eckey, 2010).

Overall, the location of firms and workers, and therefore the distribution of economic activities, incomes and wages, depends on the tension between agglomeration and dispersion forces (Fujita, 2007). When agglomeration forces are stronger than dispersion forces, firms and workers tend to concentrate in core locations with greater access to markets, as opposed to periphery ones far from markets. The resulting core-periphery economic structure can be considered to be an unintentional economic outcome where economic activities, incomes, and wages are unevenly distributed across space.

While the NEG literature emphasises the importance of demand linkages, these linkages are driven by different agglomeration and dispersion forces in various NEG models. For example, Krugman (1991) and Helpman (1998) focus on demand linkages where labour mobility in the manufacturing sector acts as a major driving force for agglomeration. However, Krugman (1991) uses immobile agricultural workers and Helpman (1998) non-tradeable housing services as the dispersion force. These two models are most often thought of as being more applicable to exploring agglomeration within countries where labour mobility is greater than across countries where labour is less mobile. Krugman & Venables (1995), Venables (1996) and Fujita et al. (1999) focus on demand linkages, where input-output linkages between firms provide an important force for agglomeration, while immobile labour provides a force for dispersion. These three models are most often thought of as being more relevant for across country agglomeration where labour mobility is constrained by migration regulations, borders and distance.

Despite the differences in agglomeration and dispersion forces, these models have similar dynamics and predictions on the distribution of economic activity, income, and wages. A key theoretical prediction common in all NEG models is that, in equilibrium, nominal wages are higher in locations with greater access to markets (Redding, 2010). A famous representation of this prediction, set out in Fujita et al. (1999), is the wage equation given by:

$$w_i = \left[\sum_r Y_r \, P_r^{\sigma-1} T_{ir}^{1-\sigma}\right]^{\frac{1}{\sigma}} \tag{2}$$

where w_i is nominal wage for region i. The right-hand side of (2) is market potential (MP_i), an index measuring the degree of accessibility to markets for each region. Thus, the market potential of region i is a function of the sum of incomes, Y_r, in all locations, weighted by trade costs between region i and r, T_{ir} and deflated by the composite price index for manufactured varieties, P_r. $\sigma > 1$ is the elasticity of substitution between manufactured varieties. Trade costs increase with increasing distance between trading regions, while the composite price index increases with the increasing fraction of manufactured varieties that are imported. To reflect the spatial interaction of workers supplying labour and firms demanding labour, as well as supplying intermediate inputs, the market potential index can be decomposed into market access, $MA_i = \sum_r Y_r \, P_r^{\sigma-1} T_{ir}^{1-\sigma}$ and supplier access, $SA_i = \sum_r n_r \, (p_r T_{ir})^{1-\sigma}$ indices (Redding & Venables, 2004; Redding, 2005). The first index measures access to potential consumers, while the second captures access to intermediate inputs.

Equation (2), thus reflects labour demand by firms in region i supplying intermediate inputs and final goods. It gives the maximum wage that a manufacturing firm in region i can afford to pay its workers, given its level of market potential (level of market and supplier access). An increase in market potential in a given region leads to firms being able to serve a larger market, thereby stimulating a rise in demand for labour, which in turn drives up wages in that region. Thus, differences in market potential can be a force for wage divergence across regions, even in the absence of significant differences in regional specific factors.

2.3.1. Measuring market potential

The above-mentioned theoretical prediction given by equation (2) has been accompanied by a significant number of empirical studies testing the validity of the prediction in various countries (Brakman, Garretsen, & Schramm, 2004; Mion, 2004; Hanson, 2005; Ottaviano & Pinelli, 2006; Fallah, Partridge, & Olfert, 2011; Paredes & Iturra, 2012)[10]. While the bulk of the studies provide evidence in support of the prediction of the wage equation, a key concern in this literature is how to measure market potential. This is mainly because some of the elements of the market potential index, such as the price index, are not observable (Kiso, 2005).

To measure market potential, three strategies have been utilised, leading to three strands in the literature (see Brakman et al. 2009, pg 211 for great details on these strategies). The first strategy is based on the Harris (1954) market potential index, which is obtained from equation (2) by assuming that regions have a similar price index (P_r=1). This strategy has been used by Ottaviano & Pinelli (2006), Paluzie, Pons, & Tirado (2009); Fallah et al. (2011), Martínez-Galarraga, Tirado, & González-Val (2015), among others. Although popular in empirical circles, the Harris market potential function has generally been criticised for its “ad hoc” nature and lack of theoretical foundations.

Two approaches based on structural wage equations derived explicitly from NEG theory have also been proposed by Redding & Venables (2004) and Hanson (1998) to improve this index. An appealing feature of these two approaches is that the estimation of the wage equation provides estimates of key structural parameters of the underlying NEG model. Apart from confirming market potential as a key determinant of regional wages, estimation of these parameters is of great importance, as they reveal the precise channels through which market potential influences regional wages. Furthermore, these parameters can also be used to check the consistency of observed results with the underlying theoretical framework.

The first approach by Redding & Venables (2004) builds on the theoretical framework of Fujita et al. (1999), which assumes the interaction of immobile workers and firms producing intermediate inputs and final goods. In their approach, Redding & Venables (2004) estimate market potential in two steps. Using bilateral trade data, they estimate a gravity trade model,

[10] Another strand of the NEG empirical literature explores the influence of access to markets on the location choice of workers (Crozet, 2004; Pons, Paluzie, Silvestre, & Tirado., 2007), while a further strand focuses on the location choice of firms (Head & Mayer, 2004).

after which the estimated model parameters are manipulated to derive a theory-based market potential index that can be decomposed into market access and supplier access indices. Based on their index, wages (proxied by GDP per capita) as given by equation (2) are positively related to market and supplier access, which are a function of transport cost-weighted income and price index of all locations. While this approach has been applied across regions within countries (Amiti & Cameron, 2007; Hering & Poncet, 2009; 2010; Fally, Paillacar, & Terra, 2010), its underlying assumption of labour immobility is best applicable to cross-country analysis, where migration is impeded by borders, distance, and language. The drawback of this approach is that it requires large amounts of bilateral trade flow data, which is not readily available at the regional level in most developing countries. This approach has also been criticised for its complexity and use of a number of arbitrary assumptions in constructing the indices (see Bosker & Garretsen, 2010)[11].

Hanson (1998) proposes an alternative approach for measuring market potential which is better suited for within-country analyses. His approach draws on the theoretical framework of Helpman (1998), which assumes the interplay between labour mobility and non-tradeable housing services. In his approach, Hanson (1998) derives a directly testable structural wage equation by replacing the unobservable price index variable in equation (2) with observable housing data. The derived theory-based market potential index is a function of income, housing stocks, and wages in each region, weighted by transport cost. While Hanson's (1998) approach is more direct and based on less restrictive assumptions than that of Redding & Venables (2004), its major limitation is its failure to account for intermediate inputs as it does not include a supplier access index.

2.4. Overview of empirical evidence on regional wage disparities

Despite the shortcomings of neoclassical and standard economic theories together with the NEG theory, these theories provide a unique contribution to the understanding of how and why wages differ across regions in many countries. Put together, these theories provide a comprehensive framework for studying regional wage disparities, and over the years a large body of empirical research with three main strands of literature has accumulated for different countries.

[11] For example, under the approach by Redding & Venables (2004) trade cost parameters are estimated separately from the market access and supplier access coefficients. By adding additional variables to the trade costs' function, the approach also introduces additional nonlinearity making it increasingly difficult to estimate all parameters of interest.

The first strand of this research examines the spatial patterns that characterise the distribution of wages across regions (Longhi, Nijkamp, & Poot, 2006; Patacchini & Rice, 2007; Huang & Chand, 2015; Breau & Saillant, 2016). This literature finds evidence of substantial wage disparities across regions in many countries, characterised by significant tendencies of wages to spatially concentrate in specific locations. The second strand of the literature tests the neoclassical convergence hypothesis to see whether regional wages are converging or diverging over time (Rosés & Sánchez-Alonso, 2004; Maza & Villaverde, 2006; Zaman & Goschin, 2014; Ferens, 2015; Goschin, 2015). Evidence from this literature is highly mixed and inconclusive, influenced in part by the underlying measures of convergence used, statistical methods applied, the country and the time-period of the analysis. The final strand of the literature examines the causes of regional wage disparities and this literature can be divided into two groups.

The first group investigates the causes of regional wage disparities using reduced-form models and finds evidence of the importance of differences in human capital (Combes, Duranton, & Gobillon, 2008; Acemoglu & Dell, 2010), local unemployment (Longhi, Nijkamp, & Poot, 2006; Ramos, Nicodemo, & Sanromá, 2015; Von Fintel, 2017), local amenities (Graves, Arthur, & Sexton, 1999; Deller, 2009; Partridge, Rickman, Ali, & Olfert, 2010), international trade (Hanson, 2003; Breau & Rigby, 2009) and market potential (Ottaviano & Pinelli, 2006; Fallah et al. 2011; Paredes & Iturra, 2012). While this literature provides evidence in support of both standard economic theory and NEG theory, reduced-form models do not discriminate between theories, as the significance of a given variable may be consistent with many theories (Brakman et al. 2009, pg 199).

To address this shortcoming, the second group explicitly tests a key theoretical insight derived from the NEG theory given by equation (2). Using the approach by Hanson (1998) and Redding & Venables (2004) a number of these studies provide overwhelming evidence in support of the NEG theory in explaining regional wage disparities in many developed countries (Mion, 2004; Brakman et al. 2004; Hanson, 2005; Pires, 2006; Knaap, 2006; Kosfeld & Eckey, 2010). However, in the case of developing countries, studies are not only limited, but the evidence is also mixed, with some studies finding evidence in support of the NEG theory (Amiti & Cameron, 2007; Hering & Poncet, 2009; 2010; Moreno-Monroy, 2008; 2011; Fally, Paillacar, & Terra, 2010), while others do not find evidence in support of the theory (Alvarado & Atienza, 2014; Paredes, 2015).

Taken together, these three strands of the literature make valuable contributions to a better understanding of the existence and persistence of regional wage disparities in many countries.

2.5. Conclusion

This chapter provides a theoretical and empirical review of regional wage disparities to better understand the distribution, convergence dynamics and causes of regional wage disparities. The aim is to provide a basis from which to derive testable hypotheses to inform the empirical chapters of this thesis. The chapter starts with a general discussion of different theoretical perspectives, including neoclassical economic theory and human capital theory, endogenous growth theory, wage curve theory, amenity theory, which we refer to as *"standard economic theory"* and NEG theory. Furthermore, the chapter provides a detailed discussion of the key theoretical insights from NEG theory, paying attention to the NEG wage equation and empirical measurement of market potential.

The chapter also presents a brief overview of the related empirical literature on regional wage disparities. The review shows that existing research appears unanimous in its conclusion that wages differ significantly across regions in many countries. However, findings from this research are inconclusive on what has happened to regional wage disparities over time. The mixed evidence is due, in part, to different measures of convergence used, country, and time-period of analysis. The review also shows that existing research finds differences in human capital, local unemployment, local amenities, international trade and access to markets to be the key determinants of regional wage disparities. This research, however, has not yet identified explicitly which of these factors best explain regional wage disparities in many countries.

Based on this review, we find that research on regional wage disparities has mainly focused attention on developed countries. In contrast, very little research on this area has been conducted in developing countries, particularly in Africa. This is mainly because of the scarcity of reliable and comparable regional economic data in most African countries (Kim, 2008). Lack of regional wage disparity studies in Africa, together with the inconclusive evidence on the changes and causes of regional wage disparities, provide substantial scope for further research in this area.

In the remainder of the thesis, we contribute to the empirical literature on regional wage disparities using the case of South Africa. Chapter 3 examines the spatial patterns that characterise the distribution of wages across regions in South Africa. Chapter 4 investigates the convergence dynamics of wages across these regions. Chapter 5 tests a key theoretical insight derived from NEG theory, to find out whether the NEG theory explains regional wage disparities in South Africa.

Chapter 3

3. The spatial distribution of wages across regions in South Africa

3. 1. Introduction

There is a growing body of empirical evidence on the spatial distribution of wages across regions in many countries (among them Patacchini & Rice, 2007; Monastiriotis, 2009; Huang & Chand, 2015; Breau, 2015; Breau & Saillant, 2016). This research has been used to differentiate between predictions of various economic theories, which in turn provide initial insights into the potential causes of regional wage disparities. It has further been used to provide guidance in the formulation of regional policy aimed at promoting regional economic development and addressing regional wage disparities. While this research provides important insights, most of the existing research has so far focused attention on advanced economies. In contrast, very little research in this area has been conducted in developing countries, particularly in Africa. The problem of scarcity of adequate wage data at the regional level in many African countries has limited such research.

As a result, little is known about the spatial distribution of wages across regions in many African economies. However, such analysis is important to provide new hypotheses and perspectives on the forces behind persistently large regional economic disparities in many African countries (see Bosker & Krugell, 2008; African Economic Outlook, 2015; Shimeles & Nabassaga, 2015). In this chapter, we circumvent the data limitations by constructing a highly disaggregated regional database using South Africa's full population census for the years 1996, 2001 and 2011. This dataset, which is geographically consistent over time across 354 magisterial districts (hereinafter regions), is then used to present the first empirical evidence on the spatial distribution of wages across regions in South Africa. To achieve this, the analysis is structured around the following specific objectives:

- To document the key stylised facts that characterise the spatial distribution of wages across regions in South Africa.
- To assess the consistency of the observed stylised facts with predictions from alternative economic theories.

The remainder of the chapter is structured as follows: Section 3.2 provides an overview of the South African spatial economy. Section 3.3 presents an overview of the related empirical literature. Section 3.4 presents the empirical framework used to document the key stylised facts in the data. The data and its construction is discussed in section 3.5. Section 3.6 presents the empirical findings. Section 3.7 concludes the chapter.

3.2. The South African spatial economy

To have a complete picture of the spatial distribution of wages across regions in South Africa, it is important to highlight the factors that have played an important role in shaping economic development in the country, leading to unequal distribution of economic activity across regions. Among them are pre-existing geographic factors, and apartheid-era and post-apartheid-era institutional policies.

The unequal development of South Africa's landscape was influenced by pre-existing geographic differences in access to waterways, climates, and natural resources. For instance, development of the port cities of Cape Town and Durban was driven by their close proximity to waterways, which gave them an important role in the country as trading posts on the shipping route between Western Europe and Asia (Gelb, 2004; Bosker & Krugell, 2008). The trade enhanced development was reinforced by favourable climatic conditions that encouraged settlement of European colonisers who established institutions to suit their stay in these port cities (Krugell & Naude, 2003)[12].

The discovery of minerals shifted attention to the development of inland regions, where mining activities stimulated the development of Gauteng (Johannesburg and Pretoria). Mining activities and resulting infrastructure established to support the mining industry generated strong economic forces that promoted rapid industrialisation, urbanization and massive migration of workers to support the growing industry in and around Gauteng (Turok, 2012). This allowed Gauteng to develop into the large urban agglomeration that we see today (Beavon, 2001) that accounts for the bulk of the country's economic activities (Stats SA, 2014).

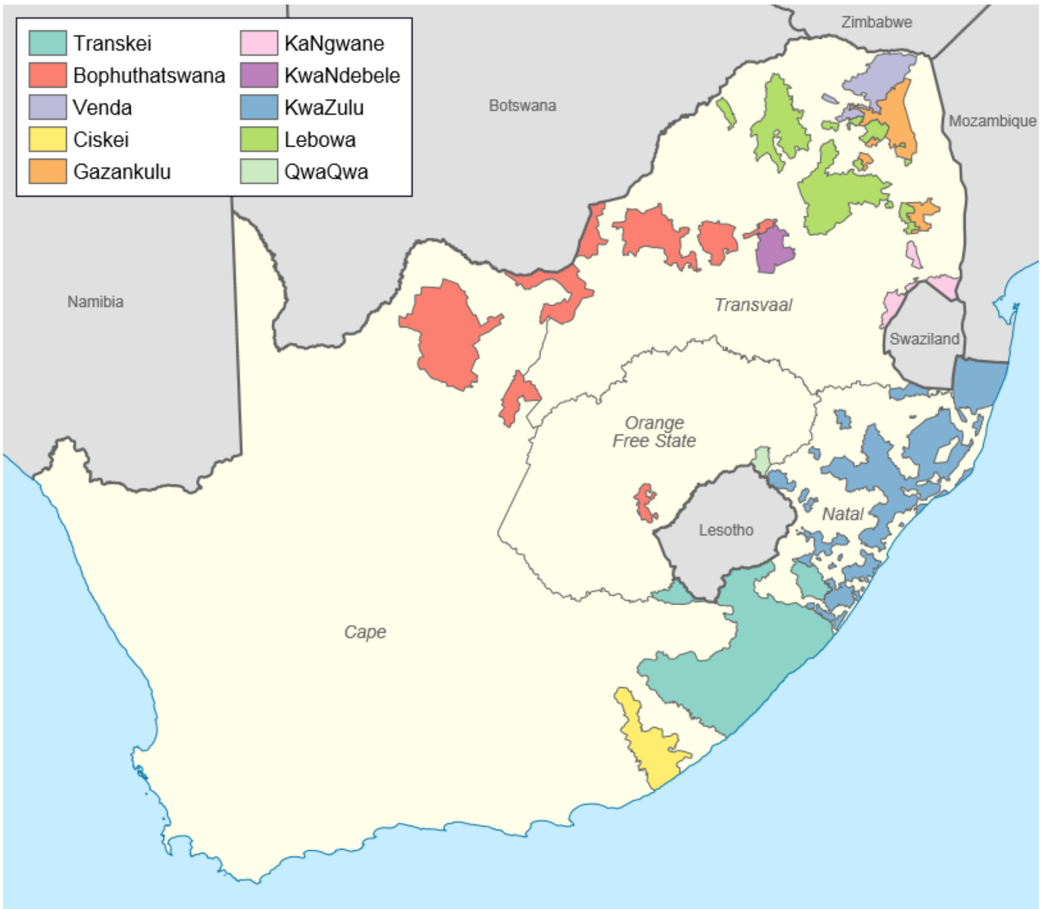
This unequal regional economic development process was further consolidated by apartheid-era institutional policies that appropriated land, wealth, and economic opportunities from blacks for the benefit of a minority white community. Among these policies, the homelands policy dispossessed blacks from their land, forcing them into 10 homeland areas[13]. As

[12] This outcome is supported by Acemoglu, Johnson, & Robinson. (2002), who note that colonies with favourable climatic conditions for European settlements are richer than other countries or regions.

[13] The ten homeland areas, which account for about 13% of South Africa's total land area, are Transkei, Bophuthatswana, Ciskei, Venda, Gazankulu, KaNgwane, KwaNdebele, KwaZulu, Lebowa, and QwaQwa are listed in Figure 3.1.

displayed in Figure 3.1, a key feature of these homelands is their location in periphery areas. They thus suffered problems associated with geographical remoteness including being far from major road networks, international airports, harbours, and major markets. The remoteness of the homelands acted as a barrier to investment, thereby limiting employment opportunities. It also enabled the apartheid government to easily implement policies of separate development and enforce movement restrictions from homeland areas through the gate-pass system. In addition, the Group Areas Act was also implemented and led to the creation of black townships such as Soweto and other informal settlements on the periphery of major urban areas[14].

Figure 3.1: Location of former homelands areas.



Source: http://commons.wikimedia.org/wiki/File:Bantustans_in_South_Africa.svg.

These settlement policies were further reinforced with employment and education laws such as the Bantu Labour Act of 1953, the Bantu Education Act 1954 and the Industrial Conciliation Act of 1956, which favoured the progression of whites in labour markets at the expense of

[14] Townships were provided with minimal infrastructure on the grounds that blacks were temporary residents in "white" South Africa and would return to the homeland areas once their services were no longer required. Inner city regions where whites lived we well serviced with good municipal infrastructure, lucrative commercial activities and employment.

blacks. The apartheid-era policies ushered in a phase of racial segregation and fragmented growth that intensified the unequal regional economic development initiated by pre-existing geographic factors.

The end of the apartheid-era rule in 1994 did not only end years of racial discrimination that heavily disadvantaged the black majority in terms of social and economic opportunities, it also led to three significant changes to the country's spatial economy. The international sanctions that were imposed on South Africa in response to apartheid-era rule were lifted, leading to the integration of the country with the global economy. The integration process was further enhanced by significant trade liberalisation, which roughly coincided with the period in which the country gained its independence (Behar & Edwards, 2006). Furthermore, the post-apartheid government implemented various regional policy initiatives aimed at promoting regional economic development and addressing regional economic disparities. Some of the most prominent initiatives include the National Spatial Development Framework (NSDF) of 1995, Spatial Development Initiatives (SDIs) of 1996, the Regional Industrial Development Strategy (DTI, 2006), and the National Spatial Development Perspective (NSDP, 2003; 2006). In addition, the government embarked on a massive re-demarcation exercise of the country's administrative boundaries in order to dissolve the racially based spatial layout created by the apartheid-era system. This led to the creation of local authorities constitutionally responsible for the development of their areas (Bosker & Krugell, 2008).

With the removal of international sanctions, significant trade liberalisation and implementation of regional policy interventions, one would expect South Africa to take a different growth path and reach a new spatial equilibrium. However, existing evidence suggests that, despite political transition and the demise of formal apartheid planning, no major shifts have taken place in the South African spatial landscape (Nel & Rogerson, 2009). Consequently, many of the spatial patterns from the past persist, with areas in former homelands remaining underdeveloped, with high unemployment rates, poverty and deprivation (Noble & Wright, 2013; Noble, Zembe, & Wright, 2014; Frame, De Lannoy, Koka & Leibbrandt., 2016). Furthermore, economic activities remain highly concentrated in six major urban cities, namely Johannesburg, Ekurhuleni (formerly the East Rand), Durban, Cape Town, Pretoria and Port Elizabeth (Naudé & Krugell, 2005). There are few economic activities in rural areas of South Africa.

In conclusion, this brief overview shows a multiplicity of factors, including geographic factors, apartheid-era and post-apartheid-era institutional policies that have contributed in shaping the unequal distribution of economic activity, wealth, income, and wages in South Africa. The next section provides a discussion of empirical insights on the spatial distribution of wages across regions.

3.3. Related empirical literature

This section provides a brief overview of international studies that examines the spatial distribution of wages across regions within a country, as well as related studies in South Africa.

Related empirical literature: International studies.

In the recent years, the availability of geographically referenced data has offered insights into the nature and extent of spatial patterns that characterise the distribution of various economic outcomes. A small but growing number of studies have provided comprehensive empirical findings on the nature and extent of the spatial distribution of wages across regions within countries[15]. These include Breau & Saillant (2016) for Canada, Patacchini & Rice (2007) for Great Britain, Monastiriotis (2009) and Larraz, Navarrete, & Pavía. (2016) for Spain, Amaral, Lemos, Simões, & Chein (2010) for Brazil, as well as Moreno-Monroy (2011) and Huang & Chand (2015) for China[16]. We discuss the key findings from these studies.

A key finding from the literature is that there are significant disparities in the distribution of wages across regions in many countries. The empirical evidence shows that, rather than being randomly distributed, these disparities show high spatial concentration (both positive and negative) of wages in space. The literature largely finds evidence of positive spatial concentration, where regions with high (low) wages are located closer to other regions with high (low) wages (Patacchini & Rice, 2007; Huang & Chand, 2015; Breau & Saillant, 2016). Thus, in many countries, adjacent regions tend on average to exhibit similar levels of wages. This finding is common for both developed (Patacchini & Rice, 2007; Monastiriotis, 2009; Larraz et al. 2016; Breau & Saillant, 2016) and developing countries (Amaral et al. 2010; Moreno-Monroy 2011; Huang & Chand, 2015; Mazol, 2016). However, in some cases, some studies also find evidence of negative spatial concentration, where regions with high (low)

[15] A large body of research has also looked into the nature and extent of spatial patterns of GDP per capita (Gallo & Ertur, 2003; Khomiakova, 2008), human capital, public investment, GDP per capita and its growth (Celebioglu & Dall'erba, 2010), income inequality (Ezcurra et al. 2007), income per capita and inequality (Rodríguez-Pose & Tselios, 2011), employment and quality of life (Rusche, 2010), as well as GDP per capita, unemployment and other economic outcomes (Monastiriotis, 2009).

[16] Other studies provide insights into the spatial distribution of value-added per hour worked (Patacchini, 2008), and value added per worker (Ezcurra et al. 2008).

levels of wages are located in close proximity to regions with low (high) levels of wages (Huang & Chand, 2015). This suggests that neighbouring regions can also register dissimilar levels of wages. As suggested by Breau & Saillant (2016), regional specific factors such as minerals, oil, gas and agricultural land are key drivers of negative spatial autocorrelation.

Another important finding in the literature is that, despite evidence of the strong spatial concentration of wages in many countries, the observed spatial patterns are highly localised with only a few regions showing evidence of significant positive and negative spatial concentration (Huang & Chand, 2015; Breau & Saillant, 2016). Furthermore, the extent and magnitude of the observed spatial concentration vary significantly across countries, driven in part by the spatial weight matrix used and the number of geographical units under study. For example, Huang & Chand (2015) use the queen contiguity spatial weight matrix and find spatial autocorrelation (concentration) figures of between 0.214 and 0.231 across 31 Chinese provinces between 2001 and 2010. Looking at 287 Canadian regions over the period 1996 and 2006, Breau & Saillant (2016) employ the queen contiguity and K-10 nearest neighbour spatial weight matrices. Their results reveal spatial autocorrelation figures of between 0.526 and 0.560 (queen contiguity), as well as 0.497 and 0.557 (K-10 nearest neighbour). However, these studies find evidence of persistently increasing spatial autocorrelation of wages of over time.

Comparing these empirical features to theoretical predictions, the literature shows that neoclassical economic theory cannot explain the observed spatial patterns of wages in space. Rather, the observed spatial patterns seem to correspond with NEG theory and standard economic theory. However, the literature clearly shows that neither NEG theory nor standard economic theory fully explains the empirical features found in the data. For instance, the revealed positive spatial concentration pattern of wages show large and well-defined patterns, confirming the existence of a dualistic economic structure with a strong rural-urban division. This economic structure is consistent with the core-periphery economic structure predicted by NEG theory (Krugman, 1991)[17]. Thus, evidence of positive spatial concentration is used as an indication of the importance of economic forces emphasised by NEG theory. On the other hand, the revealed negative spatial concentration of wages shows small and less-defined patterns. The literature acknowledges that the negative spatial concentration patterns are rare and difficult to explain (Khomiakova, 2007), as they can support both standard economic and NEG

[17] The rural areas correspond with periphery regions characterised by little economic activity, and low income and wages, while the urban areas correspond with core regions with more economic activity, and higher income and wages.

theories. However, some studies find that negative spatial concentration patterns are isolated and correspond with areas where mining, oil, gas, fertilizers and agricultural sectors are the primary industrial activities (Breau & Saillant, 2016; Mazol, 2016). This suggests the importance of region-specific factors, as postulated by standard economic theories. Nevertheless, the literature finds that the patterns are dominated by positive as opposed to negative spatial concentration. Based on this evidence, we can conclude that NEG theory offers a more plausible explanation of the observed empirical characteristics found in the data.

We can summarise this literature in two key points. Firstly, wages vary significantly across regions in many countries. The observed disparities are characterised by significant spatial concentration, showing evidence of both positive and negative spatial patterns. Secondly, the observed spatial concentration patterns cannot be attributed to a single economic theory, as the empirical features found in the data are consistent with predictions from both NEG and standard economic theories.

Related empirical literature in South Africa

In South Africa, a number of studies have documented the spatial distribution of various economic outcomes such as crime (Breetzke, 2008; Hiropoulos & Porter, 2014), dissatisfaction with the performance of local government (Cheruiyot, Wray, & Katumba, 2015), unemployment (Weir-Smith & Ahmed, 2013), socioeconomic inequality (Hakizimana & Geyer, 2014), and GDP per capita (Bosker & Krugell, 2008).

Weir-Smith & Ahmed (2013) find that, between 1991 and 2007, unemployment rates across municipalities showed evidence of significant positive spatial autocorrelation (concentration). Their results show that positive spatial autocorrelation increased over time from 0.61 in 1991 to 0.71 in 2007. The evidence of positive spatial autocorrelation suggests the spatial concentration of municipalities with similar unemployment rates. While Weir-Smith & Ahmed (2013) find evidence of increasing spatial concentration of municipalities with similar levels of unemployment, their study did not go further to reveal the geographical location of the observed positive spatial concentration. They also do not show whether the observed positive spatial autocorrelation masks patterns of negative spatial concentration.

Using a number of socioeconomic inequality indices[18] derived from 2011 census data for a sample of 14030 main places[19], Hakizimana & Geyer (2014) find that the South African spatial economy is characterised by significant spatial disparities in the distribution of the bulk of the indices. Hakizimana & Geyer (2014) find that the disparities show evidence of positive and negative spatial autocorrelation defined by four spatial cluster regimes (H-H and L-L, as well as H-L and L-H), with positive spatial autocorrelation the dominant pattern. In line with the international literature, and despite evidence of significant positive and negative spatial autocorrelation, Hakizimana & Geyer's (2014) results also reveal evidence of highly localised spatial autocorrelation, with a large number of main places registered insignificant local spatial autocorrelation.

In another study, Bosker & Krugell (2008) explore the spatial distribution of GDP per capita across a sample of 354 magisterial districts over the period 1996 and 2004. Their results show that GDP per capita varies substantially across regions, and its distribution is characterised by both positive and negative spatial autocorrelation. Cheruiyot, Wray, & Katumba (2015) carry out a spatial analysis of dissatisfaction with the performance of local government in the Gauteng City-Region in 2013. Their analysis reveals spatial clustering in levels of dissatisfaction with the performance of local government.

Taken together, these studies find evidence of significant disparities in the distribution of various economic outcomes in South Africa. The disparities show evidence of both positive and negative spatial autocorrelation. Building on this research, our study contributes to the empirical literature by documenting the key stylised facts that characterise the spatial distribution of wages across regions in South Africa. A further contribution is to reveal the specific geographical locations where wages concentrate. Finally, it assesses the consistency of the observed facts with predictions from alternative economic theories.

3.4. Empirical framework: Exploratory spatial data analysis (ESDA)

This chapter documents the key stylised facts that characterise the spatial distribution of wages across regions in South Africa. The analysis is guided by two key questions. The first is, what is the relationship between wages in one region and those of its neighbours? The second is, are wages across regions spatially concentrated in specific geographical locations? These questions

[18] These indices include the multidimensional composite index of deprivation; range ratio; relative mean deviation; standard deviation of logarithms; Gini coefficient; Kuznets ratio; Theil inequality index, mean logarithmic deviation, and the Atkinson index.
[19] Within the census, main places refer to cities or towns.

are addressed using exploratory spatial data analysis (ESDA) techniques (see Anselin, 1988, 1998; Haining, 1990 for details on the techniques). The techniques include choropleth maps, box plots, scatter plots, and measures of spatial autocorrelation (Celebioglu & Dall'erba, 2010). These allow description, visualisation and formal testing of the nature and extent of spatial patterns that characterise the distribution of various economic outcomes (Patacchini & Rice, 2007). Thus, they allow us to characterise the relationship between wages in one region and those of its neighbours. The most appealing feature of ESDA techniques is that they allow for identification of the key properties of spatially referenced data based on the spatial aspects of the data, without predetermined ideas or hypotheses.

Central to ESDA techniques is the concept of spatial autocorrelation (or concentration), which can be defined as the spatial correlation of a variable with itself in space (Altay & Çelebioğlu, 2012)[20]. This suggests the existence of a functional relationship between what happens in one region and what happens in neighbouring regions (Anselin, 1988), which in turn depends on each region's location relative to other regions. This functional relationship is an outcome of the spatial interactions between economic agents in space driven by interregional trade, factor mobility, knowledge exchange, thick labour markets, and proximity to demand and supply markets (Patacchini & Rice, 2007). The resulting relationship can lead to either positive spatial autocorrelation, which exists when high (low) values correlate with high (low) neighbouring values. Or it can lead to negative spatial autocorrelation, which exists when high (low) values correlate with low (high) neighbouring values in space. Thus, positive spatial autocorrelation exists when regions with similar values of a random economic outcome tend to locate near each other. Negative spatial autocorrelation exists when regions with dissimilar values of a random economic outcome have a high propensity to concentrate in space.

To empirically measure the nature and extent of spatial autocorrelation, the spatial arrangement of geographically referenced data need to be considered. In exploratory data analysis, the spatial arrangement of the data is captured by an $n\ x\ n$ spatial weight matrix (W), which reveals spatial units, and how they are related and influence each other. The analysis discussed in this chapter uses a distance-based spatial weight matrix, where the spatial arrangement of the data

[20] Anselin (1988) defined spatial autocorrelation as the coincidence of value similarity with locational similarity. Over the years, a number of terms such as spatial dependence, spatial association spatial relation, spatial concentration, spatial agglomeration and spatial interactions, have been used interchangeably with spatial autocorrelation. We take the same approach in this chapter and use the most appropriate term for what is being discussed at a given point.

is captured by an inverse distance function with a distance cut off threshold. This matrix allows for spatial relations among regions to decrease with increasing distance between them up to a given point, above which spatial relations between regions are assumed negligible. The matrix is defined as follows:

$$W = \begin{cases} w_{ir} = 0 \; if \; i = r \\ w_{ir} = {d_{ir}}^{-1} \; if \; d_{ir} \leq D_{max} \\ w_{ir} = 0 \; if \; d_{ir} > D_{max} \end{cases} \text{ and } w_{ir}^{*} = w_{ir} \,/\, \sum_{r=1}^{n} w_{ir} \tag{1}$$

where W is an $n \; x \; n$ spatial weight matrix, whose element w_{ir} captures the degree of spatial relation between values of a random economic outcome (in our case, wages) at one point in space and its values in other spatial units of the study area. d_{ir} is the great circle distance (in kilometres) between the geographical centres (centroids latitude and longitude) of region i and region r. D_{max} (=205 km) is the critical distance cut-off, which we set at 205 km. Our motive for using 205km as the cut off distance is that it ensures that each region has at least one neighbour[21]. w_{ir}^{*} is a standardised element of the spatial weight matrix, which ensures that relative, rather than absolute distance is considered. This transformation ensures easy use, computation and interpretation of spatial autocorrelation results. A detailed explanation of the construction of the distance-based spatial weight matrix, as well as other weight matrices used for robustness checks is provided in Appendix 3.2.

Having defined the spatial weight matrix, to investigate the nature and extent of the spatial patterns in the data, the study employs global and local spatial autocorrelation measures based on Moran's I statistic (Moran, 1948)[22]. We provide a brief explanation of these measures in the next subsection.

Global spatial autocorrelation measure.

Global spatial autocorrelation refers to the overall degree of spatial correlation between the value of a random economic outcome in each region and the values of the same economic

[21] To know the distance that ensures that each region has at least a neighbour we use the Stata command "nearstat" which generates the geographical distance between two points, as well as distance to the nearest neighbour.

[22] Global spatial autocorrelation can also be measured using Geary's c statistic (Geary, 1954) and the Getis & Ord statistic (Getis & Ord, 1992; Ord & Getis, 1995). However, Moran's I statistic is appealing because of its ability to be decomposed from a global to a local measure, which provides more information about the nature of the observed spatial patterns in the data. Further, compared to the Getis-Ord statistic, which only reveals evidence of positive spatial autocorrelation, the local Moran's I statistic provides information on negative spatial autocorrelation.

outcome in neighbouring regions. To investigate the global properties of the wage data, we use global Moran's I statistic, which captures the extent of overall spatial concentration that exists in a dataset. Formally, the statistic measures the degree of linear association between the value of an economic outcome at one location (y_i) and the spatially weighted average of the neighbouring values of the same outcome (Wy_r) (Anselin et al. 2007). The statistic is formulated as follows:

$$I = \frac{n \sum_{i=1}^{n} \sum_{r=1}^{n} w_{ir}^{*} (y_i - \overline{y})(y_r - \overline{y})}{\sum_{i=1}^{n} \sum_{r=1}^{n} w_{ir}^{*} \sum_{i} (y_i - \overline{y})^2} \tag{2}$$

where n is the total number of regions indexed by i and r. y_i and y_r is the wage value for region i and r, while $\overline{y}$ is the national wage. w_{ir}^{*} is the normalised spatial weight element, capturing the degree of spatial relations between region i and region r. Values of I ranges from -1 corresponding to perfect negative spatial correlation to +1 corresponding to perfect positive spatial correlation and 0 confirming no spatial correlation.

Thus, under the assumption of spatial randomness "no global spatial autocorrelation" ($I = 0$), we evaluate the presence of spatial autocorrelation by comparing the calculated global Moran's I statistic (I) and its theoretical mean, given by $E(I) = -1/(n-1)$. When the spatial randomness assumption is violated, the value of I is larger (smaller) that $E(I)$, which indicates positive (negative) spatial autocorrelation. To arrive at these decisions, the inference of global Moran's I statistic and hypothesis testing are based on z-scores that follow a normal distribution (Anselin, 1992)[23].

While informative, global spatial autocorrelation only detects the presence of overall spatial autocorrelation of the entire sample and does not provide information on the locations where wage values concentrate. In addition, it does not allow us to identify whether the observed positive spatial autocorrelation is due to the spatial concentration of regions with high or low-wage values. Finally, it fails to reveal whether they are spatial concentration of regions working against observed overall positive spatial autocorrelation. To overcome these shortcomings, we turn to local spatial autocorrelation measures.

Local spatial autocorrelation.

Local spatial autocorrelation refers to spatial concentrations around individual locations and pinpoints regions that contribute significantly to (or work against) overall global spatial

[23] Statistically, significant positive z-values confirm evidence of positive spatial autocorrelation, whereas negative z-values imply negative spatial autocorrelation.

autocorrelation. We measure local spatial autocorrelation using two measures, the Moran scatter plot (Anselin, 1997) and the local Moran's I statistic (Anselin, 1995).

The Moran scatterplot allows for the interpretation of global Moran's I statistic using a visual display of the linear association of Wy_r against y_i (Anselin, Sridharan, & Gholston., 2007). An appealing feature of the scatterplot is its ability to show both global and local spatial autocorrelation. While global spatial autocorrelation is reflected by the slope coefficient of the linear regression of Wy_r against y_i, local spatial autocorrelation is confirmed by four quadrants that correspond to four different patterns of local spatial autocorrelation between the wage value of each region and those of its neighbours. These quadrants are defined as follows: (1) the upper right quadrant, high-high (H-H) represents regions with high wage values that are surrounded by other regions with high wage values, (2) the upper left quadrant, low-high (L-H) captures regions with low wage values that are neighbours to regions with high wage values, (3) the lower left quadrant, low-low (L-L) displays locations with low wage values that are surrounded by other regions with low wage values, while (4) the lower right quadrant, high-low (H-L) gives regions with high wage values that are neighbours to regions with low wage values. Whereas the H-H and L-L quadrants capture positive spatial autocorrelation, L-H and H-L quadrants capture negative spatial autocorrelation. Taken together, these quadrants point to a spatially heterogeneous economy.

To reveal each region showing significant local spatial autocorrelation in each quadrant, we complement the Moran scatter plot with a local indicator of spatial association (LISA), the local Moran's I statistic (Anselin, 1995)[24]. The statistic captures the presence of significant local spatial autocorrelation (high–high or low–low) or spatial outliers (high–high or low–low) for each location (Celebioglu & Dall'erba, 2010). The local Moran's I statistic has the same properties as the global Moran's I statistic, and, their interpretations are therefore the same[25].

Taken together, the global Moran's I statistic, the Moran scatter plot and the local Moran's I statistic enable us to formally examine the nature and extent of spatial patterns that characterise the distribution of wages across regions in South Africa.

[24] In order for the local Moran's I statistic to be classified as a LISA, the following two conditions need to be satisfied: Given each observation, the LISA should give an indication on significant clustering of similar values around that observation and second the sum of the LISA for all the observations should be proportional to the global indicator of spatial association (Anselin, 1995).

[25] The local Moran's I statistic is derived from the decomposition of the global Moran's I and statistic is given by: $I_{it} = \frac{n(y_{it} - \bar{y}_t)}{\sum_{i=1}^{n}(y_{it} - \bar{y}_t)^2} \sum_{r \epsilon R_i} w_{ir} * \left(y_{rt} - \bar{y}_t\right)$, where all the other elements are defined as before and R_i denotes the set of neighbouring regions of region i.

3.5. Description and construction of the data

The assessment of regional wage disparities depends critically on the availability of adequate and reliable spatially disaggregated regional wage data. Such data is however not readily available in most developing countries, especially in Africa. For example, in South Africa a major source of wage and earnings data is household surveys such as the National Income Dynamics Study (NIDS) collected by the Southern Africa Labour and Development Research Unit (SALDRU), the Post-Apartheid Labour Market Series (PALMS) created by Kerr et al. (2013) from the October Household Surveys (OHS), the Labour Force Surveys (LFS) and the Quarterly Labour Force Surveys (QLFS) collected by Statistics South Africa (Stats SA), and Labour Market Dynamics in South Africa (LMDSA), an annual dataset created by Statistics South Africa (Stats SA) from the four waves of the QLFS.

These surveys contain detailed labour market data, including wages and earnings. Their major drawback is that the data is not spatially disaggregated to small geographic units. Data from these surveys is available mainly at the provincial and district council level. However, these geographical units are highly aggregated, which, as shown by Frame et al. (2016) and Von Fintel (2014) tends to mask significant heterogeneity across smaller geographical units within provinces and district councils.

Interestingly, population census data collected by Stats SA contains labour market information at spatially disaggregated geographical levels, over a period of time. There are, however, arguments that the censuses are not as accurate as the household surveys in measuring labour market outcomes, especially labour market income. This is because the censuses ask about total personal income rather than wages or earnings. Asking individuals their total income is highly problematic as individuals might not honestly report their income due to problems in recalling all their sources of income over time (Davern, Rodin, Beebe, & Call, 2005). Nevertheless, apart from its large sample size, the census remains the only survey to collect detailed labour market information at a detailed spatial level (including municipalities, main places (cities/towns) and sub-places (villages/suburbs). Thus, census data provides an important opportunity to undertake the sort of analysis required to understand inequalities in smaller geographical units in South Africa. For this reason, the main source of data used in this thesis comes from South Africa's full population censuses for the years 1996, 2001 and 2011.

It is important to note that census data is far from ideal (Cronje & Budlender 2004; Ardington et al. 2006) and a lot of work needs to be done to get the data into shape for analysis

(Leibbrandt, Poswell, Naidoo, Welch, & Woolard. 2005). In using census data, four major issues, namely inconsistent geographical units over time, unavailability of wage data, bracketed income and a high proportion of reported zero and missing income are worth noting. We provide a brief discussion of these issues in the next sub-sections.

Inconsistent geographical units over time

One of the advantages of the census data is that it can be aggregated to different geographical units for a given census. However, these geographical units are not consistent across the three census periods under review (see Figure 3.1A in appendix 3.1). This problem of changing area boundaries over time, which is referred to in the literature as the modifiable areal unit problem (MAUP), makes longitudinal analysis of socioeconomic outcomes using census data difficult (Weir-Smith, 2015). To address this challenge, an innovation of this thesis is the use of ArcGIS overlay tools to construct a unique geographically consistent database from the three censuses. To achieve this, we take advantage of the Geographic Information Systems (GIS) shapefiles that accompany the census data, which show the boundaries of various geographical units in South Africa. Based on 2011 sub-place and 1996/2001 magisterial district shapefiles, we use the areal-weighting interpolation technique to assign sub-place population values from the 2011 census to their corresponding 1996/2001 magisterial district population values.

This leads to a geographically consistent database across the three censuses, geo-referenced to 354 magisterial districts[26]. This database enables us to address the objective of this chapter of documenting the key stylised facts that characterise the spatial distribution of wages across regions in South Africa. It also allows us to address the objectives of the forthcoming chapters that look at convergence dynamics and causes of regional wage disparities. Appendix 3.1 provides a detailed explanation of the data construction and ArcGIS mapping exercise, as well as an explanation of why magisterial districts (hereafter – regions) are chosen as the unit of analysis. The appendix also provides a brief explanation of the creation of a homeland status variable using ArcGIS overlay tools.

Unavailability of wage (labour income) data

A major challenge with the censuses is the unavailability of wage (labour income) data which is critical in a study of regional wage disparities. To overcome this challenge, we follow

[26] Magisterial districts represent a highly disaggregated geographical level which allows us to avoid some of the aggregation biases associated with highly aggregated geographical levels (such as provinces and district councils).

existing literature (Redding & Venables, 2004; Bosker & Garretsen, 2012; Breinlich, 2006) and use regional income per worker to proxy for regional wage per worker. Regional income per worker is derived by weighting total income from employed individuals with total employed individuals in each region. Total personal income used to derive income per worker is defined as the sum of basic salary, bonuses, allowances, income from grants, transfers, remittances and any other income source received by individuals. We acknowledge that regional income per worker is an imprecise proxy for regional wage per worker, as it contains income from other sources. However, we argue that it is a good proxy in the case of South Africa, given that labour income (wages) contributes the largest share of total income of employed individuals. To support our argument, we compare income per worker and wage per worker across 53 district councils in South Africa based on NIDS 2010/2011 household survey data.

The NIDS survey collects data on various sources of individual income such as labour (wage) income, government grant income, other government income, investment income, income of a capital nature and remittance income. Of these sources, labour (wage) income is the sum of income from main and secondary job wages, casual wages, self-employment income, 13th cheque, other bonus, profit share, income from helping a friend and extra piece-rate income. From these different sources, we derive three income variables: individual income, income from employed individuals and labour (wage) income. Based on these variables, we further derive district income per worker, given as total income from employed individuals divided by total employed individuals in each district, as well as district wage per worker, given as total labour (wage) income divided by total employed individuals in each district.

Analysis of these variables shows that, on average, income from employed individuals contributes about 72.5 percent of total income across all districts, a figure close to the 73.3 percent we find from the analysis of 2011 census data. Further, on average, labour income accounts for about 69 percent of total income across districts, a figure consistent with the 70 percent normally found in most individual and household studies in post-apartheid South Africa (see Leibbrandt et al. 2010). By narrowing down to labour (wages) income and income of employed individuals only, we find that on average, labour income accounts for 94.7 percent of total income from employed individuals, across districts. This shows that the bulk of income from employed individuals comes from labour income, with roughly 5.3 percent coming from other sources (for example, income from grants, transfers, remittances.)[27]. Figure 3.3A presented in Appendix 3.3 shows this, as a highly positive relationship is revealed between district income per worker and wage per worker. With a correlation coefficient of 0.998, the data points of the 53 district councils lie very close to the correlation line, suggesting that, on average, income and wage per worker are good predictors of each other.

As a further check, we assess whether district specific factors such as industrial composition, human capital, unemployment, and geographical location have any additional effect on the spatial variation of income per worker after the effects of wage per worker are accounted for. For income per worker to be a good proxy for wage per worker, we expect district specific factors to have no additional contribution in explaining variations in district income per worker. To see whether this is the case, we regress district wage per worker on income per worker, controlling for other district-specific factors. The estimation results are presented in Appendix 3.3, Table 3.3A, where column (1) reports estimates for the association between district income and wage per worker, while column (2) reports estimates where we add district specific factors. Columns (1) and (2) show a highly positive and statistically significant association between district income and wage per worker. With the exception of the share of workers in the agricultural sector, all the other district-specific factors are not significant[28]. Based on this analysis, we conclude that income per worker is a good proxy for wage per worker in South Africa.

Bracketed income

A key challenge of the census income data which we used to derive income per worker is that it is collected in brackets (see Table 3.4A in appendix 3.3). To construct a continuous income measure, we assign the midpoint of each bracket to everyone in that bracket[29]. For the highest band, which is open-ended, we set the midpoint to two times the lower bound of the highest bracket, a rule employed by Stats SA. While the midpoint approach has been found to exaggerate income inequality (Wittenberg, 2017)[30] and reduce income variability, it has been

[27] Thus, we expect regional income per worker to overstate regional wage per worker by a small proportion.

[28] Given that the agricultural sector is highly concentrated in rural areas, the significance of the share of workers in the agricultural sector suggests that income per worker might not be a good proxy for wage per worker in rural areas. It is the case that, while workers in rural locations might receive low average wages, they might receive income from other sources, which would show higher income per worker.

[29] For example, a band of 1 to 400 rand will take the value of 200.5 Rands. Alternative approaches include weighting, multiple imputation and non-parametric techniques. For more detail on these methods see Wittenberg (2017).

[30] Although the midpoint approach can exaggerate income inequality, this exaggeration will be consistent over the various censuses, such that income inequality conclusions across the censuses will also be consistent.

found to lead to similar conclusions as other complicated techniques such as the reweighting approach, hot deck approach, mean imputation, and multiple imputation (Posel & Casale, 2005; Ardington et al. 2006; Burger & Yu, 2007; von Fintel, 2007). Its appropriateness has seen it being used widely by other researchers who analyse South Africa's survey data (Hofmeyr, 1999; Casale, Muller & Posel, 2004; Kingdon & Knight, 2004; Meth & Dias, 2004; Leibbrandt et al. 2005; Vermaak, 2005).

A further problem with the brackets is their inconsistency across the three censuses. For instance, 1996 income brackets are narrower than 2001 and 2011 brackets, which are similar. In addition, the top end of the brackets also differs significantly. The value of the top-end bracket in 1996 is set at R30 000 or more and is R204801 or more for 2001 and 2011. To reduce the possible bias due to these inconsistencies and allow comparability of income over time, we compressed the 2001 and 2011 income brackets into their 1996 real income equivalents, using the CPI values provided by Stats SA: 38.5, 52.4 and 92.6 for 1996, 2001 and 2011 census respectively (Stats SA)[31].

Table 3.1 presents the summary statistics of the resulting real income per worker (hereinafter "income per worker") variable. It can be seen that, on average, monthly income per worker has increased across regions (magisterial districts) in South Africa between 1996 and 2011. However, the Min/Max statistics show that income per worker varies significantly across regions, as well as over time.

Table 3.1: Summary statistics for monthly income per worker across regions.

Year	1996	2001	2011
Mean	1553	1953	2272
Sd	608	964	961
Min	743	599	1007
Max	4960	8230	8279

Notes: Summary statistics for monthly income per worker data derived from the full population censuses for the years 1996, 2001 and 2011 across a sample of 354 magisterial districts.

A high proportion of reported zeroes and missing income

A final challenge of the census income data is the high rate of reported zero and missing income. For example, 70.7%, 68.2% and 49.1% of the respondents in 1996, 2001 and 2011 had zero or missing income (see Table 3.5A in appendix 3.3). Interestingly, narrowing down to

[31] http://www.statssa.gov.za/publications/P0141/CPIHistory.pdf?

employed individuals who are the focus of this study significantly reduces the proportion of individuals with zero or missing income to 5%, 2.2% and 13.3% in 1996, 2001 and 2011 (see Table 3.5A in appendix 3.3). Given the employed individuals with missing and zero income, the question is how to deal with these individuals. The most common technique which we use in this study is to drop these individuals. However, dropping employed individuals with missing (missing plus zero) income information can introduce potential bias in parameter estimation (Rubin, 1987; Schafer, 1997)[32]. The extent of this bias is negligible when data is missing completely at randomly (MCAR) and to some extent when data is missing at random (MAR) but not negligible when data is missing not at random (MNAR)[33].

Given this, it is important to check whether income information is MCAR, MAR or MNAR. While it is difficult to check whether data is MAR and MNAR, we can easily check whether data is MCAR. We achieve this by running a logistic regression predicting missingness (0 = not missing, 1 = missing) from specific observed variables. Significant coefficients, either singly or jointly, would indicate a violation of MCAR. Our results presented in Table 3.6A in appendix 3.3 show highly significant coefficients for the bulk of the factors in all columns, indicating a violation of MCAR in 1996, 2001 and 2011. These results suggest that rather than missing completely at random, income information is missing systematically driven by age, education, race, gender, and location[34].

Given these results, excluding employed individuals with missing income information from our analysis might bias our results[35]. Since our analysis is at the regional level, if those workers with missing income data are concentrated at the bottom of the distribution, then the level of income per worker of a given region will be overestimated. Alternatively, if those workers with missing income information disproportionally fall at the top of the distribution, then the level

[32] We assume that all employed individuals who reported zero income have missing income because any employed individual is highly likely to receive a positive income. Thus, we drop employed individuals with missing income information.

[33] Data is MCAR if the probability of missingness does not depend on any variable, either observed or unobserved, while data is MAR if the probability of missingness depends only on observed variables and not unobserved or missing information. MCAR is a special case of MAR. Finally, data is MNAR if the probability of missingness depend on unobserved factors which are not measured by the researcher.

[34] As a robustness check, we also estimated a logistic regression model with regional specific factors like income per worker, market potential, share of workers with higher education and unemployment rate. Our results continued to reveal highly significant coefficients for these factors for all the years, indicating a violation of MCAR.

[35] Interestingly, research suggests that violation of the MCAR does not seriously bias parameter estimates (Collins, Schafer, & Kam, 2001), especially after controlling for those factors highly correlated with the variable of interest (see Allison, 2001). Accordingly, our empirical analysis will control for a number of regional specific factors highly correlated with income per worker, to reduce the potential bias due to omitted workers with missing income information.

of income per worker of a given region will be underestimated. While acknowledging this potential bias, we argue that dropping workers with missing income information will not change our overall conclusions given the small sample size of workers with missing income information in the census (5%, 2% and 13% in 1996, 2001 and 2011). While there is no established cut-off from the literature regarding an acceptable percentage of missing information for valid statistical inferences, our claim finds support from Raymond & Roberts (1987), as well as Schafer (1999) who finds that a missing rate of 5% or less is inconsequential. Moreover, Roth (1994) argued that the choice of a missing data estimation technique can have substantial implications for the parameter estimates as the portion of missing data reaches 15% to 20%. To further support our claim, we will also carry out robustness checks to see whether our main results, which exclude workers with zero income differ significantly from the results when we include workers with zero income.

3.5.1. Exploratory analysis: A Glimpse at the Data

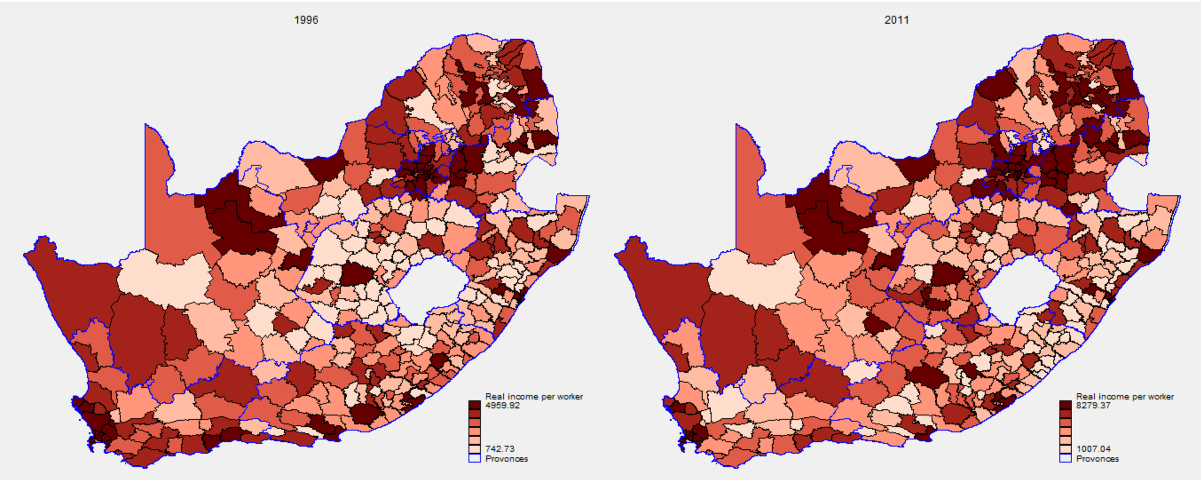
In this section, we start our analysis by visualising the distribution of income per worker across regions in South Africa. Figure 3.2 presents choropleth maps of the spatial distribution of income per worker across the 354 regions in 1996 and 2011[36]. On the maps, the darker colour indicates regions with higher levels of income per worker, while the lighter colour indicates regions with lower levels of income per worker. The two maps highlight three key stylised facts concerning the spatial distribution of income per worker. Firstly, the maps reveal that the distribution of income per worker across space is non-uniform. Secondly, the maps clearly show striking disparities in the distribution of income per worker across regions over the period 1996 - 2011.

Finally, on average, the disparities show spatial clustering of regions with similar (either high or low) levels of income per worker, as well as spatial dispersion of regions with dissimilar (either high and low or low and high) levels of income per worker. Regions with high levels of income per worker are spatially clustered in Gauteng and Western Cape, while those with lower levels of income per worker are spatially clustered mainly in former homeland areas in Free State, Eastern Cape and KwaZulu-Natal (see Figure 3.5A in appendix 3.3 which traces the boundaries for former homeland areas). Regions with dissimilar levels of income per worker

[36] The map for 2001 is given in Figure 3.4A in appendix 3.3. The distribution patterns are generally similar to those revealed in Figure 3.2 for 1996 and 2011.

are located closer to port cities, major cities in former homeland areas, townships and in some parts of Limpopo, Northwest and Northern Cape provinces.

Figure 3.2: The spatial distribution of income per worker across regions (1996 and 2011).



Notes: The map is based on a sample of 354 regions in South Africa, using 1996 and 2011 census data. The blue lines trace out the boundaries of provinces in South Africa.

Taken together, the observed spatial patterns provide evidence of wide variation in income per worker across regions. The variation is characterised by spatial clustering of regions with similar levels of income per worker, as well as spatial dispersion of regions with dissimilar levels of income per worker. To verify whether this is the case, in the next section we turn to information obtained from Moran's I statistic.

3.6. ESDA Empirical results

This section presents the stylised facts that characterise the spatial distribution of income per worker across regions in South Africa and relates these empirical features to predictions of alternative theories (NEG and standard economic theories). The analysis is in two parts. The first subsection looks at global spatial autocorrelation results, while the second subsection focuses on local spatial autocorrelation results.

3.6.1. Global spatial autocorrelation of income per worker across regions

Table 3.2 presents the results of the global Moran's I statistic of log income per worker for the years 1996, 2001, and 2011 for South Africa's 354 regions, based on the distance spatial weight matrix. The results indicate evidence of highly significant and positive global spatial autocorrelation (concentration) in the distribution of income per worker in all the years[37]. The results thus rejects the hypothesis of spatial randomness in favour of positive spatial concentration of income per worker over the whole study area in all the years. This supports the first impression from Figure 3.3, suggesting that within South Africa, spatially contiguous regions tend on average to register similar income per worker. This implies that, apart from a region's own economic conditions, income in each region is influenced by income and economic conditions of neighbouring regions. Rather than developing in isolation, as postulated by standard economic theory, regions tend to be interdependent and influence each other's development path, as suggested by NEG theory.

Table 3.2: Global Moran's I statistic for log income per worker (1996, 2001, and 2011)

Variables	1996	2001	2011
Moran's I statistic (I)	0.311	0.272	0.224
Moran's I statistic expected value - E(I)	-0.003	-0.003	-0.003
sd(I)	0.015	0.015	0.015
Z	21.124	20.517	16.708
p-value*	0.0000	0.0000	0.0000

Notes: Inference is based on a standardised z-value that follows a normal distribution. To evaluate the hypothesis of spatial randomness, $I = 0$ vs spatial autocorrelation, $I \neq 0$, we compare the value of Moran's I statistic with the expected value of Moran's I statistic. The hypothesis of spatial randomness is rejected in favour of spatial autocorrelation when the calculated Moran's I statistic is larger than the expected value, and in our case, it is rejected in all years.

The evidence of positive global spatial autocorrelation is consistent with the international literature, as well as the literature on South Africa that focuses on other economic outcomes. However, its evolution dynamics and magnitude differs significantly from what is observed in other countries. Our results confirm evidence of decreasing positive global spatial autocorrelation across regions, as the Moran's I statistic fell from 0.311 in 1996 to 0.224 in 2011[38]. However, Huang & Chand (2015) find evidence of increasing positive global spatial autocorrelation across 30 provinces in China, as the Moran's I statistic increased from 0.214 to 0.231 between 2001 and 2010. Looking at 287 Canadian regions, Breau & Saillant (2016) also

[37] Acknowledging that our results might be sensitive to the inclusion of employed individuals with zero income, we re-examined for evidence of global spatial autocorrelation using income per worker which includes workers with zero income. As shown in Table 3.6A in Appendix 3.3, the magnitude of the global Moran's I statistic of log income per worker remained similar to those when we excluded employed individuals with zero incomes. We, therefore, concluded that dropping employed individuals with zero income has no major effect on our estimates.

[38] Evidence of falling spatial autocorrelation of income per worker across regions might have significant implications for regional disparities in levels of income per worker. It might suggest an increase in dispersion of income per worker in space, which in turn might imply a fall in regional disparities in income per worker. The next chapter of this thesis will shed more light on whether this is the case by testing the convergence hypothesis.

find evidence of increasing positive global spatial autocorrelation, with the Moran's I statistic increasing between 1996 and 2006 from 0.526 to 0.560.

Spatial autocorrelation results depend heavily on the spatial weight matrix used. To check the robustness of our findings, we re-calculate the global Moran's I statistic using different spatial weight matrices. We start by checking the robustness of our results to different distance cut-off points. For this, we use 205 km as the benchmark cutoff distance and incrementally increase it up to 1000 km, before using the simple inverse distance function without a cut-off (looks at the full distance of 1795 km).

As evident in Table 3.3, regardless of the distance cut-off used, Moran's I statistic value remains positive and highly significant in all the years. However, it can be seen from the table that the intensity of global spatial autocorrelation decreases with increasing distance. This suggests that regions in close proximity to each other tend to be more spatially related than distant ones. This evidence supports Tobler's First Law of Geography, which states that *"Everything is related to everything else, but closer things are more related* (Tobler, 1970, pg 236)", as well as the NEG theory prediction of a core-periphery economic structure in which income per worker decreases gradually with increasing distance from central regions with greater access to markets (Hanson, 2005). Given this evidence, regional economies cannot be treated as independent entities that develop in isolation, as levels of income per worker in one region are influenced by income per worker of surrounding regions.

Table 3.3: Moran's I statistic under different distance cut-off points.

	1996		2001		2011	
Distance cutoff with $\alpha = 1$	I	P-value	I	P-value	I	P-value
205	0.311	(0.000)	0.272	(0.000)	0.224	(0.000)
300	0.277	(0.000)	0.267	(0.000)	0.222	(0.000)
400	0.247	(0.000)	0.242	(0.000)	0.199	(0.000)
500	0.226	(0.000)	0.223	(0.000)	0.184	(0.000)
600	0.214	(0.000)	0.213	(0.000)	0.173	(0.000)
700	0.207	(0.000)	0.208	(0.000)	0.166	(0.000)
800	0.201	(0.000)	0.203	(0.000)	0.162	(0.000)
900	0.194	(0.000)	0.196	(0.000)	0.157	(0.000)
1000	0.185	(0.000)	0.188	(0.000)	0.151	(0.000)
No cut-off	0.162	(0.000)	0.163	(0.000)	0.132	(0.000)

Notes: Our variable of interest remains income per worker. The distance is based on the inverse-distance function with a distance decay parameter, $\alpha = 1$. I gives the calculated Moran's I statistic. The calculated Moran's I statistic (I) is compared to the expected value of Moran's I statistic, which is -0.003.

Using an inverse distance weight matrix with a distance cutoff of 205 km as the benchmark weight matrix, we further check the robustness of our results to the use of different types of spatial weight matrices. Specifically, we use the binary contiguity matrix, where the presence, nature, and extent of spatial autocorrelation depend on whether regions share a common boundary. We also use the k-nearest neighbour weight matrix, where spatial autocorrelation is based on each region having exactly the same number (k) of neighbours and we consider, k equal to 5, 10, 15 and 20 nearest neighbours, respectively.

Table 3.4: Moran's I statistic under different weight matrix specifications.

	1996		2001		2011	
Matrix type	I	p-value*	I	p-value*	I	p-value*
Inverse distance ($d_{ij} < 205$)	0.310	(0.000)	0.272	(0.000)	0.224	(0.000)
Binary contiguity	0.423	(0.000)	0.438	(0.000)	0.374	(0.000)
5 neighbours	0.503	(0.000)	0.510	(0.000)	0.396	(0.000)
10 neighbours	0.467	(0.000)	0.476	(0.000)	0.365	(0.000)
15 neighbours	0.405	(0.000)	0.392	(0.000)	0.302	(0.000)
20 neighbours	0.354	(0.000)	0.331	(0.000)	0.258	(0.000)

Notes: Our variable of interest remains income per worker. The calculated Moran's I statistic (I) is compared to the expected value of Moran's I statistic which is -0.003.

As with the inverse distance weight matrix, results in Table 3.4 confirm evidence of strong and positive global spatial autocorrelation across the different spatial weight matrices. However, spatial autocorrelation is higher under the k-nearest neighbour weight matrix (where k = 5 and 10), followed by the binary contiguity matrix, and, lastly, the inverse distance matrix. Thus, regardless of the specification or type of spatial weight matrix used, our results continue to show that the spatial distribution of income per worker is characterised by spatial concentration patterns defined by significantly positive spatial autocorrelation for all the years. Thus, based on these results, the assumption of spatial randomness in the distribution of real income per worker is overwhelmingly rejected. These results also validate the appropriateness of magisterial districts as our spatial unit of analysis. Dall'erba (2005) argues that, when the spatial unit of analysis is chosen correctly, the results from the different spatial weight matrices should be similar. This is the case here. Thus, in the remaining analysis, we continue to use the inverse distance spatial weight matrix with a distance cut-off of 205 km.

3.6.2. Local spatial autocorrelation of income per worker across regions

The global Moran's I statistic provides a single value for the entire study area and cannot discriminate between spatial autocorrelation of high values and spatial autocorrelation of low values in the case of global positive spatial autocorrelation. Furthermore, it may mask regions that deviate from the global positive spatial autocorrelation. To address these limitations, we complement the global Moran's I analysis with Moran scatterplots, which show the spatial relationship between the value of income per worker (z) of each region and the standardised spatial weighted average income per worker (Wz) of all neighbouring regions.

The resulting Moran scatterplots of log income per worker for 1996, 2001, and 2011 are displayed in Figure 3.4. As explained earlier, the scatterplots consist of four quadrants[39], where the x-axis captures the value of income per worker (z) of each region, which is compared to the average value of the sample. Thus, all regions to the right of the vertical axis starting at zero have neighbouring regions with incomes for workers which are higher than the sample average. On the other hand, the y-axis captures the standardised spatial weighted average income per worker (Wz) of all neighbouring regions. Therefore, all the points below the horizontal axis starting at zero capture regions which have neighbouring regions which display incomes for workers which are lower than the sample average.

The results in all the years corroborate our earlier findings from the global Moran's I statistic of positive and significant spatial autocorrelation. This can be seen by the red line in the plots whose slope is the global Moran's I statistic value of the corresponding years. However, the picture depicted in the plots clearly shows that the spatial distribution of income per worker is more complex than the simple positive spatial autocorrelation revealed by the global Moran's I statistic.

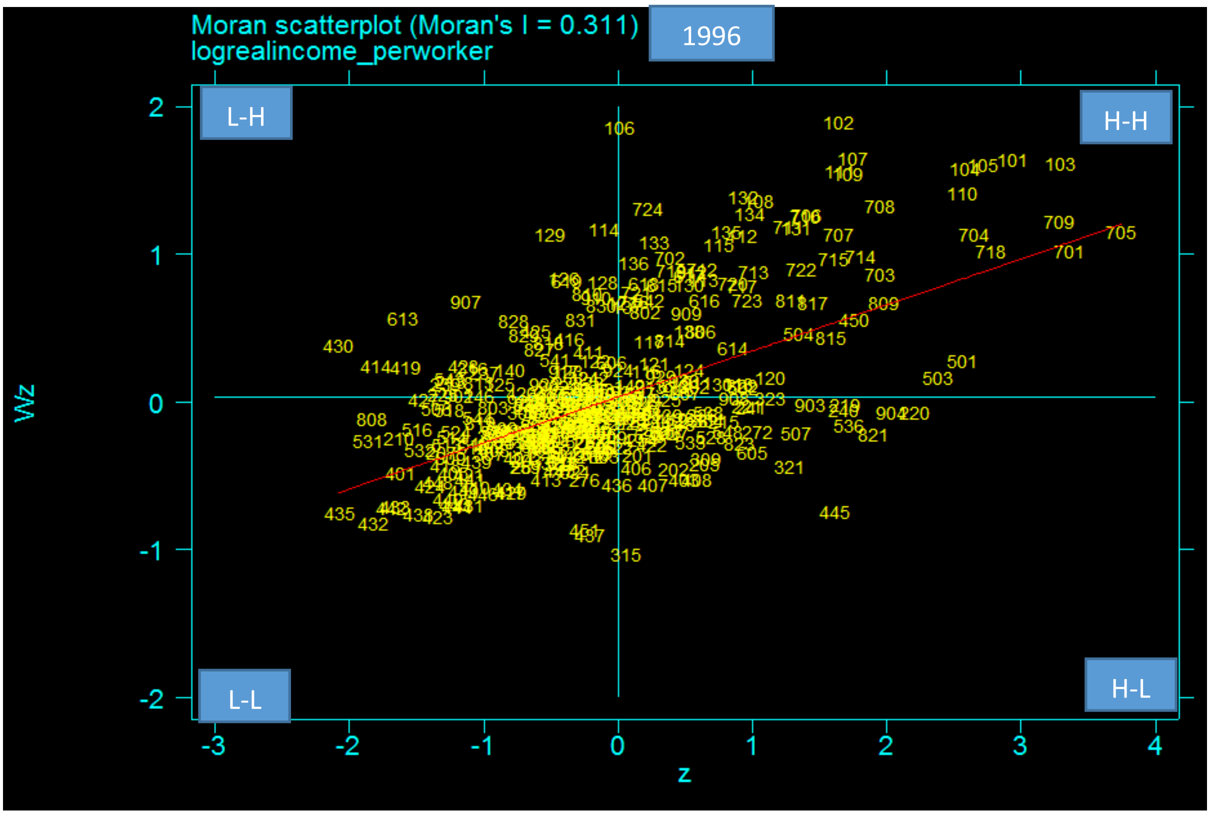
The overall positive global spatial autocorrelation pattern masks significant spatial heterogeneity, as workers' incomes spatially cluster in specific geographical locations. These locations are defined by four spatial cluster regimes: H-H, L-L, H-L and L-H. The H-H regime comprises mainly regions in Gauteng and Western Cape. The L-L regime contains mostly regions in former homeland areas[40] in KwaZulu-Natal, Eastern Cape, and Free State. These

[39] A recap of these quadrants: (1) the upper right quadrant, high-high (H-H) represents regions with high wage values that are surrounded by other regions with high wage values, (2) the upper left quadrant, low-high (L-H) captures regions with low wage values that are neighbours to regions with high wage values, (3) the lower left quadrant, low-low (L-L) displays locations with low wage values that are surrounded by other regions with low wage values, while (4) the lower right quadrant, high-low (H-L) gives regions with high wage values that are neighbours to regions with low wage values.

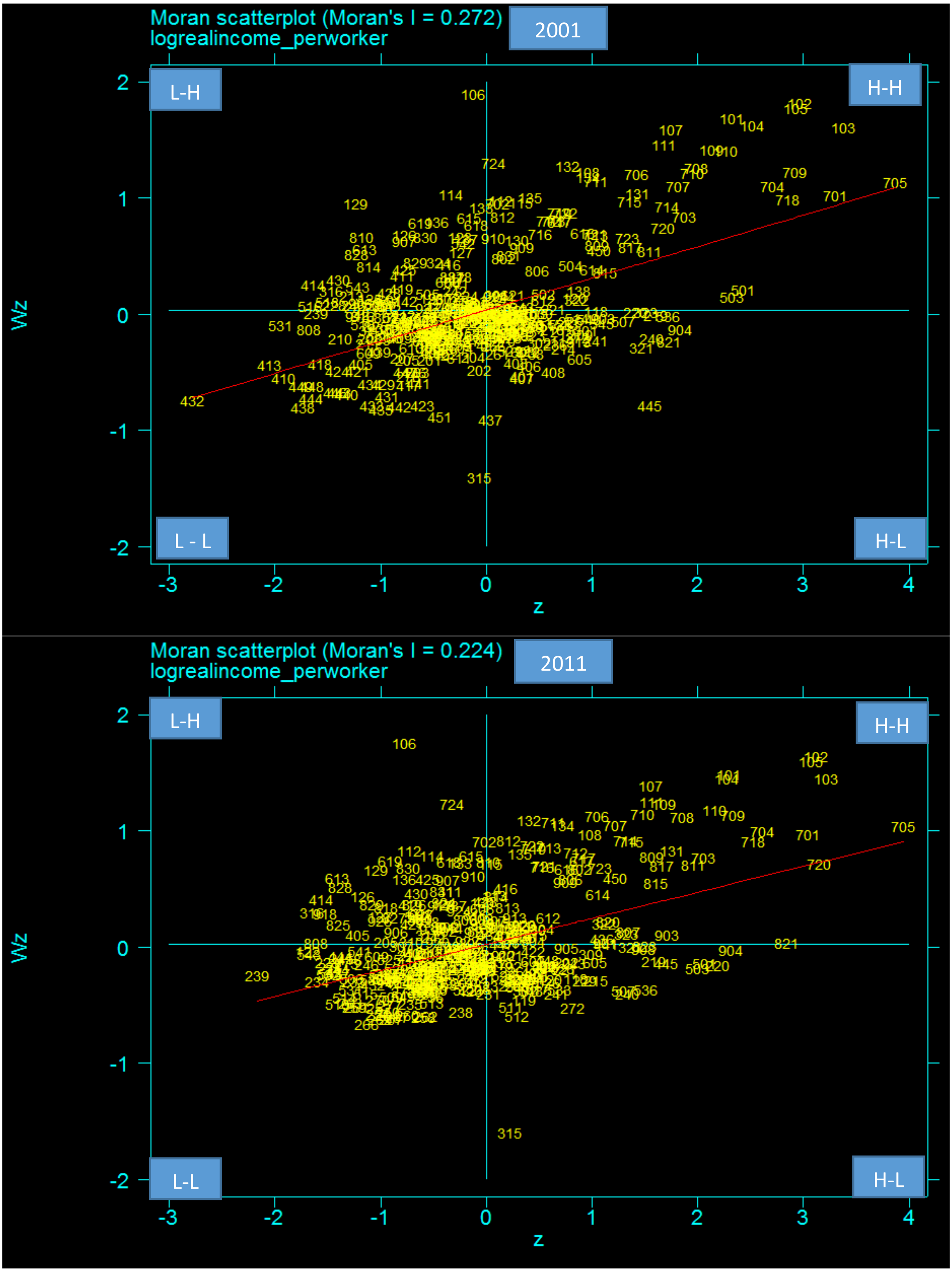
[40] A summary of all the regions showing statistically significant local spatial autocorrelation by homeland status shows that, of the 44, 37, and 45 regions in 1996, 2001 and 2011, respectively that fall in the L-L quadrant, 20, 16 and a massive 41 falls in former homeland areas. These number clearly confirms that regions with low levels of income per worker are concentrated in the former homeland areas.

regimes reflect spatial clustering of regions with similar (high or low) levels of income per worker, which drive positive spatial autocorrelation. On the other hand, the H-L and L-H regimes comprise mainly wealthy regions scattered around the country[41], as well as poor regions (townships) in major urban centres. These regimes capture spatial clustering of regions with dissimilar (high and low or low and high) levels of income per worker, which drive negative spatial autocorrelation.

Figure 3.3: Moran's scatterplot for log income per worker (1996, 2001, 2011)



[41] Existence of these regions supports earlier results by Bosker & Krugell (2008), who find that South Africa is characterised by rich regions that are scattered around the country that act as local growth poles, absorbing economic activity from their poor neighbours.



Notes: The three scatter plots are based on the inverse distance spatial weight matrix with a cut-off distance of 205 km across the 354 magisterial districts.

Complementing the Moran scatter plot with local Moran's I statistic, Table 3.5 indicates the number of regions confirming evidence of significant local spatial autocorrelation. Of the 354 regions, 113, 103 and 117 regions show evidence of significant local spatial autocorrelation in 1996, 2001 and 2011, respectively. Tables 3.8A – 3.10A in Appendix 3.3 list the names of all these regions. An important observation from Table 3.5 is the high number of regions not confirming evidence of significant local spatial autocorrelation in all the years[42]. This is a common finding in other studies. For instance, Breau & Saillant (2016) find that 111 and 122 out of 287 regions show no evidence of significant local spatial autocorrelation in the distribution of wages in Canada in 1996 and 2006, respectively. Nevertheless, a look at the different spatial regimes show that the bulk of the regions confirming significant local spatial autocorrelation belong to the H-H and L-L regimes, while the H-L and L-H regimes contain the least number of regions. The dominance of the L-L and H-H regimes confirm the existence of a dualistic economic structure for South Africa, showing a strong rural-urban division, where the rural areas correspond with the L-L spatial regime, while urban locations are in line with the H-H spatial regime.

Table 3.5: Summary of regions confirming significant local spatial autocorrelation.

Spatial regime	Spatial autocorrelation type	1996	2001	2011
H-H	Positive	51	45	46
L-L	Positive	44	37	45
H-L	Negative	7	7	10
L-H	Negative	11	14	16
Not Significant	Zero	241	251	237
Total significant		113	103	117
Total regions		354	354	354

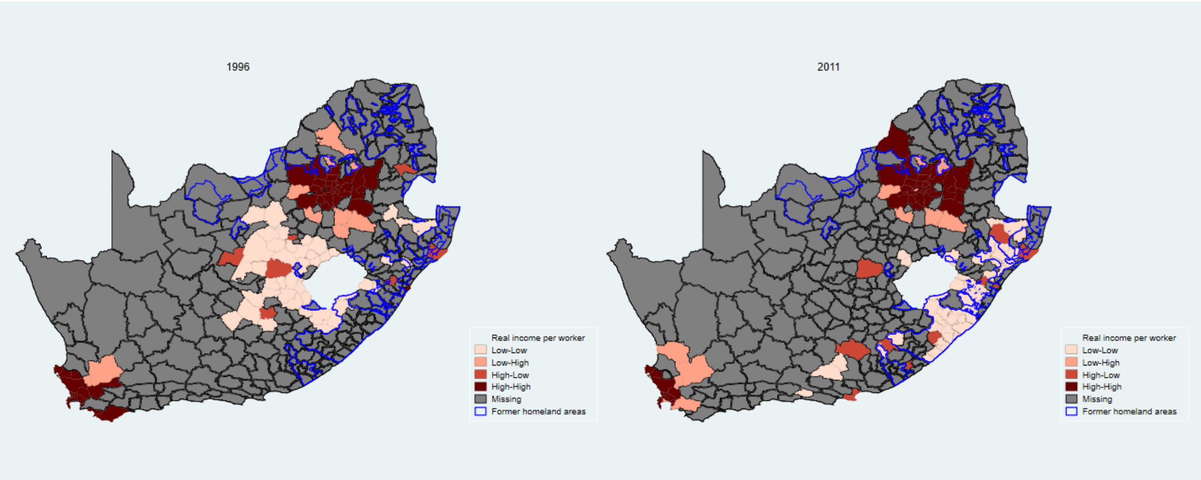
Notes: The table provides a summary of all the regions with statistically significant local spatial autocorrelation coefficients for regional income per worker.

Figure 3.4 shows clearly that the H-H and L-L regimes have large and well-defined patterns, with a strong rural-urban division, whereas the H-L and L-H regimes reveal small, less-defined and scattered patterns across the country. Furthermore, Figure 3.5 shows that, while the H-H regime remains located in Western Cape and Gauteng, the L-L regime shifted from the central part of the country, which is Free State to the former Transkei found in KwaZulu-Natal. This shift is rather surprising given that between 1996 and 2011 KwaZulu-Natal's economy

[42] Figure 3.4 shows clearly that the bulk of the regions showing insignificant evidence of local spatial autocorrelation in 1996 fall in former homelands, while in 2011 some regions in former homelands, particularly in KwaZulu-Natal show evidence of positive spatial autocorrelation characterised by clusters of regions with low levels of income per worker.

performed well compared to Free State[43]. Taken together, these four regimes point to an economy characterised by significant disparities in the distribution of incomes for workers across regions, which tend to persist over time. Furthermore, they point to an economy characterised by highly localised spatial autocorrelation processes, which, according to Monastiriotis (2009), suggests the existence of a highly fragmented and heterogeneous spatial economy that lacks economic forces to promote strong linkages between distant regions.

Figure 3.4: Distribution of regions showing significant local spatial autocorrelation.



Note: In Figure 3.4, the legend shows the colour scheme for the different spatial regimes for regions showing significant local spatial autocorrelation in the distribution of income per worker in South Africa in 1996 and 2011.

Comparing the above empirical facts to theoretical predictions, it is apparent that neither NEG theory nor standard economic theory predictions fully match all the empirical features found in the data. Rather, the observed facts (spatial patterns) are partially consistent with both theories. For instance, the large and well-defined positive spatial patterns captured by the H-H and L-L regimes are consistent with the core-periphery economic structure predicted by NEG theory (Krugman, 1991; Krugman & Venables, 1995). As would be expected based on NEG theory, regions in Gauteng and Cape Town are located in the H-H quadrant and constitute the geographical core. These regions seem to have good access to markets, which in turn, promotes concentration of economic activities and higher incomes and wages in these regions. For example, our data shows that 24 (out of 354) regions in Gauteng are less than 1.5% of the country's total land area, but account for 22 percent of the national population, 33 percent of national employment, and a massive 40 percent of national income in 2011. By comparison,

[43] For example, while KwaZulu-Natal registered an annual real economic growth of 3.6% and contributed 15.7% to the South African economy in 2011, Free State recorded an annual real economic growth of 2.5% and contributed 5.3% to the South African economy in 2011.

regions in the Free State, Eastern Cape and KwaZulu-Natal located in the L-L regime correspond with the geographical periphery and seem to be characterised by poor access to markets, which in turn leads to low levels of economic activity, incomes, and wages.

The empirical literature acknowledges that negative spatial autocorrelation patterns such as those revealed by the H-L and L-H regimes are hard to explain (Khomiakova, 2007). However, given the identity of the regions in these quadrants, these patterns are consistent with predictions of the standard economic theory that emphasize the importance of region-specific factors. For instance, the H-L quadrant consists of regions with greater access to waterways (Durban), greater tourism activities, driven by wildlife (Nelspruit), and highly endowed with mineral resources (Pietersburg)[44]. These factors seem to explain the above-average incomes of workers in these regions. The L-H quadrant consists of regions such as Mitchell's Plain and Soweto, which coexist with wealthy regions in major urban areas. This suggests that their low incomes for workers are a result of apartheid-era spatial and settlement planning, which led to the creation of poor and highly marginalised townships to accommodate black workers within urban areas.

Thus, it seems clear that the positive and negative spatial concentration patterns that characterise the distribution of income per worker in South Africa are consistent with predictions from both NEG and standard economic theories. However, with these patterns largely dominated by positive spatial autocorrelation (H-H and L-L regimes) rather than negative spatial autocorrelation (H-L and L-H), it is clear that NEG theory offers a plausible explanation for the observed empirical characteristics found in the South African data.

3.7. Conclusion

In this chapter, we carry out an exploratory analysis of the spatial distribution of income per worker (used as a proxy for a regional average wage) across regions in South Africa. To conduct the analysis, the chapter constructs a highly disaggregated regional database for 354 regions (magisterial districts) using South Africa’s population census for the years 1996, 2001 and 2011. The data is then used to document the key stylised facts that characterise the spatial distribution of income per worker across regions in South Africa and assesses their consistency

[44] While the study uses magisterial district names given in the 1996/2001 census, we acknowledge that places like Nelspruit and Pietersburg have since changed names to Mbombela and Polokwane respectively.

with predictions from alternative theories. To achieve this, the study utilises exploratory spatial data analysis techniques. Below we discuss the main findings to emerge from the analysis.

The analysis finds striking disparities in the distribution of income per worker across regions, which persist over time. In line with evidence in the international literature, the disparities are characterised by positive global spatial autocorrelation, which is robust to the use of different spatial weight matrices. This means that, rather than randomly distributed across space, income per worker spatially concentrates in specific locations, with spatially adjacent regions registering on average similar income per worker figures. More specifically, regions with high (low) income per worker are close to other regions with high (low) income per worker figures. This finding points to regional interdependence, as incomes of workers in one region tend to be influenced by those in neighbouring regions. Over time, the study reveals evidence of decreasing trends in the extent of positive global spatial autocorrelation. This suggests a reduction in the spatial concentration of incomes of workers, which in turn might suggest a reduction in disparities in income per worker across regions. This issue will be discussed in the next chapter.

Another important finding from the analysis is that the positive global spatial autocorrelation masks significant spatial heterogeneity in the distribution of income per worker defined by four spatial cluster regimes: H-H, L-L, H-L and L-H. On one hand, the H-H regime, which is comprised mainly of the city regions of Gauteng and Western Cape, and the L-L regime, which contains mainly rural regions in former homeland areas in KwaZulu-Natal, Eastern Cape, and Free State, reflect the spatial concentration of regions with similar (high or low) incomes of workers. These two regimes contain regions driving positive spatial autocorrelation. The H-L and L-H regimes, which comprise mainly cities scattered around the country and townships in major urban centres, capture spatial concentration of regions with dissimilar (high and low, or low and high) levels of income per worker. The two regimes reflect regions driving negative spatial autocorrelation. Thus, the study shows that the spatial distribution of incomes of workers is characterised by persistently positive and negative spatial autocorrelation, but dominated by positive spatial autocorrelation. The dominance of positive spatial autocorrelation confirms the existence of a dualistic structure for the South African spatial economy, with a strong rural-urban division. Taken together, these four spatial regimes confirm the existence of significant disparities in the distribution of income per worker across regions in South Africa between 1996 and 2011.

Another key finding from the study consistent with existing evidence from other countries is that, despite evidence of strong positive spatial autocorrelation, only a few regions confirm evidence of statistically significant local spatial autocorrelation. The results show that, of the 354 regions, 113, 103 and 117 confirm evidence of significant local spatial autocorrelation in 1996, 2001 and 2011, respectively. This suggests highly localised spatial autocorrelation processes, that, according to Monastiriotis (2009), confirm the existence of a fragmented and heterogeneous spatial economy that lacks economic forces to promote strong linkages between distant areas. Given this evidence, we can conclude that many regions are not integrated into the economy. It appears necessary for policy to focus on interventions that promote strong linkages across regions to ensure greater integration of regional economies.

This chapter also assesses the consistency of the empirical features found in the data to predictions of NEG and standard economic theories. The results show that, while none of the theories fully match all the empirical features found in the data, both NEG and standard economic theory features exist in the data. For instance, evidence of positive spatial autocorrelation revealed by the H-H and L-L spatial cluster regimes, with a large and well-defined rural-urban division, is consistent with the core-periphery economic structure predicted by NEG theory (Krugman, 1991; Krugman & Venables, 1995). On the other hand, the existence of negative spatial autocorrelation shown by the H-L and L-H spatial regimes, with small, isolated, and less-defined spatial patterns, is consistent with a standard economic theory explanation. However, with the spatial patterns dominated by positive spatial autocorrelation, NEG theory offers a more plausible explanation for the observed spatial features in the South African data.

The exploratory analysis presented in this chapter provides important insights into the spatial distribution of income per worker across regions in South Africa. These insights are an initial step in a better understanding of the South African spatial economy. In Chapter 4 we deepen our understanding of the South African spatial economy further by explicitly examining the extent to which real income per worker has converged or diverged across regions over the period 1996 – 2011. We do this by testing the neoclassical convergence hypothesis, which states that, over time, regional economic disparities tend to decrease, leading to regional economic convergence.

Chapter 4

4. Regional convergence dynamics of wages in post-apartheid South Africa

4.1. Introduction

In this chapter, we extend the analysis discussed in the previous chapter by examining regional convergence dynamics of wages in post-apartheid South Africa. In general terms, convergence refers to the progressive diminishing trend in the differences in levels of wages or other economic outcomes between rich and poor regions (Laurini, Andrade, & Pereira, 2005). The issue of whether wages are converging or diverging over time and across regions within a country is important for a number of reasons.

Firstly, it helps to distinguish between predictions of alternative economic theories. For instance, evidence of regional wage convergence can be seen as an empirical validation of the neoclassical economic theory convergence hypothesis. This hypothesis states that, in a well-functioning economy, diminishing returns to capital, factor mobility, and interregional trade promote convergence of factor prices (wages), and consequently convergence of income and standard of living (Hofer & Wörgötter, 1997; Alexiadis, 2010; Ferens, 2015). On the other hand, evidence of regional wage divergence can be seen as supporting alternative economic theories, among them, endogenous growth theory and human capital theory, that predict persistent or even increasing regional economic disparities driven by human capital differences. Evidence of regional wage convergence or divergence can also be in line with the new economic geography (NEG) theory, which emphasises the importance of access to markets, driven by increasing returns to scale, transport costs, and consumers' love of variety.

Secondly, it assists in evaluating the effectiveness of regional policy initiatives implemented to promote regional economic development and address regional wage disparities. Evidence of regional wage convergence can be interpreted as a sign of the effectiveness of existing regional policy initiatives (Petrakos et al. 2005). In addition, it helps in the design of regional policy initiatives by identifying the factors promoting or hampering regional wage convergence[45]. Thirdly, it sheds light on why regional wage disparities continue to persist even across regions in countries with high labour mobility and interregional trade. Finally, a study of regional wage

[45] Implementation of effective policy initiatives is a fundamental objective of most governments given that persistent wage disparities can lead to wide disparities in welfare, which in turn can be a source of social tension.

convergence can provide available information about the future trends of overall income inequality.

Thus, a growing number of studies have empirically examined whether wages are converging or diverging across regions in many countries (Carlino & Mills, 1996; Tavernier & Temel, 1997; Bukenya, Davis, Banerjee, & Gyawali, 2011; Moazzami, 1997; Ferens, 2015;Naz, Ahmad, & Naveed, 2017). Evidence has been mixed, due in part to the measures of convergence used, statistical methods employed, and country and time-period of analysis. Furthermore, little attention has been paid to countries in Africa, even though there are important differences in the distribution of economic activities, wages, incomes and growth performances within countries in Africa. In addition, Africa has the poorest and some of the most unequal countries in the world[46].

This chapter investigates the extent to which wages have converged or diverged across regions in South Africa. Using the neoclassical convergence hypothesis as the basis of the analysis, the study addresses the following two questions:

- Are regional wage disparities in post-apartheid South Africa converging or diverging over time?
- What are the potential drivers of the observed patterns of convergence or divergence?

The chapter's contribution is threefold. First, it contributes to the regional science literature by examining regional wage convergence, in the context of Africa, a region with few studies of this nature. Secondly, it introduces a study of regional wage convergence as an added dimension to the existing convergence literature on South Africa that has so far focused attention on regional convergence dynamics of GDP per capita. Thirdly, it adds to the literature by updating convergence analyses to include the most recent period (1996-2011), which has so far been neglected in regional convergence studies in South Africa. The current body of research focuses on regional convergence dynamics in levels of GDP per capita over the period 1990 – 2004 (Naudé & Krugell, 2003, 2005, 2006; Bosker & Krugell, 2008). While this

[46] Africa is the world's second most unequal region after Latin America (Ravallion & Chen, 2012), with six of the ten most unequal countries in the world in Africa and more specifically in Southern Africa in 2010 (African Development Bank, 2012). The six countries are Angola, Botswana, Comoros, Lesotho, Namibia, South Africa, and Swaziland. The most striking increase in inequality was found to be in South Africa, whose Gini coefficient rose from 58 to 67 between 2000 and 2006 (African Development Bank, 2012). Secondly, available evidence suggests that spatial inequality (regional disparity) contributes close to 40% of overall asset inequality in Africa (Shimeles & Nabassaga, 2015).

research finds little to no evidence in support of regional convergence between 1990 and 2004, this study assesses whether the trend continues to hold over the same period, as well as beyond 2004, using a different measure of regional economic disparity (wages). Finally, it contributes to the practical policy debate in South Africa, where mitigating regional economic disparities and ensuring regional equalisation of living standards is a key objective of government regional policies, such as the National Spatial Development Framework (NSDF) of 1995, Spatial Development Initiatives (SDIs) of 1996, Regional Industrial Development Strategy (DTI, 2006), and National Spatial Development Perspective (NSDP, 2003; 2006).

The rest of the chapter is organised as follows: Section 4.2 provides a discussion of the key theoretical insights on regional wage convergence. Section 4.3 offers a discussion of the alternative measures of economic convergence. Section 4.4 provides a review of the related empirical literature; section 4.5 presents the empirical framework. Section 4.6 provides a description of the data. Section 4.7 and 4.8 present the empirical results, and section 4.8 offers conclusions.

4.2. Theoretical insights on regional wage convergence

Economic convergence is an implication of the neoclassical growth theory set out by Solow (1956) and Swan (1956). In this theory, income per capita (worker) for a group of economies (countries or regions) with similar preferences and access to similar technology converges over time, irrespective of initial conditions. The theory suggests that poor regions with low capital-labour ratios tend to have higher (lower) returns to capital (labour), while richer regions with higher capital-labour ratios tend to have lower (higher) returns to capital (labour). These conditions provide incentives for factors of production to move between regions, as workers migrate away from low-wage regions towards high wage regions, while firms move away from the more affluent regions where returns to capital are low, towards lagging regions with higher returns. The migration of workers and firms in response to regional differentials in factor returns erodes regional disparities in factor prices, leading to the equalisation of wages, incomes, and consequently standard of living across regions (Alexiadis, 2010; Ferens, 2015).

The neoclassical growth theory convergence prediction is known in the growth and convergence literature as the “convergence hypothesis” (Quah, 1996; Monastiriotis, 2014). This hypothesis finds support from neoclassical trade models. These include the comparative advantage model, which highlights the importance of interregional trade (Heckscher, 1919; Ohlin, 1933; Samuelson, 1949), and the factor movement model, which emphasises the importance of free factor movement (Rybczynski, 1955)[47]. Thus, neoclassical economic theories suggest that, while regional economic disparities initially arise in the process of reallocation of resources, in a perfectly competitive market, diminishing returns to capital, factor mobility, and interregional trade promote equalisation of factor prices. This leads to regional convergence in wage levels, GDP per capita, and, in the long run, living standards (Alexiadis, 2010; Ferens, 2015). Thus, proponents of the neoclassical economic theory argue against the implementation of costly regional policy interventions, as market forces ensure regional economic convergence in the long run.

Empirical evidence supports the neoclassical convergence hypothesis (see Magrini, 2004 for a review of the literature), but indicates that limitations to diminishing returns to capital, factor mobility, and interregional trade can arise from the existence of region-specific factors, such as human capital, mineral resources, local institutions, and local unemployment rates. This view is supported by other schools of thought, such as endogenous growth theory (Romer, 1986; Lucas, 1988), human capital theory (Becker, 1962), wage curve theory (Blanchflower & Oswald, 1995) and amenity theory (Roback, 1982) that predict regional economic divergence. These theories see regional policy as necessary for reducing regional wage disparities, because region-specific factors can sustain regional wage disparities even across regions in countries where factor mobility and interregional trade is high.

Another school of thought, the New Economic Geography (Krugman 1991), predicts neither regional economic convergence nor divergence, but argues that access to markets driven by increasing returns to scale, transport costs and consumers' love of variety determine a region's economic activity, income, and wages. The main insight from this theory is that the spatial concentration of firms and consumers creates positive and negative externalities that raise local wages in those regions that offer greater access to markets (Mion & Naticchioni, 2009).

[47] The factor movement model predicts that free factor movement promotes regional factor price convergence as low-wage, less advanced regions attract capital, while high-wage, more advanced regions attract labour. The comparative advantage model states that interregional trade promotes regional convergence in both product and factor prices, as regional economies specialise in the production and exporting of goods that utilise intensively their abundant and cheap factor of production, while importing those goods that require an intensive use of their scarce and expensive factor of production. This model suggests that interregional trade can substitute migration of factors of production such that regional factor price convergence occurs even without regional factor mobility (O'Rourke & Williamson, 1999; Zaman & Goschin, 2014).

In summary, from this brief review, it is clear that economic theory does not give a complete answer on whether regional economic convergence or divergence will be observed over time. This implies that occurrence of regional convergence or divergence in various economic outcomes can only be verified empirically. In the next sections, we provide a review of the approaches used to measure convergence, as well as a review of the related empirical literature.

4.3. Measures of economic convergence

Economic theory is clear on what regional economic convergence means. However, measuring economic convergence is not a trivial task (Juessen, 2009). As a result, various alternative approaches have been used to measure economic convergence. One such approach is the distributional dynamics approach, consisting of kernel density and Markov chain estimators. This approach originates with Quah (1993, 1996a, 1996b), and evaluates the convergence hypothesis by examining the changes over time in the external shape and intra-distribution dynamics of the entire distribution of wages (or GDP per capita). An alternative approach is σ-convergence pioneered by Barro & Sala-i-Martin (1991). This approach provides a summary measure of the extent of the dispersion of wages across regions (or countries) at a given point in time. There is evidence of σ-convergence when the dispersion in wages across regions, measured by indices such as the standard deviation, coefficient of variation, and the Gini coefficient, decrease over time, while an increase in these three indicators point to σ-divergence (Barro & Sala-i-Martin, 1991; Ferens, 2015).

Another approach introduced by Baumol (1986) is β-convergence, which is understood as the tendency of initially poor regions to grow faster than initially rich ones, thereby catching-up with them (Barro & Sala-I-Martin, 1991, 1992; Ferens, 2015). One can distinguish between two types of β-convergence: Unconditional and conditional β-convergence (Sakamoto & Islam, 2008; Zaman & Goschin, 2014). Under unconditional β-convergence, regional economies are assumed to be homogeneous (have identical structural characteristics), such that, in the long run, economies converge to the same steady state level of wages and GDP per capita (Ghosh, 2008; Sakamoto & Islam, 2008; Artelaris, Arvanitidis, & Petrakos, 2011)[48]. On the other hand, under conditional β- convergence, regional economies are assumed to be heterogeneous, such that, in the long run, economies converge to their own steady state level

[48] A steady-state refers to a situation where the growth rates of all variables are constant.

of wages and GDP per capita, in line with their underlying structural characteristics (Sakamoto & Islam, 2008).

Empirically, there is evidence of unconditional β-convergence when the relationship between the growth rate and initial wages is negative. Evidence of conditional β-convergence is shown when the same relationship is negative, but after controlling for regional differences in structural factors (Barro, 1991; Barro & Sala-i-Martin, 1992). By controlling for regional structural characteristics, conditional β-convergence can be said to be assumed, as regional disparities in wages and GDP per capita can persist, or even increase over time (Artelaris et al. 2011). In the empirical literature, studies testing for evidence of unconditional and conditional β-convergence have utilised cross-sectional, panel, and time-series data models.

4.4. Related empirical evidence

Several studies have tested the validity of the neoclassical convergence hypothesis across countries, as well as across regions within a country (for surveys of this literature see Islam, 2003; Magrini, 2004; Durlauf, Johnson, & Temple, 2005). The bulk of the earlier studies tested the convergence hypothesis in terms of GDP (income) per capita. In more recent years, however, several studies have shifted attention to testing the hypothesis in terms of wages, especially across regions within a country. While a number of factors justify this shift, two of them can be considered to be most relevant.

First, the neoclassical growth models from which the convergence hypothesis derives are based on production functions, whose implications relate more closely to wages, or GDP per worker, than GDP per capita (Durlauf et al. 2005). Second, although GDP per capita is an important measure of living standards, it does do not perfectly reflect differences in living standards across regions (Juessen, 2009). The literature acknowledges that wages tend to be a better indicator of the economic well-being of the vast majority in any society (Williamson, 1998; Zaman & Goschin, 2014). This is expected, given that wages are the major contributor to total household income, accounting for 50 to 80 percent of income for households with at least one member of working age (see Global Wage Report, 2014/15). This section provides a brief overview of regional studies that test the convergence hypothesis in terms of wages.

Overview of International studies

A growing number of studies provide comprehensive empirical evidence on the convergence dynamics of wages across regions in both developed and developing countries. Table 4.2A and

4.3A in Appendix 4 summarise some of the most significant studies. Organised by the approach used to measure convergence, the following are the key findings from this research.

The first strand of the empirical literature employs the distributional dynamics approach. Measuring convergence with stochastic kernel density (Maza & Villaverde, 2006) and Markov chain (Webber, 2001) indicators, these studies find evidence of wage convergence across regions in Spain and regions within EU countries, respectively.

A second strand of the literature analysing σ-convergence reveals two key findings. The first finds that wages vary significantly across regions in both developed and developing countries. However, looking at the most recent year of data in these studies, the reported σ coefficients are much higher for developing countries, such as Romania (Zaman & Goschin, 2014); China (Huang & Chand, 2015); Brazil (Estanislau, et al. 2013) and India (Collins, 1999) than developed countries like Spain (Rosés & Sánchez-Alonso, 2004; Maza & Villaverde, 2006); Poland (Ferens, 2015); United States (Carlino & Mills, 1996; Bukenya, Davis, Banerjee, & Gyawali, 2011); and Sweden (Enflo, Lundh, & Prado., 2014). This suggests that wage dispersions are much wider across regions in developing countries than in developed countries.

The second finding from this literature is that in most developed countries the trends over time of the reported σ coefficients generally confirm evidence of σ-convergence (Carlino & Mills, 1996; Maza & Villaverde, 2006; Bukenya et al. 2011; Enflo et al. 2014; Ferens, 2015). By contrast, the trends in developing countries show evidence σ-convergence (Vakulenko, 2013; Estanislau et al. 2013) and σ-divergence (Zaman & Goschin, 2014; Goschin, 2015) or both σ-convergence and σ-divergence (Collins, 1999; Rosés & Sánchez-Alonso, 2004; Huang & Chand, 2015).

A final finding from this literature is that wages for different workers across regions in the same country display different convergence trends over time. This is confirmed by Rosés & Sánchez-Alonso (2004) for Spain and Collins (1999) for India. Rosés & Sánchez-Alonso (2004) find evidence of σ-divergence for agrarian and unskilled urban workers and evidence of σ-convergence for skilled industrial workers. Collins (1999) provides evidence in support of σ-convergence for horse-keepers and postman and evidence of σ-divergence for agricultural workers.

The third strand of the literature analysing β-convergence provides a number of insights on regional wage convergence. Firstly, the literature provides mixed findings, affected in part by the statistical methods employed (cross-sectional, panel data and time- series growth regression models). For example, with the exception of Moazzami (1997), studies utilising time - series models tend to reject the hypothesis of regional wage convergence (Carlino & Mills, 1996; Bukenya et al. 2011; Chen et al. 2016). In contrast, the majority of studies using cross-section (Tavernier & Temel, 1997; Rosés & Sánchez-Alonso, 2004; Maza & Villaverde, 2006; Šlander & Ogorevc, 2010; Ferens, 2015) and panel data (Enflo et al. 2014; Vakulenko, 2016; Naz et al. 2017) models provide evidence in support of the hypothesis of regional wage convergence. This finding is common across studies in both developed and developing countries.

The second finding from this literature is that panel data studies generally reveal higher convergence rates than cross-section studies. For example, for panel data studies the highest reported convergence rate is 27.5% per year (Enflo et al. 2014), while for cross-section studies the highest is 7.2% per year (Rosés & Sánchez-Alonso, 2004). Notwithstanding this evidence, a general finding from both cross-section and panel data studies is that the rate of convergence can vary significantly across regions in different countries. The rate also varies depending on whether convergence is unconditional or conditional, with conditional convergence rates higher than unconditional convergence rates in all cases. For instance, Maza & Villaverde (2006) find evidence of an unconditional convergence rate of 4.1% per year and a conditional convergence rate of 5.3% per year. The findings by Enflo et al. (2014) reveal an unconditional convergence rate of 13.5% per year and a conditional convergence rate of 27.5% per year.

Based on these studies, conditional convergence rates are in all cases higher than unconditional convergence rates. This finding is important, as it points to the existence of significant differences in region-specific characteristics, which in turn play a pivotal role in promoting or hampering regional economic growth, and therefore regional wage convergence. This literature highlights the importance of such factors as human and physical capital (Rosés & Sánchez-Alonso, 2004), labour productivity, unemployment and rural populations (Tavernier & Temel , 1997), capital intensity and production structure (Šlander & Ogorevc, 2010), average schooling and experience (Estanislau et al. 2013), migration, urbanization, industrialization and infant mortality (Enflo et al. 2014). Apart from these factors, other studies highlight the importance of spatial effects (Šlander & Ogorevc, 2010; Niebuhr, Granato, Haas, & Hamann, 2012; Vakulenko, 2016). While these factors play a critical role in explaining regional wage disparities and convergence, the lack of a standard set of factors shows that the existing literature is far from conclusive on the right set of conditioning factors.

It is evident that, while some of the studies find similar convergence trends to the three approaches for measuring convergence (Maza & Villaverde, 2006; Enflo et al. 2014; Ferens, 2015), some studies find contradictory trends even for the same country and dataset (Zaman & Goschin, 2014; Huang & Chand, 2015). The contradictory σ-convergence and β-convergence findings of Zaman & Goschin (2014) and Huang & Chand (2015) confirm the arguments in the literature that β-convergence is a necessary but not sufficient condition for σ-convergence to occur (Barro & Sala-i-Martin, 1991; Gezici & Hewings, 2004; Young et al. 2008). That is, β-convergence may exist without leading to σ-convergence. These contradictory findings highlight the need for empirical research that uses different approaches to measure convergence to ensure that results are not contingent upon a specific approach.

Evidence from South Africa

There is a shortage of studies analysing regional wage convergence in developing countries, particularly in Africa. Such work has been impeded by the unavailability of regional wage data. The bulk of existing convergence studies in Africa have focused primarily on convergence dynamics of GDP per capita across countries (among them McCoskey, 2002; Cuñado & Gracia, 2006; Charles, Darne, & Hoarau, 2012; Solarin & Sahu, 2013; Asongu, 2014), rather than convergence dynamics of average wages across regions within a country.

There are few convergence studies focusing on regional wage convergence in South Africa. Rather, a number of existing studies have examined the convergence dynamics of GDP (income) per capita across regions in South Africa (among them Naudé & Krugell, 2003, 2005, 2006; Krugell, Koekemoer, & Allison, 2005; Bosker & Krugell, 2008; Naudé, Krugell, & Matthee, 2010; Bastos & Bottan, 2014). In line with the international literature, this research applies the three approaches for measuring convergence discussed above, namely σ-convergence, β- convergence, and the distributional dynamic approach.

Studies analysing σ- convergence and measuring dispersion with the standard deviation of log of income per capita, show changing trends in the estimates of σ over time. Naudé & Krugell (2003) find evidence of σ – convergence over the period 1990 – 1996 as σ decreased from 0.61 to 0.53 and σ – divergence between 1996 and 2000 as σ increased to 0.55. In another study, Naudé & Krugell (2005) find no evidence in support of σ – convergence for income per worker, as σ remained stable at 0.55 from 1998 to 2002. Based on this evidence, it can be concluded that the South African spatial economy experienced a decrease in regional dispersions of income per capita in the early to mid-1990s, followed by an increase in dispersions in the early

2000s. Studies employing a kernel density estimator and markov chain estimator find clear evidence of increasing dispersions in GDP per capita across regions between 1996 and 2004 (Krugell et al. 2005; Bosker & Krugell, 2008)[49].

Finally, studies analysing β-convergence find no support for unconditional convergence, but do find evidence in support of conditional convergence in GDP per capita across regions in South Africa (Naudé & Krugell, 2003; 2005, 2006). Of these studies, Naudé & Krugell (2003) use panel data regression models, and find evidence of slow conditional convergence over the period 1990 – 2000. Their results show that, conditional on initial human capital, local employment rate and distance from main markets and harbours, GDP per capita of initially poor regions were catching-up to those of their richer counterparts, at a rate of 1.2% per year. Utilising panel data regression models, Naudé & Krugell (2005, 2006) find no evidence of unconditional convergence, but do show conditional convergence of GDP per capita between 1998 and 2002. Their results show that convergence is conditional on access to internal markets, human capital, capital stock, and export propensity. In a related study and using cross regional regression models, Bastos & Bottan (2014) find strong evidence of spatial convergence of income per capita among marginalised communities (communities in former homelands, just-inside and just-outside the former homelands) with higher initial exposure to resource rents, between 1996 and 2011. Their results are robust to the inclusion of several community characteristics, like access to infrastructure, market access, proximity to main roads, and migration, among other factors.

While we learn a lot from regional convergence studies in South Africa, they have a number of shortcomings. First, while these studies focus on regional convergence dynamics of GDP (income) per capita, we are unaware of a study that examines regional convergence dynamics of wages. Secondly, with the exception of Bastos & Bottan (2014), the bulk of the studies rely on regional data covering the period 1990 to 2004 and therefore do not provide insights into more recent trends. Thirdly, although the study by Bastos & Bottan (2014) covers the more recent period (1996-2011), the study does not cover the whole country, as it focuses on communities in former homelands only. Finally, most of the studies utilise GDP per capita data drawn from the Regional Economic Explorer (REX) database compiled by Global Insight

[49] In their study, Bosker & Krugell (2008) conclude that the South African economy is characterised by substantial regional disparities in economic development. They further conclude that South Africa's disparities are much larger than those observed in Europe and the United States. They are also shown to be larger than those observed in countries that are at a similar stage of economic development, such as India, Brazil, and China.

Southern Africa. The main weakness of this database is that it is based on the reconciliation of data from many different sources across the spatial dimension (Naudé et al. 2010). Given that this regional database is not based on direct statistical inferences from survey results, there is room for error in the reconciliation process, as numerous assumptions and imputations are made.

This chapter moves beyond these limitations and extends the convergence literature by examining regional convergence dynamics of wages across regions in South Africa using data covering the whole country and a more recent period (1996 – 2011).

4.5. Empirical strategy

The empirical strategy used in this chapter combines three different but complementary approaches for measuring convergence, namely the distributional dynamic (kernel density estimator), σ-convergence, and β-convergence. We use these approaches to provide a descriptive and econometric analysis of the convergence hypothesis.

Firstly, we apply the distributional dynamic approach, where we use a kernel density estimator to derive univariate density functions of average wages for 354 regions for the years 1996, 2001 and 2011. From the large family of kernel density estimators, we utilise the Gaussian kernel with optimal bandwidth, according to Silverman's (1986) rule-of-thumb. Following Maza, Hierro, & Villaverde (2012), the kernel density estimator of a series Y at a point y is given by:

$$f(y) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{y-Y_{it}}{h}\right) \tag{1}$$

where n is the number of regions, $K(\bullet)$ is the Gaussian kernel function, h denotes the bandwidth parameter which determines the degree of smoothing and Y_{it} is the data point. To evaluate the convergence hypothesis, we focus on the external shape dynamics of the entire cross-section wage distribution given by equation (1). We concentrate on the changes in the spread and skewedness of the distribution over time. One advantage of using the kernel density estimator is that we do not have to assume any form of the density functions, as the densities are estimated from the data.

However, a major drawback of the kernel density estimator is that the analysis relies heavily on subjective judgement based on visual impressions. Thus, we complement the analysis with a study of σ-convergence, where we measure dispersion using the standard deviation of the logarithm of regional average wages given by:

$$\sigma_t = \sqrt{\frac{1}{n}[\sum_{i=1}^{n}(y_{it} - \overline{y_t}]^2} \tag{2}$$

where σ_t is the standard deviation in year t, y_{it} and $\overline{y_t}$ denotes the logarithm of real wages of region i at time t and its average level at time t and n is the number of regions. The higher the absolute value of σ_t, the greater the extent of regional wage dispersion. There is evidence of σ-convergence over a period t to $T + t$ years when $\sigma_{t+T} < \sigma_t$, while σ-divergence occurs when $\sigma_{t+T} > \sigma_t$. Thus, a fall in the standard deviation over time confirms evidence of regional wage convergence, while an increase points to regional wage divergence. An alternative measure for dispersion that we also use is the coefficient of variation, calculated as the standard deviation of regional wage relative to average wage of all the regions ($\sigma_t/\bar{y}_t$).

To consolidate the analysis, we reinforce the descriptive work with an analysis β – convergence. An appealing feature of a study of β - convergence is that it allows for explicit calculation of the rate at which regional economies converge or diverge over time (Alexiadis, 2010). Even more important, it reveals the key factors influencing regional wage convergence or divergence. These can be targeted by policy initiatives aimed at stimulating regional economic development and promoting regional equalisation of wages and living standards.

In the analysis, we first test for evidence of unconditional β – convergence by estimating a growth-initial wage regression, where initial regional average wage is used as the only explanatory variable. Following Barro & Sala-i-Martin (1991), we specify the model as follows:

$$\frac{1}{T}\ln(y_{it}/y_{i0}) = \alpha + \beta\ln(y_{i0}) + \varepsilon_{it} \tag{3}$$

where y_{i0} and y_{it} are the initial and final level of average wage in region i. $\frac{1}{T}\ln(y_{it}/y_{i0})$ is the yearly growth in average wage in region i during the T year period under study. α is a constant term representing the steady-state point of convergence that is assumed to be the same for all regions. ε_{it} is the error term which is assumed to be independently and identically distributed. β is the regression coefficient to be estimated.

In addition, we test for evidence of conditional β-convergence by augmenting equation (3) with regional specific factors that potentially explain the convergence process, and therefore regional wage growth. Following Mankiw, Romer, & Weil. (1992), we estimate the following model:

$$\frac{1}{T}\ln(y_{it}/y_{i0}) = \alpha + \beta\ln(y_{i0}) + \sum_{n=1}^{N}\gamma_i X_{i0} + \varepsilon_{it} \qquad (4)$$

Where all the other variables are as defined in equation (3). X_{it} is a vector of regional controls incorporated to act as proxies for each region's steady state, that is, determinants for regional average growth rate of wages. The variables to be considered as controls will be discussed in the data section[50]. γ_i is the regression coefficient to be estimated.

In estimating equations (3) and (4), we are interested in the sign, magnitude, and significance of β, a coefficient that enables us to evaluate whether the dynamics reveal evidence of wage convergence or divergence. Thus, evidence of both unconditional and conditional β – convergence is confirmed by a negative and significant coefficient on initial regional average wage (β). A negative β coefficient implies that regions with initially low average wages grow faster than those with initially high average wages. It validates the β-convergence hypothesis. Interestingly, when the β-convergence hypothesis holds, the annual rate of convergence (b) and the half-life (h), which is the time needed to reduce the gap in variation in average wages between rich and poor regions by half, can be inferred from the estimated β parameter, using the relationship β=$-(1 - e^{-bT})/T$ (see Barro & Sala-i-Martin, 2004). The resulting b and h parameters are given by:

$$b = -\frac{log(1+\beta T)}{T} \quad \text{(a)}, \qquad h = \frac{\log(2)}{b} \quad \text{(b)} \qquad (5)$$

A larger β-coefficient in absolute terms (more negative) implies a greater speed of convergence, which in turn indicates fewer years necessary to reduce the regional wage gap by half.

4.6. Data description

This study draws on the geographically consistent dataset we construct in chapter 3. For this dataset, we use South Africa's full population census data for the years 1996, 2001, and 2011. In this dataset, data is aggregated to 354 magisterial districts (hereinafter – regions). As

[50] Exclusion of Solow growth model variables such as investment and capital is influenced mainly by unavailability of regional data of these key variables.

explained in chapter 3, regional income per worker is used to proxy regional wage per worker. To eliminate the influence of inflation on income, we convert nominal income for 2001 and 2011 to its 1996 real income equivalent, using the national consumer price index (CPI) provided by Stats SA[51]. Thus, for kernel density and σ-convergence we use regional real income per worker as the key variable of analysis, while for β-convergence analysis we use regional average yearly growth rate of real income per worker.

To test for conditional β-convergence, we include several variables grouped into 6 categories. The share of workers (of the total working age population) in each region who have tertiary education is used to capture regional differences in human capital (skilled workers)[52]. To account for the composition effects of workers across regions, we include the share of workers in the public sector. The regional unemployment rate is incorporated, to capture the effects of local labour market conditions, such as labour market rigidities. To account for the influence of regional industrial structure, we include the share of agricultural and manufacturing sector workers in each region. Population density and market potential are used to capture agglomeration effects. Population density is given as total regional population relative to total regional area. Market potential is based on the Harris (1954) index given by:

$$MP_{it} = \sum_{r=1}^{R} Y_{rt} \, (d_{ir})^{-1} \tag{6}$$

where Y_r is the total personal income for each region and d_{ir} is the bilateral distance between region i and r, calculated as the great-circle distance (in kilometres) between two points[53]. To account for internal distance for each region, we follow Redding & Venables (2004) and model internal distance of each region as $d_{ii} = \frac{2}{3}\sqrt{area_i/\pi}$, where $area_i$ is the size of a given region expressed in square kilometres.

In addition, the share of each region's area that falls in a former homeland area is included to differentiate regions in the former homeland and non-homeland areas. This variable is used to capture the effects of historical events, in particular, the apartheid-era system. This information

[51] The CPI values are 38.5 for census 1996, 52.4 for census 2001, and 92.6 for census 2011 (Stats SA). http://www.statssa.gov.za/publications/P0141/CPIHistory.pdf?

[52] The definition of human capital is subject to debate. We therefore check the robustness of our definition using different cut-off levels for education, such as considering all workers with at least a matric qualification (12 years of schooling). Regardless of the definition used, the importance of human capital remains evident in our analysis.

[53] To calculate distance, we first use Geographic Information Systems (GIS) data (shapefiles) provided by Stats SA with the census data to derive centroids (central latitudes and longitudes points) for each region, using the ArcGIS software programme. Using the centroids for each region, we then employ the 'haversine' formula, and calculate distance as the great-circle distance (in kilometres) between two points, region i and r.

is not readily available. We, therefore, use ArcGIS overlay tools to map magisterial district boundaries to former homeland boundaries[54]. From the resulting mapping, we create a ratio based on the union area of the two boundaries, which ranges between 0 and 1. A ratio of 1 indicates that a given region falls completely in a former homeland area, while a ratio of 0 denotes that the region was not part of a former homeland.

Finally, the share of mining sector workers in each region is used as a proxy for valuable mineral resource endowments. In addition, we obtain temperature and rainfall variables from the climatic data produced by Harris, Jones, Osborn, & Lister. (2014) from the Climatic Research Unit (CRU) at the University of East Anglia, to capture differences in climatic conditions. The climate data combines data from more than 4000 weather stations around the world and satellite data, to get estimates of monthly average temperature and rainfall over the period 1901-2016. The advantage of this database is that it is provided at fine spatial resolution (0.5x0.5 degree) grids which allows us to get climate estimates disaggregated at the magisterial district level. Combining this data with the centroid (latitudes and longitudes) information of each magisterial district obtained from the shapefiles provided by Stats SA, we obtain the average yearly temperature and rainfall for each magisterial district, by adding the monthly rates and dividing them by the number of months in a year. We use these variables to capture effects of local amenity conditions.

4.7. Descriptive analysis of the convergence dynamics of real income per worker

To give an initial indication of the dispersion of our key variable, Table 4.1 presents the summary statistics[55] of real income per worker (hereinafter – income per worker) across regions for the years 1996, 2001, and 2011. The data shows, first, that on average, income per worker across regions increased from 1553.13 Rands in 1996 to 1953.40 in 2001, then 2271.58 in 2011. Second, as shown by the Min and Max statistics, as well as the ratio between these statistics (Max/Min), the data shows that the regional average figure masks significant heterogeneity across regions. For instance, income per worker in the richest region (Max) is 6.68, 13.73 and 8.22 times that of the poorest region in 1996, 2001, and 2011 respectively.

To examine the extent of the observed dispersion in the data, we plot the kernel density estimates of the entire cross-section distribution of income per worker in the years 1996, 2001, and 2011, in Figure 4.1. For easy comparison, the horizontal axis shows income per worker for

[54] For a detailed explanation of the ArcGIS mapping exercise, see chapter 3, appendix 3.1.
[55] The summary statistics of the other variables are presented in Table 4.1A in appendix 4.

each region compared to national income per worker (measured in log). The resulting relative income per worker distribution shows the extent to which workers' income in each region deviates from the national average. Values equal to one (=1) indicate that income per worker of a given region is equivalent to the national average. On the other hand, values less than one (<1) indicate that income per worker of a given region is less than the national average, while values greater than one (>1) indicate that it is greater than the national average. The vertical axis gives the density of regions at different relative income per worker.

Table 4.1: Summary statistics for monthly regional income per worker (Rands).

Year	1996	2001	2011
Regional Average	1553.13	1953.40	2271.58
Std Dev	608.48	964.26	961.09
Min	742.73	599.17	1007.04
Max	4959.92	8229.52	8279.37
Max/Min	6.68	13.73	8.22

Notes: To eliminate the influence of inflation on income, we convert nominal income for 2001 and 2011 to their 1996 real income equivalent using the national consumer price index (CPI) provided by Stats SA.

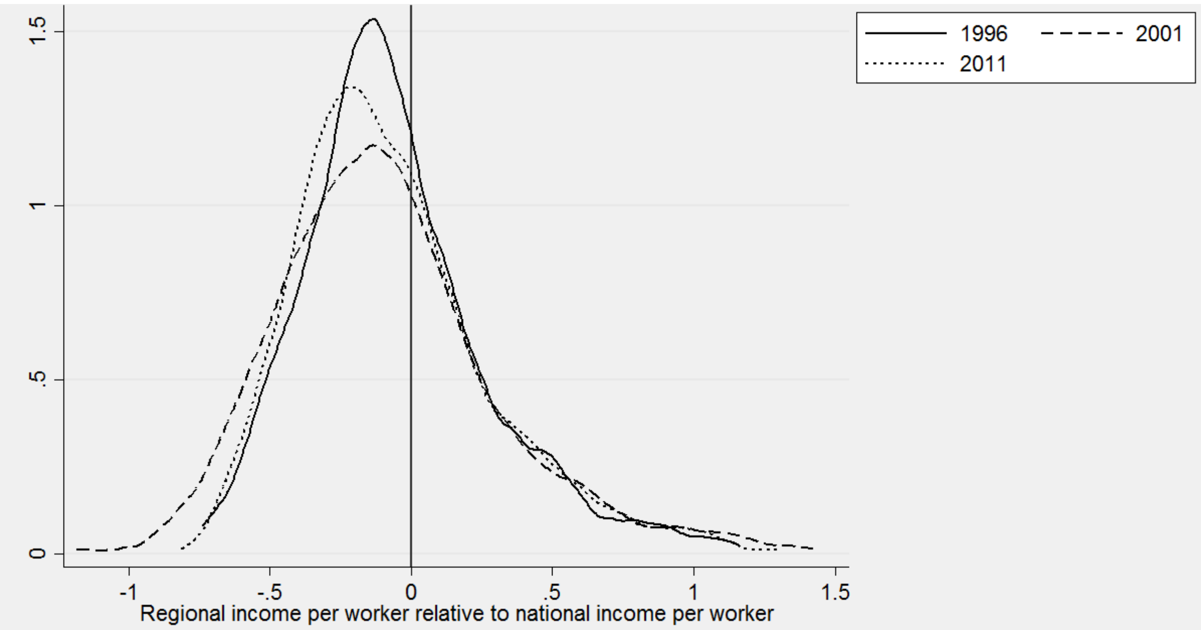
Three key features are evident from the kernel density estimates. The first feature is the positive skewedness of the distributions over the entire study period, showing that the bulk of the regions are near the bottom part of the distribution. This suggests that the country is characterised by many regions which are poorer than the national average and few regions with income per worker far above the national average[56]. This result is consistent with findings by Krugell et al. (2005) and Bosker & Krugell (2008) who examine regional disparities in GDP per capita in South Africa.

The second striking feature is the spreading out of the distributions over the entire study period. This suggests that the South African economy is characterised by substantial dispersions in the distribution of income per worker across regions between 1996 and 2011. However, the shape of the distributions show that the 1996 distribution is narrower, more peaked, and more concentrated than the 2001 distribution. The 2011 distribution is also narrower, more peaked and more concentrated than the 2001 distribution. This indicates wider dispersions in levels of income per worker in 2001 and 2011 than in 1996. The evidence also shows wide dispersions in income per worker in 2001 compared to 2011. These changes suggest an increase in regional disparities in income per worker between 1996 and 2011 that mask two contradicting trends:

[56] Incomes in most regions are concentrated around 0, which is the national average. Of the 354 regions in our sample, 42 in 1996, 35 in 2001 and 39 in 2011 incomes per worker above the national average.

An increase in regional disparities between 1996 and 2001, which was followed by a decrease in regional disparities from 2001 to 2011[57].

Figure 4.1: Distribution of relative income per worker across regions.



Notes: The kernel densities are based on real income per worker across the 354 magisterial districts.

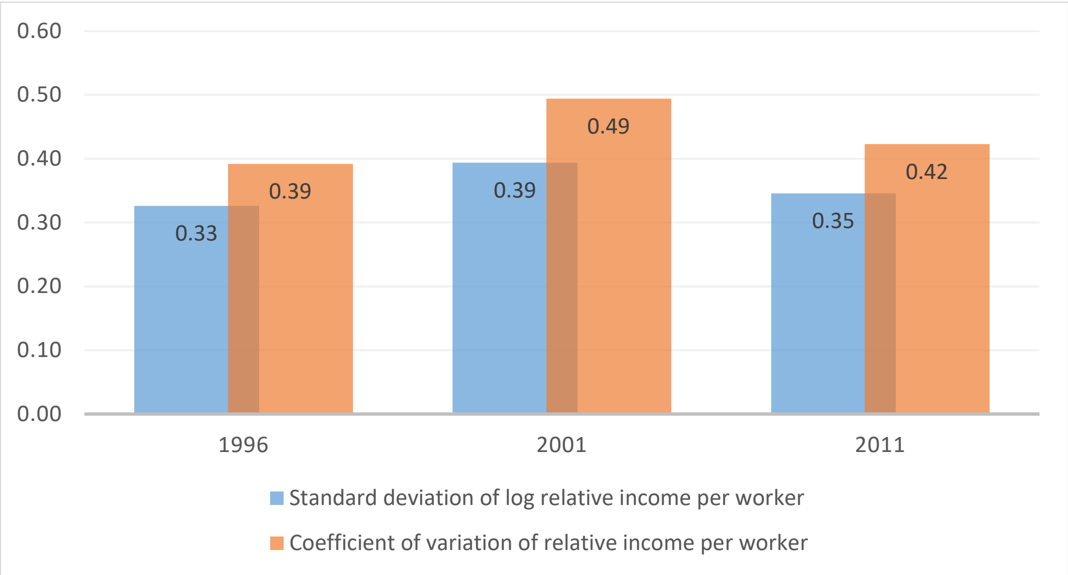
The final feature relates to shifts in the distributions over time. While the upper parts of the 2001 and 2011 distributions shifted slightly over time, the lower parts of these distributions shifted significantly. It is evident that the 2001 distribution shifted to the left of the 1996 distribution, indicating that poor regions became poorer in terms of income per worker between 1996 and 2001. On the other hand, the 2011 distribution shifted to the right of the 2001 distribution (but not as far as the 1996 distribution), confirming that poor regions became richer during the period 2001 to 2011.

To consolidate the descriptive analysis, and further illustrate the convergence dynamics of income per worker across regions in South Africa, we plot the results for σ-convergence analysis, in Figure 4.2. This figure displays the standard deviation of log relative income per

[57] While we used the mid-point approach to derive a continuous income measure and also dropped employed individuals with missing and zero income, it's interesting to note that our findings are consistent with earlier findings by Wittenberg (2017). In his analysis, Wittenberg (2017) finds that the Gini coefficient of wage inequality increased over the period 1994–2011. It is also evident in Figure 1 of Wittenberg's (2017) analysis that the increase in wage inequality between 1994 and 2011 masks two contradicting trends: An increase in wage inequality between 1994 and 2002, which was followed by a decrease in wage inequality from 2002 to 2011. This evidence is consistently reviewed across different imputation methods (including the mid-point approach) used to account for bracketed reporting on wages and the influence of zeros and missing wage data.

worker, and the coefficient of variation of relative income per worker across regions in 1996, 2001, and 2011. The results show unmistakable evidence of an increase in regional dispersions in levels of income per worker over time. This is shown by the increase in the standard deviation from 0.33 in 1996 to 0.39 in 2011 (a 6.13 % increase) and an increase in the coefficient of variation from 0.39 in 1996 to 0.42 in 2011 (a 7.91 % increase). Thus, both indicators confirm evidence of σ-divergence, as the relative income per worker distribution has become more dispersed over time. We further characterise the dispersion of income per worker across regions, after splitting the study period into two subperiods. This reveals that the standard deviation and coefficient of variation increased between 1996 and 2001, before decreasing over the 2001 - 2011 period. This is evidence of σ-divergence between 1996 and 2001, followed by σ - convergence from 2001 to 2011.

Figure 4.2: Cross-section dispersion of relative income per worker across regions.



Source: Author's calculations based on income per worker across the 354 magisterial districts.

Overall, the results in this section consistently show evidence of regional divergence in levels of income per worker from 1996 to 2001, followed by regional convergence from 2001 to 2011. The evidence of regional divergence in income per worker over the period 1996-2001 is consistent with findings by Naudé & Krugell (2003) who focus on convergence dynamics of GDP per capita across regions in South Africa. However, our evidence of regional convergence in incomes for workers from 2001 to 2011 is a new finding. It shows that the more recent years, the South African economy has demonstrated regional convergence in terms of income per worker. However, it's important to note that regional dispersions in income per worker remain very large in South Africa. The extent of these dispersions is much larger than in other countries (both developed and developing countries). For instance, at the end of the study period (2011), the coefficient of variation is 0.42. It is 0.15 in Romania in the same year (Zaman & Goschin, 2014), 0.11 in Poland in 2013 (Ferens, 2015), 0.23 in Brazil in 2009 (Estanislau et al. 2013) and 0.27 in China in 2010 (Huang & Chand, 2015). However, differences in economic indicators used to measure dispersions, the number of regions under study, and time-period of analyses makes comparison with other countries difficult.

4.8. Econometric analysis of the convergence dynamics of income per worker

The descriptive analysis discussed in the previous section raises interesting questions about regional convergence dynamics of income per worker. The first question relates to the magnitude and statistical significance of the observed dynamics of income per worker across regions in South Africa. The second question relates to the potential drivers of the observed dynamics. To provide insights into these issues, in this section we present cross-section estimates from β-convergence analysis based on equations (3) and (4) for the entire study period (1996-2011) and then for the two sub-periods (1996-2001 and 2001-2011). We estimate these models using ordinary least squares (OLS) method.

One concern in estimating these models is the potential bias due to reverse causality. However, using initial period variables as explanatory factors reduces this problem. Further, after confirming evidence of heteroscedasticity across regions using the White and Breusch-Godfrey tests, all standard errors are corrected for heteroscedasticity. We present the results in the next sections.

Unconditional β-convergence analysis

As a starting point, we test for evidence of unconditional β-convergence by estimating equation (3), where we examine the association between regional average growth rate and initial income per worker:

$$\frac{1}{T}\ln(y_{it}/y_{it-1}) = \alpha + \beta\ln(y_{it-1}) + \varepsilon_{it} \tag{7}$$

The hypothesis we test is that initial log of income per worker is negative and statistically significant. We, therefore, expect $\beta < 0$.

Table 4.2 presents the estimates (β) for the log of initial income per worker for the entire study period (1996-2011) and then for the two subperiods (1996-2001 and 2001-2011) based on equation (7). In the first specification (1996-2011), the estimate for the log of initial income per worker is negative and statistically significant, suggesting evidence of unconditional β-convergence across regions. The revealed β-estimate of 0.01 suggests a convergence rate of 1.07% per year, which means that it could take 65 years to reduce the gap in levels of income per worker between rich and poor regions by half. While the evidence of unconditional convergence is consistent with the theoretical and empirical literature, the convergence rate is far below the so-called "iron law of convergence", where the gap among economies is eliminated at a rate of around 2% per year (see Barro, 2015). The apparent evidence of unconditional β-convergence in income per worker across regions contradicts our earlier findings from the descriptive analysis of regional divergence in levels of income per worker over the study period[58].

Table 4.2: Absolute $\boldsymbol{\beta}$-convergence test, OLS.

	1996-2011	1996-2001	2001-2011
VARIABLES	(1)	(2)	(3)
Log initial income per worker	-0.010***	0.011*	-0.031***
	(0.002)	(0.006)	(0.003)
Constant	0.108***	0.064	0.237***
	(0.019)	(0.053)	(0.027)
Observations	354	354	354
R-squared	0.052	0.009	0.245
F-test	18.58	3.162	103.9
Convergence rate (%)	1.07	Diverge	3.66
Half-life (years)	65	Diverge	19

Asterisks indicate the level of significance, where: *** p<0.01, ** p<0.05, * p<0.1 and the values in parentheses are heteroscedasticity-consistent standard errors. The dependent variable is the average annual growth rate of income per worker.

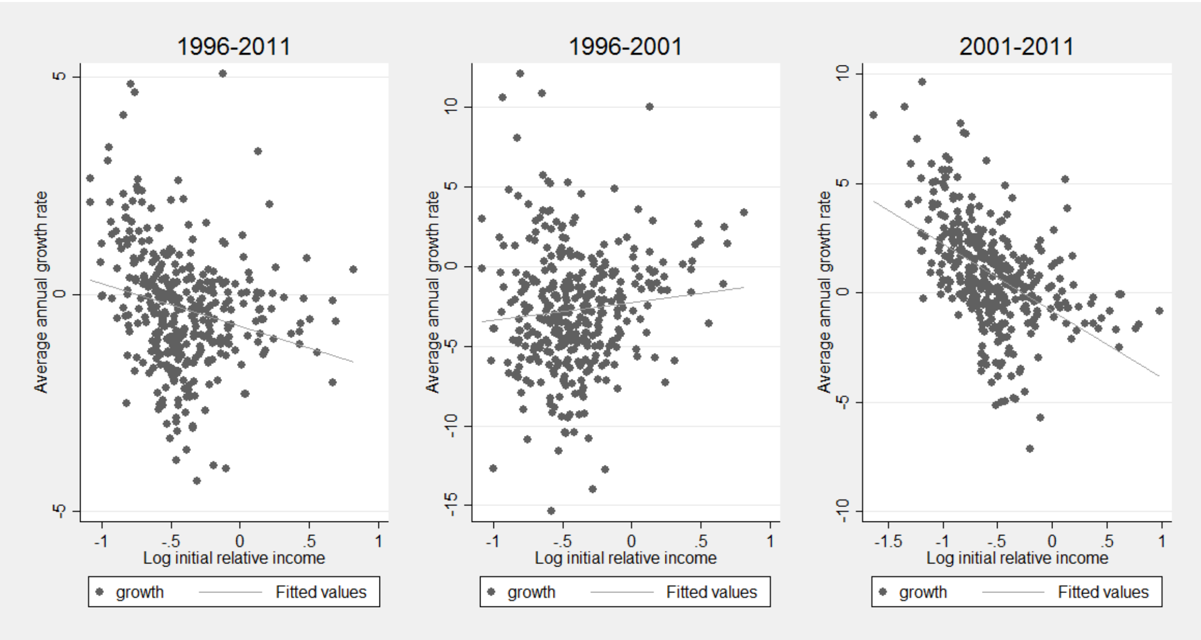
However, re-estimation of equation (7) separately for each sub-period confirms the convergence dynamics revealed in the descriptive analysis. From 1996 to 2001, the estimate of the log of initial income per worker is positive and marginally significant, indicating no evidence of unconditional β-convergence. This result supports findings by Naudé & Krugell (2003, 2005, 2006) of no evidence of unconditional β- convergence from 1996 to 2002 for regional GDP per capita. However, during the period 2001-2011, the estimate for the log of initial income per worker is negative and highly significant, confirming evidence of unconditional β-convergence. The revealed estimate of 0.031 suggests a regional convergence

[58] This result confirms that β-convergence is a necessary, but not sufficient condition for σ-convergence.

rate among of 3.66% per year, which means that it could take 19 years to reduce the gap in levels of income per worker across regions by half.

A graphical display of the above results is provided in a scatterplot (Figure 4.3) of the relationship between average annual growth rate and initial income per worker across regions. The negative slope of the fitted line for the 1996-2011 and 2001-2011 periods reaffirms evidence of unconditional β-convergence, while the positive slope of the fitted line for the 1996-2001 attests to the lack of evidence of unconditional β-convergence.

Figure 4.3: Association between average growth rate and initial income per worker.



Source: Author's calculations based on real income per worker data from censuses for the 354 regions.

Taken together, these results show evidence of unconditional convergence in levels of income per worker between 1996 and 2011 and reflect two trends: no evidence of unconditional convergence between 1996 and 2001, and evidence of unconditional convergence between 2001 and 2011. This suggests that richer regions tended to grow faster than poor ones over the 1996-2001 period, increasing regional dispersion in incomes for workers. From 2001 to 2011, however, poor regions grew faster than rich ones, thereby decreasing regional dispersion in workers' incomes over this period. The divergence and convergence trends in different periods are consistent with our earlier findings from the descriptive analysis based on kernel density and σ-convergence analysis[59]. However, the fit of the estimated models is very poor as indicated by the low R squared values in all columns.

Conditional β-convergence analysis

Evidence of lack of unconditional β-convergence across regions over the 1996-2001 sub-period, and evidence of unconditional β-convergence across regions over the 1996-2011 period and 2001-2011 sub-periods seems to be well-established. However, one fundamental concern with unconditional β-convergence approach is that omitted variables might cause bias in the model estimates. In this sub-section, we elaborate on how we test for conditional β-convergence that enables us to address the potential problem of omitted variables affecting unconditional β-convergence. We achieve this by controlling for various region-specific factors that potentially explain regional growth in workers' income other than initial income per worker. We do not control explicitly for the characteristics of the Solow- Swan (1954; 1954) growth model, from which the convergence hypothesis derives. However, we include regional characteristics based on economic theory and the empirical literature[60].

One compelling source of differences in growth of incomes of workers by region is differences in human capital, which determine the importance of worker's skills and the composition of workers across regions. The importance of human capital differences in driving regional economic growth finds support from the theoretical literature (Becker, 1962; Romer, 1986), and the empirical literature, which find a positive effect of human capital on levels and growth of income and wages across regions (see Ottaviano & Pinelli, 2011; Bai et al. 2015; Huang & Chand, 2015)[61].

Regional variation in local labour market conditions, such as the extent of unemployment, is another potential influencing factor. According to the regional wage curve theory, a negative association exists between the local unemployment rate and regional wage levels (Blanchflower & Oswald, 1990; Card, 1995) and this relationship has been confirmed empirically in different countries, including in South Africa (Magruder, 2012; Von Fintel, 2017). Another notable cause of differences in regional economic growth is variation in local amenities. According to local amenity theory (Roback, 1982, 1988), differences in climatic

[59] Unlike some international studies, we do not find mixed convergence trends from the different approaches of measuring convergence (see Rosés & Sánchez-Alonso, 2004; Zaman & Goschin, 2014).

[60] We exclude Solow growth model variables, such as investment and capital, because regional data is not available for these key variables.

[61] As discussed in the previous section, there is evidence of a positive effect of initial human capital on regional economic growth in South Africa (see Krugell & Naude, 2003, 2005, 2006).

conditions, natural resources, institutional quality, and cost of living have an effect on levels of productivity, one of the main driving forces behind regional wage and income disparities (Maza & Villaverde, 2006). Regions with favourable local amenities, such as valuable natural resources, good access to waterways, favourable climatic conditions, and well-developed institutions and infrastructure can be more productive, which in turn may raise levels of income per worker and therefore growth in these regions.

Differences in regional industrial structure may also lead to differences in regional economic growth rates, as industries are not evenly distributed across regions. Regions in urban areas may have higher growth rates than those in rural areas because of the former's advantages of industrial agglomeration and associated productivity gains (Fujita & Thisse, 2002). In contrast, wage levels may be lower for regions in rural areas as they lack economic forces to attract and hold economic activities. Another notable cause of uneven regional economic growth is local externalities because of spatial agglomeration of economic activities in given locations. Agglomeration of firms and consumers creates positive externalities, such as large markets. It can as also create negative externalities (such as high housing costs) that affect regional wages, income, and economic growth. Capturing agglomeration effects with such measures as population density and market potential, Crozet & Koenig (2005) find that market potential has a positive effect on regional economic growth, while Ottaviano & Pinelli (2006) find population density has a negative effect.

Uneven regional economic growth may also be explained by historical events such as, in South Africa the apartheid system, in place from 1948 to 1994. The apartheid regime implemented several racially segregatory policies, which forcefully relocated black South Africans to "homeland" areas ostensibly based on ethnicity. These areas were highly marginalised, overcrowded, and distant from economic centres. The advent of democracy in 1994 led to the end of the apartheid-era rule and the legal reintegration of all homeland areas into South Africa. In addition, numerous policies were implemented to promote regional economic development and address regional economic disparities created by years of apartheid-era rule. Despite the implementation of these policies, the legacy of apartheid-era racial segregatory policies may still shape regional wages and incomes. This claim finds support in a growing body of research that suggests that distant historical events have long-lasting effects that shape economic developments for long periods (Nunn, 2009; Acemoglu & Dell, 2010; Becker, Boeckh, Hainz, & Woessmann, 2016).

Taken together, region-specific factors may either promote or hamper regional convergence of income per worker over time. We examine their effects and, in the process, test for conditional β-convergence by augmenting equation (7) to capture the effects of human capital, local amenities, local labour market conditions, agglomeration effects, and historical events, as follows:

$$\frac{1}{T}\ln(y_{i,t+T}/y_{i,t}) = \alpha + \beta\ln(y_{i,t}) + \sum_{n=1}^{N}\alpha_i X_{i,t} + \varepsilon_{i,t} \quad (8)$$

where all the other variables are as defined in equation (4). $X_{i,t}$ is a vector of regional controls, which includes skilled workers, market potential, population density, temperature and rainfall, local unemployment rate, share of workers in the agricultural, manufacturing, mining, and public sectors, and share of each region in former homeland areas. Of these variables market potential, population density, temperature and rainfall are introduced in logarithms, while human capital, homeland status, local unemployment rate, share of workers in the agricultural, manufacturing, mining and public sectors are incorporated as shares. We expect skilled workers, market potential, share of manufacturing, public and mining sector workers to have a positive effect. A negative effect is expected for homeland status, population density, share of agricultural workers, and local unemployment rate. For temperature and rainfall, we expect either a positive or negative effect[62].

The estimation results from equation (8), for the entire study period and the two sub-periods, are presented in Table 4.3[63]. Notwithstanding the few variables that are insignificant in some columns, the inclusion of region-specific factors lead to significant improvements in the fit of the estimated model. This is shown in all columns by higher values of the R squared statistic. In addition, the estimate of interest, log of initial income per worker is negative and statistically significant in all three periods. This suggests that, after accounting for the effects of initial

[62] For instance, high population density and market potential can affect regional wage growth positively through their association with greater intensity of economic activities, driven by agglomeration economies (Fujita et al. 1999; Krugman, 1996), but higher densities and market potential can also have negative effects, through increased population pressures on scarce resources and agglomeration costs (Liu & Yamauchi, 2014).

[63] Given the important role that apartheid-era policies played in shaping the South African spatial economy, we examined whether including the homeland variable as the only control could change our initial results on regional convergence dynamics. We achieved this by re-estimating equation (8) with the homeland variable as the only control. To check the importance of other controls, we also included other controls individually. Our results in Tables 4.4A – 4.6A show that including the homeland variable together with initial income per worker as the only controls does not change our conclusions for both the unconditional and conditional models in all the 3-time periods (1996-2011, 1996-2001 and 2001-2011). Our analysis shows that the unconditional divergence over the period 1996 – 2001 is not a result of the peculiarity of the former homeland areas. Our conclusions also remain valid in relation to other controls.

regional conditions, we now find evidence of β-convergence even for 1996-2001. However, the magnitude of the estimate changes significantly once we account for region specific factors. The estimates in all columns, shown in Table 4.3, become much larger (in absolute terms) than those reported in Table 4.2.

The differences in the magnitude of the estimate of the log of initial income per worker in these Tables points to the existence of important structural factors that significantly affect the growth of real income per worker across regions. The bulk of the included regional factors are statistically significant, confirming that they play a vital role in determining the growth of income per worker across regions. The regression results show a positive and highly significant association between the share of skilled workers and the average growth rate of income per worker. This suggests that regions with a greater share of skilled workers tend on average to have higher growth of income per worker. This finding is in line with those from other regional studies in South Africa, such as Naudé & Krugell (2003, 2005, 2006), who find that regions with greater initial human capital tend to have a higher GDP per capita growth rate over the period 1990 – 2002[64].

Turning to measures of agglomeration effects. Market potential and population density have a statistically significant effect (except for column 2 for population density), indicating that these factors play a key role in determining regional growth of income per worker. Market potential has a positive effect, while population density has a negative effect. These results highlight the tension between various forces linked to agglomeration, which can either promote or hamper regional economic growth. From our results, and consistent with the existing literature, it is evident that greater market potential promotes growth of income per worker (see Ottaviano & Pinelli, 2006; Holl, 2012; Bai et al. 2012), while higher population density hampers growth of income per worker (see Naudé & Krugell, 2006).

Looking at the effects of local labour market conditions, we see that, while the estimate of the unemployment rate is not significant in column 1, it turns out significant in column 2 and 3. However, the estimate shows that unemployment has a mixed effect over time. It is positively associated with the growth of income per worker over the period 1996-2011 and negatively associated over the period 2001-2011. Highlighting the importance of regional industrial

[64] The statistical significance of the human capital and share of workers in the agricultural sector variables also highlights the importance of regional composition of workers. However, share of workers in the public sector turns out positive but insignificant in all columns. As a result, we choose to drop the public-sector variable.

structure, the estimates of the share of workers in the agricultural sector is negative and significant in all columns. This suggests that a higher share of agricultural workers is associated with a lower growth of income per worker. The share of workers in the manufacturing sector is negative and significant (only column 1).

Table 4.3: Conditional β-convergence test, (OLS).

Period	1996-2011	1996-2001	2001-2011
Log real income per worker	-0.040***	-0.052***	-0.073***
	(0.004)	(0.014)	(0.004)
Regional specific factors			
Human capital			
Skilled workers (%)	0.343***	0.851***	0.377***
	(0.066)	(0.219)	(0.053)
Agglomeration effects			
Log market potential	0.003***	0.007**	0.005***
	(0.001)	(0.003)	(0.001)
Log population density	-0.002***	-0.001	-0.004***
	(0.001)	(0.002)	(0.001)
Labour market conditions			
Unemployment rate (%)	-0.012	0.082***	-0.023**
	(0.008)	(0.026)	(0.012)
Industrial structure			
Share of agricultural workers (%)	-0.045***	-0.070***	-0.051***
	(0.007)	(0.024)	(0.012)
Share of manufacturing workers (%)	-0.029**	-0.002	0.002
	(0.014)	(0.046)	(0.021)
Local amenities			
Log average rainfall	0.008***	-0.000	0.014***
	(0.002)	(0.005)	(0.003)
Log average temperature	0.005	0.018	0.002
	(0.005)	(0.018)	(0.008)
Share of mining workers (%)	-0.008	0.007	0.015
	(0.007)	(0.023)	(0.013)
Historical events			
Share of area in homelands (%)	-0.017***	-0.038***	-0.018***
	(0.003)	(0.009)	(0.004)
Constant	0.242***	0.220**	0.421***
	(0.031)	(0.102)	(0.043)
Observations	354	354	354
R-squared	0.410	0.124	0.586
F-test	19.75	4.029	40.28
Convergence rate (%)	6.08	6.08	13.04
Half-life (years)	11	11	5

Asterisks indicate the level of significance, where: *** p<0.01, ** p<0.05, * p<0.1 and the values in parentheses are heteroscedasticity-consistent standard errors. The dependent variable is the average annual growth rate of income per worker.

Finally, the proxy for historical events, the share of each region in former homeland areas is negative and highly significant. This means that being in a former homeland is associated with a lower growth rate of income per worker. This implies that the legacy of apartheid-era spatial planning and institutional policies continues to negatively affect regional economic development in post-apartheid South Africa, 15 years after its abolishment (15 years in line with our data). This result is consistent with Becker et al. (2011) who find that historical events leave a legacy that continues to shape current regional economic performance through cultural norms, values, beliefs, and formal institutions. Given that homeland areas are more likely to have more unemployment and workers in the agricultural sector, the cumulative effect of a region being in former homeland area is likely to be large.

Accounting for region-specific factors increases the rate of convergence from between 1.07% and 3.66% per year (in Table 4.2) to between 6.08% and 13.07% per year (in Table 4.4). Accordingly, the number of years required to reduce the gap in income per worker between regions by half decreases from between 19 and 65 years to between 5 and 11 years[65]. The conditional convergence rates are in line with the findings from existing empirical studies in other countries, which show a large variability in convergence rates. However, South Africa's maximum conditional rate of convergence of 13% per year is much higher than that of other countries. For example, the United States has a maximum rate of 8% (Tavernier & Temel, 1997), Romania - 5.4% (Zaman & Goschin, 2014), Spain - 10.5% (Rosés & Sánchez-Alonso, 2004; Maza & Villaverde, 2006), India - 2.7% (Collins, 1999), and Brazil 4.2% (Estanislau et al. 2013). This highlights that there are important structural differences across regions in South Africa that play a significant role in explaining regional disparities and growth of incomes of workers. Thus, to promote economic growth and reduce workers' income disparities across regions, it is necessary to pay more attention to the quality of human capital, access to markets, population density, unemployment, the agricultural sector, and economic conditions of regions in former homeland areas.

[65] This implies that rather than 19 years if differences in regional specific factors had been eliminated, it could have taken about 5 years to reduce the worker income gap between rich and poor regions by half between 2001 and 2011.

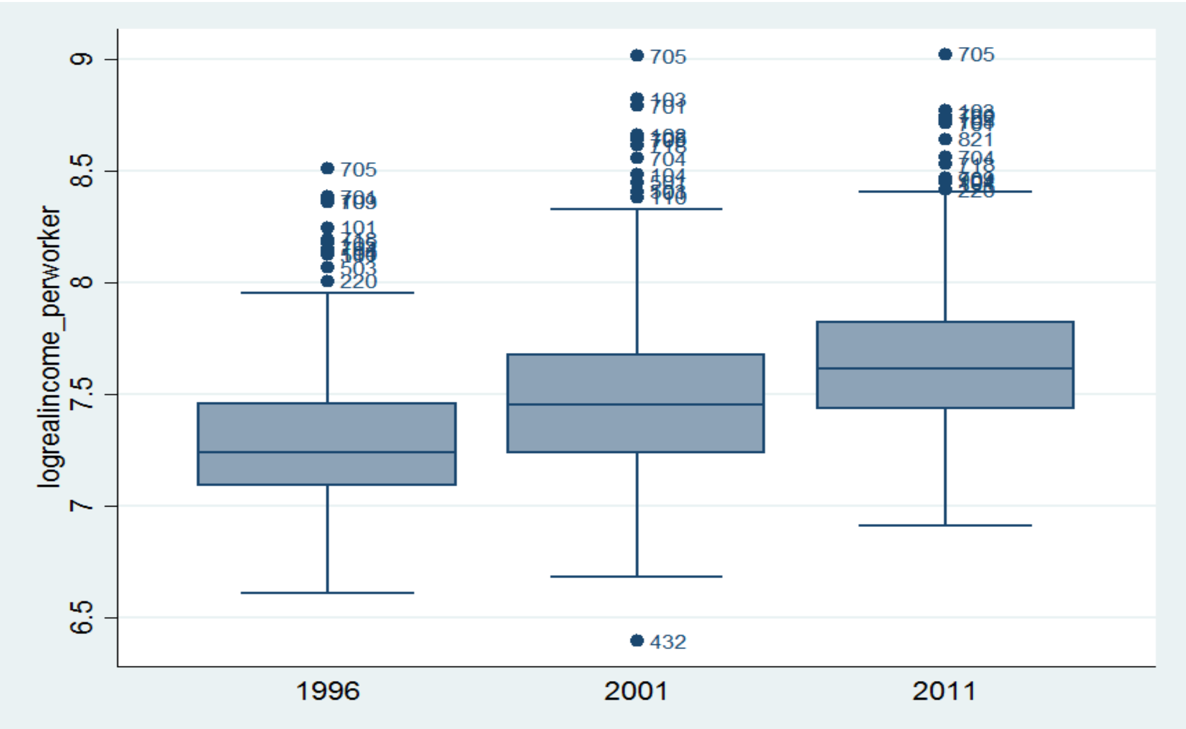
4.8.1. Robustness tests

There are four major concerns with the results discussed in the previous section. These include the influence of outlier regions, alternative samples, alternative measures of regional economic disparity, and alternative econometric models. This section presents robustness and sensitivity tests we carry out on the results reported in Table 4.2 and 4.3 to assess the potential bias due to these issues.

Influence of outlier regions

The first concern is that our results might be driven by outlier regions with extremely high or low incomes of workers. To identify whether we have outlier regions in our dataset with respect to levels of income per worker, we use box plots. Specifically, we plot the box plots of log real income per worker in 1996, 2001 and 2011 in Figure 4.4. As shown, the box plots consist of three main elements: the box, the whiskers and the extreme values, where the box shows the lower quartile and upper quartile *(*the *interquartile range* or *IQR)* values of income per worker, while the horizontal line within the box represents the median value of income per worker. The lines on the end of the lower and upper whiskers capture the lowest and highest values, while points below the lower and above the upper values reflect outlier regions with extremely low or high levels of income per worker.

Figure 4.4: Distribution of income per worker across regions (1996 - 2011).



Note: The boxplots are based on income per worker data derived from 1996, 2001 and 2011 census for a cross-section of 354 magisterial districts.

To identify these outlier regions, we calculate the lower and upper values of regional income per worker using the following formulas: lower value= (upper percentile quartile - 1.5IQR) and upper value = (upper percentile quartile + 1.5IQR). Using these formulas, we find that the lower value for 1996, 2001 and 2011 is log 6.6, 6.6 and 6.9, while the upper value for the same time period is log 8, 8.3 and 8.4. Based on these values, we see that while there is only one outlier region below the lower value in 2001, hence extremely poor, there are a number of outlier regions that are above the upper value in all the years (13 regions in 1996 and 2001, and 14 in 2011), hence extremely rich. These extremely rich regions are located mainly in Gauteng and Western Cape and might be driving our results.

With this concern in mind, the robustness of the β-convergence results in the previous section is tested by re-estimating equation (7) and (8), excluding outlier regions. The results are presented in Table 4.4. Notwithstanding the slight change in the magnitude of the estimates of the log of initial income per worker, and some controls, that turn out insignificant, the results continue to reveal no evidence of unconditional β-convergence but conditional β-convergence among regions over the 1996-2001 period. The analysis also continues to show robust evidence of both unconditional and conditional β-convergence over the 1996-2011 and 2001-2011 periods. Comparing the results, the estimates of convergence rates increased with the removal of outliers from between 1.07% and 3.66% per year to between 1.3% and 5.3% per year for unconditional convergence and from between 6% and 13.2% to between 7.6% and 14.2% for conditional convergence.

Alternative sample: Unskilled and skilled workers

Another issue to consider is whether the convergence dynamics revealed in Table 4.2 and 4.3 differ across different groups of workers. We re-estimate equation (7) and (8) using a restricted sample of unskilled and skilled workers only. We define unskilled workers narrowly as workers with no schooling, and skilled workers as workers with a tertiary education. The results of the unconditional β-convergence tests are presented in Table 4.5, and those for the conditional β-convergence tests are reported in Table 4.7A in appendix 4[66].

[66] In Table 4.5, column (1) – (3) report estimates for unconditional β-convergence for unskilled workers for the period 1996 – 2011, 1996-2001 and 2001-2011, respectively, while column (4) – (6) present estimates for unconditional β-convergence for skilled workers over the same time-period. In Table 4.7A, column (1) – (3) report estimates for conditional β-convergence for unskilled workers for the period 1996 – 2011, 1996-2001 and 2001-2011, respectively, while column (4) – (6) present estimates for conditional β-convergence for skilled workers over the same time-period.

Table 4.4: β-convergence test excluding outlier regions, (OLS).

	Unconditional			Conditional		
Period	1996-2011	1996-2001	2001-2011	1996-2011	1996-2001	2001-2011
Log income per worker	-0.017***	-0.001	-0.041***	-0.046***	-0.063***	-0.076***
	(0.003)	(0.008)	(0.003)	(0.004)	(0.015)	(0.004)
Regional specific factors						
Skilled workers				0.429***	0.876***	0.369***
				(0.083)	(0.272)	(0.064)
Log market potential				0.003***	0.007**	0.005***
				(0.001)	(0.003)	(0.001)
Log population density				-0.003***	-0.004	-0.004***
				(0.001)	(0.002)	(0.001)
Unemployment rate				-0.004	0.090***	-0.019
				(0.008)	(0.027)	(0.012)
Share of agricultural workers				-0.044***	-0.078***	-0.051***
				(0.007)	(0.024)	(0.012)
Share of manufacturing workers				-0.028**	0.010	-0.003
				(0.013)	(0.047)	(0.020)
Log average rainfall				0.008***	0.002	0.012***
				(0.002)	(0.006)	(0.003)
Log average temperature				0.006	0.024	0.001
				(0.005)	(0.018)	(0.008)
Share of mining workers				-0.003	0.016	0.017
				(0.006)	(0.023)	(0.013)
Share of area in homelands				-0.017***	-0.037***	-0.018***
				(0.003)	(0.009)	(0.004)
Constant	0.146***	0.044	0.321***	0.266***	0.257**	0.437***
	(0.019)	(0.055)	(0.025)	(0.030)	(0.106)	(0.041)
Observations	337	340	337	337	340	337
R-squared	0.106	0.000	0.298	0.471	0.122	0.600
F-test	39.91	0.00987	142.5	26.26	4.138	44.28
Convergence rate (%)	1.9	-	5.3	7.8	7.6	14.2
Half-life (years)	36	-	13	9	9	5

Asterisks indicate the level of significance, where: *** p<0.01, ** p<0.05, * p<0.1 and the values in parentheses are heteroscedasticity-consistent standard errors. The dependent variable is the average annual growth of real income per worker. We test for unconditional convergence for the periods 1996-2011, 1996-2001 and 2001-2011 in column (1) - (3) respectively, while in columns (4) – (6), we test for conditional convergence for the same time periods.

Notwithstanding the changes in the magnitude of the estimates of the log of initial income per worker, the main conclusions reached in Tables 4.2 and 4.3 continue to hold. The results show no support for unconditional but conditional β-convergence among regions from 1996 to 2001. The analysis also shows evidence of both unconditional and conditional β-convergence over the 1996-2011 and 2001-2011 periods. The results, however, suggest that within both the unskilled and the skilled groups, convergence is higher than for the pooled group of workers.

For example, the unconditional convergence rate increased from between 1.07% and 3.66% per year to between 4.6% and 9.7% per year for unskilled workers, and to between 2.8% and 10.6% for skilled workers. On the other hand, the conditional convergence rate increased from between 6.1% and 13.0% per year to between 9.5% and 22.0% per year for unskilled workers and to between 5.7% and 22.5% for skilled workers. Market potential, unemployment, and being in former homeland are key drivers of conditional convergence for skilled workers. The share of agricultural workers and unemployment play a key role for unskilled workers. These results suggest that, differences in pay for the same skill level are being eliminated across regions much quickly. Moreover, because the speed of convergence is lower for the pooled group of workers than within the groups, convergence in worker's incomes across skills categories is not taking place to the same degree. These finding highlights the structural impediments (skills barriers) in the South African economy that sustain inequality within the labour market across skill categories.

Table 4.5: Unconditional β-convergence test for a restricted sample of workers, (OLS)

	Unskilled workers			Skilled workers		
VARIABLES	1996-2011	1996-2001	2001-2011	1996-2011	1996-2001	2001-2011
Log initial income per worker	-0.033***	-0.003	-0.062***	-0.023***	-0.011	-0.065***
	(0.003)	(0.007)	(0.003)	(0.003)	(0.013)	(0.003)
Constant	0.236***	0.053	0.426***	0.232***	0.157	0.603***
	(0.017)	(0.047)	(0.023)	(0.022)	(0.112)	(0.025)
Observations	354	354	354	354	354	354
R-squared	0.303	0.001	0.475	0.190	0.002	0.611
F-test	153.3	0.200	318.8	82.59	0.682	552
Convergence rate (%)	4.6	-	9.7	2.8	-	10.6
Half-life (years)	15	-	7	25	-	7

Asterisks indicate the level of significance, where: *** p<0.01, ** p<0.05, * p<0.1 and the values in parentheses are heteroscedasticity consistent standard errors. The dependent variable is the average annual growth rate of income per worker for unskilled and skilled workers.

Alternative measure of regional economic disparity

Another concern is that our baseline results might be driven by the measure of regional economic disparity (income per worker) employed in this study. As a robustness check, we re-estimate equations (7) and (8) using an alternative measure, real income per capita (hereafter – income per capita). We derive income per capita by dividing total personal income by the total population in each region[67]. Table 4.6 presents the results, where columns (1) – (3) report

[67] Income per worker is an important measure for regional disparity, as it closely relates to the economic theory from which the convergence hypothesis is derived. However, income per capita is also an important measure, as it relates closely to regional policy, which is mainly concerned with levelling out living standards across regions.

estimates for unconditional β-convergence over the period 1996 – 2011, 1996-2001 and 2001-2011 respectively, while columns (4) – (6) present estimates for conditional β-convergence over the same period.

Table 4.6: β-convergence test for income per capita across regions, (OLS).

	Unconditional			Conditional		
Period	1996-2011	1996-2001	2001-2011	1996-2011	1996-2001	2001-2011
Log income per capita	-0.015***	0.006*	-0.027***	-0.026***	-0.025**	-0.054***
	(0.001)	(0.004)	(0.001)	(0.003)	(0.013)	(0.004)
Regional specific factors						
Skilled workers				0.306***	0.682***	0.359***
				(0.058)	(0.218)	(0.051)
Log market potential				0.002***	0.008**	0.003***
				(0.001)	(0.003)	(0.001)
Log population density				-0.003***	-0.002	-0.004***
				(0.001)	(0.003)	(0.001)
Unemployment rate				0.003	0.038	-0.035***
				(0.010)	(0.039)	(0.013)
Share of agricultural workers				-0.021***	-0.024	-0.022**
				(0.008)	(0.029)	(0.011)
Share of manufacturing workers				-0.005	0.076	0.026
				(0.015)	(0.055)	(0.020)
Log average rainfall				0.008***	0.001	0.010***
				(0.002)	(0.007)	(0.003)
Log average temperature				-0.002	0.050**	-0.018**
				(0.006)	(0.021)	(0.007)
Share mining workers				-0.023***	-0.018	-0.000
				(0.007)	(0.027)	(0.013)
Share of area in homelands				-0.014***	-0.045***	-0.017***
				(0.003)	(0.011)	(0.004)
Constant	0.188***	0.045**	0.282***	0.187***	-0.088	0.406***
	(0.006)	(0.021)	(0.008)	(0.029)	(0.108)	(0.037)
Observations	354	354	354	354	354	354
R-squared	0.378	0.009	0.574	0.527	0.124	0.686
F-test	213.6	3.131	473.4	34.67	4.383	67.91
Convergence rate (%)	1.7	-	3.2	3.3	2.7	7.9
Half-life (years)	41	-	22	21	24	9

Asterisks indicate the level of significance, where: *** p<0.01, ** p<0.05, * p<0.1 and the values in parentheses are heteroscedasticity-consistent standard errors. The dependent variable is the average annual growth rate of income per capita. Columns (1) - (3) tests for unconditional convergence for the periods 1996-2011, 1996-2011 and 2001-2011 respectively, while columns (4) – (6) tests for conditional convergence for the same periods.

The first observation from Table 4.6 is that the revealed estimates for the log of initial income per capita in all columns lead to similar conclusions as those shown in Table 4.2 and 4.3. As with income per worker, these results show no evidence of unconditional β-convergence but do reveal conditional β-convergence over the 1996-2001 period, as well as robust evidence of both unconditional and conditional β-convergence from 1996 to 2011 and 2001 to 2011. Second, like income per worker, conditional convergence of income per capita is driven mainly by differences in initial skills levels, market potential, population density, unemployment, share of agricultural workers and former homeland areas.

The third observation is that unconditional convergence rates of income per worker and income per capita are similar, but their conditional convergence rates differ significantly, with higher rates observed for income per worker. This result highlights the heterogeneous nature of the South African local labour market and shows that regional structural factors play a greater role in explaining convergence dynamics of income per worker than income per capita. However, despite the variability in the conditional convergence rates, the results, shown in tables 4.3 and 4.6, lead to similar conclusions. This provides evidence that our conclusions in the previous section are not being driven by the indicator used to measure regional economic disparity[68].

Alternative econometric model

The cross-section OLS econometric models estimated thus far ignores spatial autocorrelation. However, spatial autocorrelation is an important force in the process of regional convergence and ignoring it in the estimation could, therefore, result in model misspecification (Rey & Montouri, 1999), leading to biased estimates (Fingleton, López-Bazo, & Lopez-Bazo, 2006). The findings in the previous chapter confirm evidence of strong spatial autocorrelation in the distribution of income per worker across regions. We check whether the revealed spatial autocorrelation affects our results by testing for the presence of spatial autocorrelation in the OLS estimates reported in Table 4.2 and 4.3. To achieve this, we use three different tests, namely the Moran's I, the Lagrange Multiplier (LM) and the Robust Lagrange Multiplier (RLM) tests (see Anselin, 1988; Anselin, Le Gallo and Jayet, 2008 for more details on these tests). All tests are based on a row standardised spatial weights matrix based on the great-circle distance between the geographic centres of the regions.

The test results for the unconditional β-convergence models (in Table 4.2) are presented in Table 4.7. The Moran's I test results show significant spatial autocorrelation in the residuals of

[68] However, it is important to note that our results based on income per capita are likely to be more biased compared to estimates based on income per worker due to the high prevalence of individuals with reported zero and missing income in the censuses. Nevertheless, we include the analysis based on income per capita as it allows us to check consistency with your earlier results, as well as with other literature that uses GDP per capita.

all the OLS estimations. While the Moran's I statistic is commonly used in the literature, it does not show whether the revealed spatial autocorrelation is due to the omission of a spatial lag variable or due to overlooked spatial effects in the error term. This problem is addressed by the other two tests, the Lagrange Multiplier (LM) and the Robust Lagrange Multiplier (RLM). These tests check for the omission of a spatial lag variable and presence of spatial dependence in the error term. According to the decision rule by Anselin & Florax (1995), there is need to include a spatial lag variable if the LM-test for spatial lag dependence (LM$_{Lag}$) is more significant than the LM-test for spatial error dependence (LM_{Error}) and the RLM-test for spatial lag dependence (RLM$_{Lag}$), which is robust against the presence of spatial error dependence is significant[69].

Both the LM tests for the omission of a spatial lag variable (LM_{Lag}) and presence of spatial dependence in the error term (LM_{Error}) strongly reject the null hypothesis of no spatial autocorrelation for the period 1996-2011 and 2001-2011. In addition, while the RLM test for the omission of a spatial lag variable (RLM$_{Lag}$) is not significant, the RLM test for the presence of spatial dependence in the error term (RLM_{Error}) is still significant. This implies that the presence of spatial dependence in the error term is robust to the presence of spatial dependence in the lag variable. These test results also hold for conditional β-convergence models for which we did not present the results as the interpretation is similar to the one provided above.

Table 4.7: Test for presence of spatial autocorrelation in the unconditional OLS model

	1996-2011		1996-2001		2001-2011	
Test	Statistic	p-value	Statistic	p-value	Statistic	p-value
$Moran's\ I$	12.783	0.000	1.7160	0.086	12.426	0.000
LM$_{Error}$	132.713	0.000	1.661	0.197	126.254	0.000
RLM$_{Error}$	9.153	0.002	0.1800	0.672	62.484	0.000
LM$_{Lag}$	125.721	0.000	1.898	0.168	63.823	0.000
RLM$_{Lag}$	2.161	0.142	0.416	0.519	0.053	0.819

Notes: These tests are based on the OLS estimations presented in Table 4.2 and 4.3.

Given these findings, a spatial model with a spatially autocorrelated error term might provide a better fit to our data than simple OLS models for the period 1996-2011 and 2001 -2011[70].

[69] On the other hand, there is presence of spatial dependence in the error term if the LM-test for spatial error dependence (LM$_{Error}$) is more significant than the LM-test for spatial lag dependence (LM_{Lag}) and the RLM-test for spatial error dependence (RLM$_{Error}$), which is robust against the presence of spatial lag dependence is significant.

[70] However, estimation of a spatial model is rejected completely for the period 1996-2001.

Accordingly, we modify equation (7) and (8) and estimate a spatial error model (SEM) which incorporates spatial effects in the error term as follows:

$$\frac{1}{T}\ln(y_{it}/y_{it-1}) = \alpha + \beta\ln(y_{it-1}) + \sum_{n=1}^{N} \alpha_i X_{i,t} + (1 - \rho W\ )^{-1}\mu_{it} \tag{9}$$

where W is an $n\ x\ n$ spatial weight matrix, whose element w_{ir} captures the degree of spatial relations between two regions (i and an r)[71]. ρ is the spatial error coefficient which measures the intensity of the spatial dependence of the residuals across regions. To estimate equation (9) we use the Maximum Likelihood approach as using OLS method to estimate spatial models may lead to inefficient parameter estimates (yet unbiased), which in turn biases the parameter variances (Anselin & Bera, 1998; Elhorst, 2003). The estimation results of the SEM with and without controls for the period 1996-2011 and 2001-2011 are presented in Table 4.8. Columns (1) and (2) report estimates for the unconditional β-convergence models, which we get by assuming that $\alpha_i = 0$ in equation (9), while columns (3) and (4) present estimates for conditional β-convergence models based on equation (9).

The results reveal a positive and highly significant spatial error coefficient in all columns, confirming positive correlation of the error terms across space. The parameter of interest, log initial income per worker remains negative and statistically significant in all the columns. This suggests that, after accounting for spatial effects in the error term, we continue to find evidence of unconditional and conditional convergence over the 1996 - 2011 and 2001 – 2011 periods. Even more interesting, comparing the spatial error model (SEM) results in Table 4.8 to the OLS results in Table 4.2 and 4.3, we see that the estimates of all variables are of similar magnitude. Thus, we confidently conclude that failure to control for spatial effects in the error term does not affect our OLS estimations.

In summary, although the magnitude of the estimates of the log initial income per worker from the various sensitivity tests above shows some variability, the general conclusions reached, and shown in Table 4.2 and 4.3, continue to hold[72].

[71] We define the spatial weight matrix W based on distance, where distance is the inverse of the great-circle distance between the geographic centres of two regions. More details on this spatial weight matrix are provided in Chapter 3 of this thesis.

[72] Interestingly, as shown in appendix 4, Table 4.8A, the conclusions reached in Table 4.2 and 4.3 continue to also hold even when we test for unconditional and conditional β-convergence using income per worker which includes workers with zero income.

Table 4.8: $\boldsymbol{\beta}$-convergence test of income per worker, Spatial Error Model (SER)

	Unconditional		Unconditional	
	1996-2011	2001-2011	1996-2011	2001-2011
Log initial income per worker	-0.009***	-0.031***	-0.038***	-0.070***
	(0.002)	(0.003)	(0.004)	(0.004)
Skilled workers (%)			0.320***	0.356***
			(0.063)	(0.048)
Log market potential			0.002***	0.005***
			(0.001)	(0.001)
Log population density			-0.003***	-0.004***
			(0.001)	(0.001)
Unemployment rate (%)			-0.012	-0.027**
			(0.008)	(0.012)
Share of agricultural workers (%)			-0.038***	-0.044***
			(0.007)	(0.011)
Share of manufacturing workers (%)			-0.014	0.006
			(0.014)	(0.020)
Log rainfall			0.007***	0.012***
			(0.002)	(0.003)
Log temperature			0.002	0.000
			(0.006)	(0.008)
Share of mining workers (%)			-0.007	0.009
			(0.007)	(0.013)
Share of area in homelands (%)			-0.015***	-0.017***
			(0.003)	(0.004)
Constant	0.091***	0.248***	0.233***	0.410***
	(0.020)	(0.027)	(0.033)	(0.043)
lambda	0.926***	0.930***	0.892***	0.913***
	(0.072)	(0.068)	(0.104)	(0.085)
Observations	354	354	354	354
Variance ratio	0.049	0.257	0.397	0.563
Log likelihood	1032.4	879.5	1110.3	980.8
Convergence rate (%)	1.023	3.748	5.6	11.99
Half-life (years)	68	19	12	6

Asterisks indicate the level of significance, where: *** p<0.01, ** p<0.05, * p<0.1 and the values in parentheses are heteroscedasticity-consistent standard errors. The dependent variable is the average annual growth of real income per worker. We test for unconditional convergence for the period 1996-2011 and 1996-2011 in columns (1) and (2) respectively, while we test for conditional convergence for the same time periods in columns (3) and (4).

4.9. Conclusion

This chapter undertook an empirical examination of the extent to which wages have converged or diverged across regions in South Africa between 1996 and 2011. The chapter draws on the same regional dataset we construct and discuss in Chapter 3. We use regional income per worker to proxy for regional wage per worker. The study is a contribution to the literature on regional wage convergence, which is currently limited in the context of Africa. Using both descriptive and econometric methods, this study shows our findings are consistent across the three measures of convergence employed. This is in contrast to some of the existing convergence literature, which reveals mixed evidence based on the kernel density estimator, σ-convergence and β-convergence methods (see Magrini, 2004; Young et al. 2008).

The descriptive analysis based on the kernel density estimator and σ-convergence analysis provides evidence of increasing regional disparities in workers' incomes from 1996 to 2001, and decreasing regional disparities in workers' incomes from 2001 to 2011. This result points to the presence of a process of regional divergence over the period 1996-2001, which was followed by a process of regional convergence over the period 2001-2011. Secondly, the study finds that, despite, evidence of regional convergence in more recent years, regional disparities in income per worker remains high in South Africa. These disparities are much larger than observed in other countries, both developed and developing countries. Thirdly, the cross-sectional β-convergence econometric analysis reveals no evidence of unconditional convergence among regions over the 1996-2001 period, but shows strong unconditional convergence across regions for 2001-2011. The unconditional convergence rate for 2001-2011 is estimated at 3.7% per year. This indicates that it could take 19 years to reduce the worker income gap between rich and poor regions by half.

Further analysis from β-convergence econometric analysis shows that there is strong evidence of conditional convergence over the 1996-2001 period, as well as from 2001 to 2011. The analysis reveals a higher rate of conditional convergence of 6.1% per year between 1996 and 2001, which increases to 13% per year between 2001 and 2011, after controlling for homeland status and initial differences in numbers of skilled workers, unemployment, market potential, population density and share of agricultural workers. This implies that if differences in these regional structural factors were eliminated, it takes between 5 and 11 years to reduce the worker income gap between rich and poor regions by half. Thus, region-specific factors, some of which can be influenced directly by policy, are constraining regional income convergence. The results from the econometric analysis are robust to exclusion of outlier regions, use of an alternative sample of workers, an alternative measure of regional economic disparity and an alternative econometric model.

Taken together, the evidence of regional convergence of income per worker (capita) during 2001-2011 offers an encouraging message to policymakers concerned with addressing regional economic imbalances in South Africa. However, the study also shows that workers' income disparities are still observed across regions, and decreasing slowly. The study lends support to policy initiatives aimed at improving the underlying conditions of lagging regions, especially in former homeland areas. For instance, policy measures promoting human capital accumulation, and greater access to markets by improving transport infrastructure, as well as addressing the problems of high unemployment, population density and low productivity in the agricultural sector, can have significant effects on regional economic development and overall convergence of income and wages across regions.

In this chapter, we have empirically highlighted the key determinants of growth of income per worker across regions in South Africa. However, to fully understand the South African spatial economy and why differences in worker's incomes remain large, there is a need for an explicit examination of the causes of these regional disparities in worker's incomes. In the next chapter, we turn our attention to an empirical test of a key theoretical prediction of the new economic geography (NEG) theory on the causes of regional wage disparities.

Chapter 5

5. Can the New Economic Geography explain regional wage disparities in South Africa?

5.1. Introduction

Regional wages vary significantly in both developed and developing economies and show a strong core-periphery economic structure. The New Economic Geography (hereafter, NEG) theory pioneered by Krugman (1991) provides a theoretical explanation for how and why such disparities exist and persist over time. NEG theory emphasises the importance of access to markets in the determination of a region's wages. It predicts that greater access to markets leads to higher local wages. A representation of this prediction set out in Fujita, Krugman, & Venables (1999), is a "wage equation", that posits a positive relationship between regional wages and market potential, that is, an index measuring accessibility to markets (Head & Mayer, 2004).

An extensive body of research has empirically tested the validity of the wage equation in various countries, including Italy (Mion, 2004), Germany (Brakman, Garretsen, & Schramm, 2004; Kosfeld & Eckey, 2010), the United States (Hanson, 2005; Fallah, Partridge, & Olfert, 2011), Brazil (Fally et al. 2010), Indonesia (Amiti & Cameron, 2007), China (Hering & Poncet, 2009, 2010) and Chile (Paredes, 2015). While this research covers both developed and developing economies, few studies have tested the validity of the wage equation in African economies[73]. However, the income and wage inequalities in African countries are highly conditional on natural resource exploitation, historical institutional settings, and peculiar labour market conditions, and such studies might open new lines of research in this area. The main objective of this chapter is to test the validity of the wage equation for South Africa, one of the most unequal countries in the world.

The study of regional wage disparities and the empirical validation of the wage equation in South Africa has several motivations. Firstly, to the best of our knowledge, at the inception of this thesis, analysis of wage disparities at the regional level has rarely been the main subject of research in South Africa. The existing literature has discussed the trends and causes of wage inequality at the individual level (Burger, 2015; Burger & Yu, 2007; Ntuli & Kwenda, 2014; Wittenberg & Pirouz, 2013; Wittenberg, 2014, 2016, 2017). A few exceptions that have looked

[73] A notable exception is the work by Bosker & Garretsen (2012) which estimates a theory-based wage equation across a group of African countries. Using GDP per worker to proxy wages, Bosker & Garretsen (2012) find evidence in support of the wage equation.

at the causes of wage disparities at the regional level include Kingdon & Knight (2006), Magruder (2012) and Von Fintel (2014; 2017)[74]. Secondly, a number of studies have used NEG theory to explain regional disparities in GDP per capita (Naudé & Krugell, 2003, 2005, 2006; Naudé, Krugell, & Matthee, 2010), output per worker (Krugell & Rankin, 2012), manufacturing (Fedderke & Wollnik, 2007) and export performance (Gries & Naude, 2008; Matthee & Naudé, 2008; Naudé & Gries, 2009). While this research has shown that market potential matters, the evidence is based on the estimation of reduced-form equations that do not show clearly the precise mechanisms through which market potential operates.

Finally, as discussed in chapter 3 and 4, South Africa is characterised by significant disparities in the distribution of income per worker (a proxy for wage per worker) across regions. The disparities show evidence of substantial positive spatial concentration patterns consistent with the core-periphery economic structure predicted by NEG theory. The question is whether the observed positive spatial concentration patterns have anything to do with the mechanisms postulated by NEG theory or whether they are an outcome of alternative factors put forward by competing theories. NEG theory explains regional wage disparities based on very narrow economic forces generated by the interplay of the manufacturing and transport sectors. However, these sectors are less developed in South Africa than in developed countries, where the theory has largely been tested.

Like many emerging economies, South Africa has a robust primary sector highly dependent on natural resources (minerals, agricultural land, access to waterways and favourable climate). This factor is neglected by the NEG theory. However, natural resources, together with historical institutional settings and peculiar labour market conditions might be key determinants of regional wages in South Africa and many other emerging countries. The possible tension between these factors on one hand, and NEG economic forces, on the other hand, provides a unique testing ground of the validity and robustness of the NEG theory for explaining regional wage disparities in emerging economies.

A key empirical question addressed in this chapter, therefore, is whether the prediction of the NEG wage equation is consistent with the observed regional wage disparities in South Africa,

[74] Magruder (2012) uses magisterial districts as the unit of analysis to examine the effects of bargaining councils on labour market outcomes (employment, employment by firm size and wages by industry). Kingdon & Knight (2006), and von Fintel (2017) use individual workers identified by their location (360 clusters, magisterial districts, district councils and provinces) as the unit of analysis to examine the effects of local unemployment on individual wages. Using individuals as the unit of analysis is empirically appealing as it enables them to control for within-region heterogeneity due to differences in worker characteristics.

a country where regional wage disparities seem to be driven by several factors. To address this question, the study estimates a structural wage equation for the years 1996, 2001, and 2011 based on the Helpman-Hanson model. Apart from showing the relationship between regional wages and market potential, estimating the Helpman-Hanson model is empirically appealing as the estimated structural wage equation allows us to reveal the precise channels through which market potential drives regional wage disparities. It also allows us to check the consistency of the estimated results with the predictions of the underlying theoretical framework, as well as related studies from other countries.

The chapter contributes to the empirical literature on regional wage disparities in several ways. First, it provides an empirical validation of the NEG wage equation in the context of Africa where studies are still limited. Second, it estimates for the first time a structural wage equation for South Africa based on the Helpman-Hanson model. Third and in line with the overall objective of the thesis, the study contributes to a better and deeper understanding of the causes of regional wage disparities in South Africa by augmenting the Helpman-Hanson model with other potential explanatory factors. Finally, the study contributes towards the practical policy debate on regional wage disparities in South Africa. The goal of addressing regional wage disparities and creating a more equitable labour market may benefit from policies that are well-informed in the determinants of regional wage disparities, in particular, the precise mechanisms through which market potential affects regional wage levels.

The remainder of this chapter is structured as follows. The next section discusses the Helpman-Hanson theoretical framework which underlines the empirical analysis. Section 5.3 provides a brief review of the related empirical literature. This is followed by section 5.4 that presents the empirical framework. Section 5.5 describes the data. The results of the empirical findings are reported in section 5.6. Lastly, section 5.7 offers conclusions and policy implications.

5.2. The New Economic Geography (NEG) theory

There are several theoretical models for explaining the causes of regional wage disparities, among them, the human capital theory, the amenity theory, the regional wage curve theory, and the NEG theory. These models are all important. The NEG theory, however, provides a solid theoretical explanation for how and why regional wage disparities exist and persist over time. This explanation is based on micro-foundations of increasing returns to scale, transport costs, and consumer’s love of variety, which, together, create pecuniary externalities that influence the location decisions of economic agents (Fujita & Mori, 2005; Fujita & Thisse, 2009).

The NEG literature consists of several theoretical models[75]. In this chapter we discuss how we empirically test one of these, the Helpman (1998) model, derived from the well-known core-periphery NEG model by Krugman (1991). Although the Krugman (1991) and the Helpman (1998) models are similar in most aspects, they differ in their definition of the competitive sector that acts as the dispersion force. In both models, labour mobility in the manufacturing sector acts as the driving force for agglomeration. However, Krugman (1991) uses immobile agricultural labour and Helpman (1998) non-tradeable housing services as the dispersion force.

In using immobile agricultural labour as the dispersion force, Krugman (1991) has generally been criticised for failing to capture some of the spatial characteristics of agglomeration that have been found to be relevant empirically, such as prices of non-tradable – housing services (Brakman et al. 2004). By incorporating housing services, the model designed by Helpman (1998) seems to be more suitable, as it captures the key localisation factors (congestion costs - higher land prices) affecting location decisions of firms and consumers. The model is also preferred because of the less extreme nature of its equilibria. While Krugman (1991) predicts complete agglomeration, which is hardly ever observed, Helpman (1998) allows for partial agglomeration, as high prices of non-tradable services push some economic agents away from agglomeration areas. Thus, by allowing some agents to spread out, the model captures closely the spatial characteristics of most economies.

We provide an explanation of the structure of the Helpman (1998) model. Given that the theoretical framework of this model has been derived many times (see Brakman et al. 2004; Hanson, 2005; Kiso, 2005; Bruyne, 2010; Kosfeld & Eckey, 2010), the discussion focuses on the key dynamics of the model that leads to the Helpman-Hanson model tested in this chapter.

5.2.1. Helpman-Hanson model

The economy is assumed to have R regions where each region has two sectors, each producing one good. All consumers have identical Cobb-Douglas preferences and maximise utility by consuming homogeneous, non-tradable housing services, and a variety of differentiated

[75] As discussed in chapter 2, some of these models include, Krugman (1991), Krugman & Venables (1995), Venables (1996) and Fujita, Krugman, & Venables (1999). For a review of these NEG models, their main features and the NEG empirics, see Overman et al. (2001), as well as Head & Mayer (2004).

tradable manufactured goods. In the economy, housing stocks are produced in a perfectly competitive market and the supply is fixed in each region (Hanson, 2005; De Bruyne, 2010). In this scenario prices for housing services will tend to be high in densely populated areas and low in sparsely populated areas (Kosfeld & Eckey, 2010).

The manufactured good is produced in a monopolistic competitive market by a firm operating under increasing returns to scale using mobile labour as the only factor of production. Manufactured goods are traded across regions at a cost modelled in the form of an "iceberg" transport cost, meaning that only a fraction of the shipped good arrives at the final destination. The good can be thought of as a composite of differentiated varieties, and the consumption of each variety is determined by its price, as well as the elasticity of substitution of the manufactured varieties (σ).

In deciding where to locate in space, firms and consumers choose locations that maximise profits and utility, respectively. Since manufacturing firms are confronted with increasing returns to scale, they prefer to concentrate production in just one region with greater access to markets - "the home market effect" - to minimise transport costs and benefit from large-scale production (Redding, 2013). On the other hand, because of their love of variety and the need to avoid paying higher transportation costs in importing manufactured goods, consumers also favour locating close to regions with greater access to large markets that offer a wide variety of manufactured goods at lower prices - "the price index effect" (Redding, 2010). The interplay of the market access effect and the price index effect is mutually strengthening and generates agglomeration forces that stimulate firms and consumers to concentrate close to regions with greater access to markets.

However, the resulting concentration of firms and consumers has associated congestion costs, as demand for housing increases, leading to an increase in housing prices - "the crowding effect" (Kosfeld & Eckey, 2010). This cost acts as a dispersion force, pushing for the spreading out of economic activity. Overall the spatial distribution of economic activity and resultant wages depend on the tension between agglomeration and dispersion forces (Fujita, 2007), whose strength, in turn, depends on transportation costs (Redding, 2013). When agglomeration forces are stronger than dispersion forces, an optimal economic result is the concentration of firms and workers in core locations with good access to markets. This concentration usually leads to regional wage disparities, with wages higher in core regions where economic activities agglomerate than peripheral ones (Kiso, 2005).

Under these conditions, the long-run spatial equilibrium of the economy can be summarised by five simultaneous equations related to real wage, housing expenditure, income, price index, and nominal wage. The nominal wage equation is central in examining the relationship between market potential and local wages. The most prominent representation of this equation is set out in Fujita, Krugman, & Venables (1999):

$$w_r = \left[\sum_i Y_i T_{ri}^{1-\sigma} {P_{iM}}^{\sigma-1}\right]^{\frac{1}{\sigma}} \quad (1)$$

where w_r is wage for region r. Equation (1) is the NEG wage equation, and gives the average wage that firms in region r are willing to pay their workers. The average wage is a function of income in all other regions (Y_i), the price index for manufactured varieties in all regions (P_{iM}), transport costs between region r and i (T_{ri}) and the elasticity of substitution between manufactured varieties (σ) which satisfies $\sigma > 1$. While equation (1) is central in the empirical validation of the NEG wage equation, it cannot be estimated directly, since P_{iM} is an implicit price that cannot be observed directly (Kiso, 2005).

To arrive at a testable equation, Hanson (1998) utilises two additional equilibrium conditions of the Helpman (1998) model[76]. In particular, Hanson (1998) uses the equilibrium condition of the housing market given by:

$$P_{rH} H_r = (1-\mu) Y_r \quad (2)$$

where P_{rH} is the price of housing services in region r. H_r is the fixed housing stock in region r. Y_r is total income in region r. Equation (2) suggests that, in equilibrium, the market value of housing services supplied in region r equals the share of income spent on housing services in the region, given by $1-\mu$. μ is the share of income devoted to consumption of manufactured goods.

In addition, Hanson (1998) uses the real wage equalisation assumption, where free labour mobility equalises real wages across regions. While nominal wages can vary across regions in the short-run, labour mobility ensures that in the long-run real wages are equalised as

[76] Another commonly used strategy to eliminate the manufactured price index is provided by Redding & Venables (2004) who use estimated parameters from a gravity trade model based on bilateral trade data to derive a theory-based market potential function consisting of two indices: market access and supplier access indices. This strategy is less appealing for regional analysis because of the unavailability of regional-level trade data in most countries, and also because of its underlying assumption of labour immobility.

manufacturing workers migrate from regions with low wages to locations with higher wages to realise higher utility levels. Real wages are equalised across regions by deflating nominal wages by the regional cost-of-living price index that is a function of housing price P_H and manufactured goods price index P_M. Workers have no motive to migrate when real wages are equalised across regions,

$$\frac{w_r}{P_{rH}^{1-\mu}P_{rM}^{\mu}} = \frac{w_i}{P_{iH}^{1-\mu}P_{iM}^{\mu}} = \omega \tag{3}$$

Utilising the housing market equilibrium condition (2) and the real wage equalisation condition (3), Hanson (1998) eliminates the unobservable manufacturing goods price index variable in equation (1) to obtain the following testable model:

$$w_r = \left[\sum_i Y_i^{\frac{\sigma(\mu-1)+1}{\mu}} H_i^{\frac{(1-\mu)(\sigma-1)}{\mu}} w_i^{\frac{\sigma-1}{\mu}} {T_{ri}}^{1-\sigma}\right]^{\frac{1}{\sigma}} \tag{4}$$

where all the other variables and parameters are defined as before. Equation (4) is now generally referred to as the *"Helpman-Hanson model"*. The right-hand side of the equation now gives a modified market potential index that is a function of income (Y_i), housing supply (H_i) and wages (w_i) in all other regions, and transport costs (T_{ri}) between regions (r and i). Hanson (1998) captures transport costs with an exponential distance decay function, $T_{ri} = e^{-\tau d_{ir}}$, where d_{ri} is the distance between region r and i. Thus, according to equation (4), the average wage in region r is higher when market potential is higher, shown by higher income (Y_i), housing stocks (H_i) and wages (w_i) in surrounding regions, as well as low transport costs between trading regions ($e^{-\tau d_{ir}}$). Thus, wages are systematically higher in regions with greater market potential (greater access to markets), and they are lower in regions with less market potential (low access to markets).

An appealing feature of equation (4) is that it is directly derived from theory. The implications of the NEG theory are fully captured by the key structural parameters of the model, namely, the elasticity of substitution among manufactured varieties (σ), the transport costs parameter (τ) and the share of income devoted to consumption of manufactured goods (μ) or housing services ($1 - \mu$). These parameters capture the tension between the forces of agglomeration (access to markets) and dispersion (higher housing costs) that shape the location decisions of economic agents, and consequently the distribution of wages across space[77].

Depending on the level of transport costs, the Helpman-Hanson model is said to be valid when agglomeration forces dominate dispersion forces. This happens when σ is low, μ is high and τ is high. The rationale is that low σ allows the exploitation of economies of scale gains by concentrating production in few locations. high μ supports a larger concentration of firms in need of high consumer demand, and high τ encourages concentration of firms and consumers in the same locations to avoid incurring high transport costs[78]. For agglomeration forces to dominate, the model parameters need to satisfy the following constraints: $\sigma > 1$, $0 \leq \mu \leq 1$ and $\tau \geq 0$.

In addition, two relations between model parameters should be satisfied. The first is the market power condition, given by $\sigma/(\sigma - 1)$, a ratio showing the mark-up of prices over marginal costs. The condition holds when $\sigma/(\sigma - 1)$ >1 and it implies that firms in a given region are operating under increasing returns to scale, a key feature in the NEG theory. The second is the no black hole condition, which is given by $\sigma(\mu - 1)$. The condition holds when $\sigma(\mu - 1) < 1$ and it's critical in the overall validation of the Helpman-Hanson model. When the condition holds, it implies that in the determination of regional wage, agglomeration and dispersion forces are interacting in a way that is consistent with the Helpman-Hanson model. In such as case and depending on the level of transportation costs, economic activities will either agglomerate or disperse in space (Mion, 2004). For instance, the interaction of increasing returns to scale at the firm level and consumers' love of variety with increasing transport costs leads to more agglomeration of economic agents, and widening of regional wage disparities (Helpman, 1998). However, when the condition does not hold, it implies that the NEG theory is irrelevant in explaining the distribution of economic activities and wages across regions (Kiso, 2005). Rather, the spatial distribution of economic activities and wages is determined by exogenous regional factors such as housing stocks and natural resources (De Arcangelis & Mion, 2002).

[77] Thus, these parameters reveal the precise channels through which market potential influences the spatial distribution of economic activities and wages across regions.

[78] Economic agents tend to disperse in space, leading to regional wage convergence when σ is high, μ is low and τ is low. The rationale is that high σ erodes the benefits from economies of scale gained by concentrating production in few locations; low μ, implies a higher share of income is being devoted to housing services, which is turn drives economic agents to locations with low housing costs. Low τ implies firms can serve different markets, while consumers can import goods from different markets at relatively low transport costs.

5.3. Related empirical literature

This section focuses on the empirical literature that tests the wage equation based on the Hanson (1998) approach, which leads to the Helpman-Hanson model given by equation (4)[79]. Hanson (1998) was the first to empirically test the validity of the Helpman-Hanson model, using data from 3075 US counties for 1970-80 and 1980-90. In a revised version, Hanson (2005) augments equation (4) with region-specific characteristics to control for other potential explanations. Measuring regional wages with average annual earnings per worker, Hanson (2005) finds a significantly positive relationship between regional wages and market potential. His results show that the higher the personal income, wages, and housing stocks in proximate locations and the lower the transport costs to those locations, the greater the local wage. His results are robust to the inclusion of human capital, demographic composition of working age population, and exogenous amenities.

Hanson (1998) also finds highly significant structural parameters ($\sigma; \mu$ and τ) that are consistent with the underlying Helpman (1998) model. More precisely, his estimates of the elasticity of substitution, σ, range between 4.9 and 7.6 and satisfy the restriction that $\sigma > 1$. The implied mark-up of prices over marginal cost ($\sigma/(\sigma$ -1)) ranges from 1.15 to 1.26, suggesting that firms operate under increasing returns to scale and have some degree of monopoly power. Estimates of the share of income devoted to manufactured goods, μ, range between 0.91 and 0.98. These estimates are consistent with the theoretical restriction of $0 \leq \mu \leq 1$, although the values are somewhat high. The estimates of the transport cost parameter, τ , fall in the range 1.6 to 3.2, which is also in line with the theoretical expectation that $\tau \geq 0$. He further finds that the interplay of the parameters shown by $\sigma(1 - \mu)$ reveal values between 0.084 and 0.653, showing that the no black hole condition holds. Hanson's (2005) findings therefore provide overwhelming evidence in support of the validity of the wage equation in the US[80].

[79] A large body of literature has tested the wage equation based on another approach by Redding & Venables (2004). See chapter 2 of this thesis for a detailed explanation of these two approaches, and why the approach by Hanson (1998) is preferred in this chapter.

[80] Hanson (2005) finds that the theory-based wage equation based on the Helpman (1998) model has greater explanatory power than the simple ad hoc wage equation based on the Harris (1954) market potential model that does not control for regional variation in housing stocks.

Researchers have subsequently tested the validity of the Helpman-Hanson model in many other countries. These include Roos (2001); Brakman et al. (2000; 2004) and Kosfeld & Eckey (2010) for Germany; Bruyne (2010) for Belgium; Mion (2004) for Italy; Cieślik & Rokicki (2016) for Poland, and Kiso (2005) for Japan. Amongst these studies, Brakman et al. (2004), extend the Helpman-Hanson model by including a variable for land prices along with housing stocks. They also relax the real wage equalisation assumption. Kosfeld & Eckey (2010) extend the Helpman-Hanson model by complementing housing stocks with a price index for manufactured goods, in line with the initial NEG wage equation (2). Kiso (2005) modifies the Helpman-Hanson model by including intermediate inputs and building stocks. Another group of studies test the validity of the Krugman (1991) model. These include Turgut (2014) for Turkey, Pires (2006) for Spain and Niebuhr (2006) for regions across EU countries.

The first point to notice is that the estimated parameters differ across these studies, suggesting varying strength of demand linkages in different countries. For instance, Roos (2001) estimates values of σ, μ and τ to equal 6.2, 0.86 and 0.003, respectively. De Bruyne (2009) estimates σ, μ and τ to be 5.5, 0.81 and 0.003, respectively. Mion (2004) finds estimates of σ, μ and τ of 1.92, 0.87 and 0.19 respectively. Brakman et al. (2004) find estimates of σ, μ and τ ranging from 3.1 to 4.9, 0.54 to 12.48 and -0.001 to 0.01, respectively. Despite these differences, it is important to note that all these studies reach similar conclusions, as they provide evidence in support of the wage equation. These studies thus provide further evidence validating the Helpman-Hanson model in explaining regional wage disparities in different countries, even after controlling for alternative explanations (Kosfeld & Eckey, 2010) and its extension or modification (Brakman et al. 2004; Kiso, 2005).

The studies cited above focus on developed economies. Extension of this research to developing countries is still very limited. A few notable exceptions include Moreno-Monroy (2008; 2011) for China, Moncarz (2007) for Argentina, and Paredes (2015) for Chile[81]. Following Brakman et al. (2004), Moreno-Monroy (2011) relax the real wage equalization assumption and find evidence in support of the Helpman-Hanson NEG model with estimates of σ ranging from 2.63 to 3.86 and τ from 0.45 to 0.97. In contrast, Moncarz (2007) and

[81] A few studies have tested the validity of the NEG theory based on the Hanson (1998) approach, which led to the development of a Helpman-Hanson model for developing countries. A number of studies have tested the NEG theory in different developing countries, based on Redding & Venables (2004) approach, which estimates a market potential index using estimates from a gravity trade model estimated using bilateral trade data. While the index can be decomposed into market access and supplier access indices, its greatest limitation for a regional analysis is its assumption of labour immobility.

Paredes (2015) fail to find strong evidence in support of the Helpman-Hanson model in Argentina and Chile, respectively. Paredes (2015) finds estimates of σ of between 41.8 and 46.01, μ of between 0.76 and 0.86 and τ of between 0.001 and 0.029. While these parameter estimates are within the ranges specified by the Helpman-Hanson model, the no-black-hole condition, which is critical for overall assessment of the validity of the model, is rejected. Paredes (2015) show that natural resources and exogenous endowments play a more important role than market potential in the determination of wages across regions in Chile. He concludes that the case of Chile is poorly explained by the NEG theory.

Overall, while studies using data for developed economies provide overwhelming evidence in support of the NEG theory, based on the Paredes (2015) findings, it seems clear that, this theory is not necessarily well-suited for some emerging countries. The following question then arises: Can the NEG theory explain regional wage disparities in emerging economies like South Africa, where a complex set of factors seem to influence regional wage levels?

5.4. Empirical framework

This chapter empirically tests the Helpman-Hanson model given by equation (4). To estimate the equation, a transport cost function needs to be defined. Hanson (1998; 2005) captures transport costs with an exponential distance decay function given by $T_{ir} = e^{-\tau d_{ir}}$. In our analysis discussed in this chapter, we use a distance power function, given by T_{ir}=$d_{ir}^{-\tau}$. As argued by Mion (2004), the distance power function is empirically appealing, because of its strong theoretical foundations within gravity trade models, that have been used to provide insights into NEG models[82]. Inserting the distance power function, taking logs, and imposing restrictions[83] to equation (4) parameters, we derive the following reduced form equation that we use as the baseline model:

$$\log(w_r) = \alpha_0 + \alpha_1 log\left[\sum_{i=1} Y_i^{\frac{1}{\alpha_1}-\alpha_2} H_i^{\frac{1}{\alpha_1}-1-\alpha_2} w_i^{\alpha_2} d_{ri}^{\alpha_3}\right] + \varepsilon_r \quad (5)$$

The dependent variable, $\log(w_r)$ is log wage per worker in regions r. α_0 is a function of constants (σ, μ, τ, f) and the equilibrium real wage, ω. Y_i, H_i and w_i capture income, housing

[82] For robustness checks, we also estimate the model using the exponential distance decay function. While the magnitude of the estimates differs between the functions, we get similar conclusions from the two functions.
[83] The following restrictions are imposed on equation (4) to get the following reduced-form parameters: $\alpha_2 = (\sigma - 1)/\mu$, $\frac{1}{\alpha_1} - \alpha_2 = \sigma + \frac{1-\sigma}{\mu}$; $\frac{1}{\alpha_1} - 1 - \alpha_2 = (\mu - 1)(\sigma - 1)/\mu$ and $\alpha_3 = -\tau(\sigma - 1)$. These restrictions are important, as they reduce the high nonlinearity of the model, which can affect convergence of the model.

stocks, and wages in region i respectively. d_{ir} is the distance between region i and r, that is used to proxy transport costs under the distance power function and ε_r is the regression error term.

As with equation 4, equation 5 captures the notion of a spatial wage structure, where wages increase as one moves closer to centres of production characterised by high market potential. While the importance of market potential is confirmed by a positive and significant α_1 coefficient ($\alpha_1 > 0$), estimation of the other reduced-form coefficients, α_2, and α_3 enable us to explicitly derive the structural parameters (σ, μ, τ) of the Helpman-Hanson model. A positive and significant wage coefficient ($\alpha_2 > 0$) and a negative and significant estimate for distance coefficient ($\alpha_3 < 0$) is expected.

Given the nonlinearity of equation (5), we follow existing literature (Roos, 2001; Brakman et al. 2004; Hanson, 2005) and estimate the equation using nonlinear least-squares (NLS) method. The main advantage of the NLS method is that it allows direct estimation of the nonlinear relationship between regional wages and market potential without having to linearise the model (Cieślik & Rokicki, 2016). The method also takes into account the constraints due to the links between the model parameters.

We test the robustness of the baseline model in various ways. First, we extend the baseline model by controlling for various region-specific factors, to fully account for the causes of regional wage disparities. As controls, we include measures of human capital, natural resources, climatic conditions, labour market conditions, and historical events[84]. Second, we test the robustness of the results from the extended model to potential bias, due to (1) reverse causality and (2) the inclusion of non-competitive sectors.

5.5. The Data

This study draws on the geographically consistent dataset we discuss in chapter 3, which we construct from the full population censuses, as well as the climate data discussed in chapter 4 produced by Harris, Jones, Osborn, & Lister. (2014) at the Climatic Research Unit (CRU) at the University of East Anglia. Data in this dataset is aggregated to the 354 magisterial districts (hereafter – regions) of South Africa for the years 1996, 2001, and 2011.

[84] A description of the variables is provided in the data section, while the motivation for their inclusion is given in the empirical results section.

The estimation of equation (5) requires data on wage per worker (w_r), income (Y_r), housing stocks (H_r), and distance between region (d_{ri}). As explained in chapter 3, regional income per worker is used to proxy regional wage per worker. Thus, in estimating equation (5), regional income per worker for 1996, 2001, and 2011 is used as the dependent variable. Due to the unavailability of wages data, this approach has also been used Redding & Venables (2004) who proxied wages with GDP per capita and Bosker & Garretsen (2012) who employed GDP per worker as a proxy for wages. For the components of the market potential index, we use total personal income to proxy for each region's income (Y_r), number of rooms in a dwelling to proxy for regional housing stocks (H_r) and we calculate distance (d_{ri}), as the great-circle distance (in kilometres) between two points (region i and r)[85].

To test the robustness of the relationship between regional income per worker and market potential, we include the following region factors: First, the share of workers (in total working-age population) in each region with a tertiary education is included, to capture regional differences in human capital (skilled workers)[86]. Second, the share of workers in the mining sector in each region is included, to proxy mineral resource endowments. Third, each region's average yearly temperature and rainfall is used to proxy regional climatic differences. These variables are included to capture the effects of local amenities. Fourth, the regional unemployment rate is included, to capture differences in local labour market conditions. Finally, homeland status, given by the share of each region that falls in a former homeland area, is used to capture persistent effects of Apartheid land policies. The summary statistics of all variables are presented in Table 5.1A in appendix 5.

5.5.1. The spatial distribution of key model variables

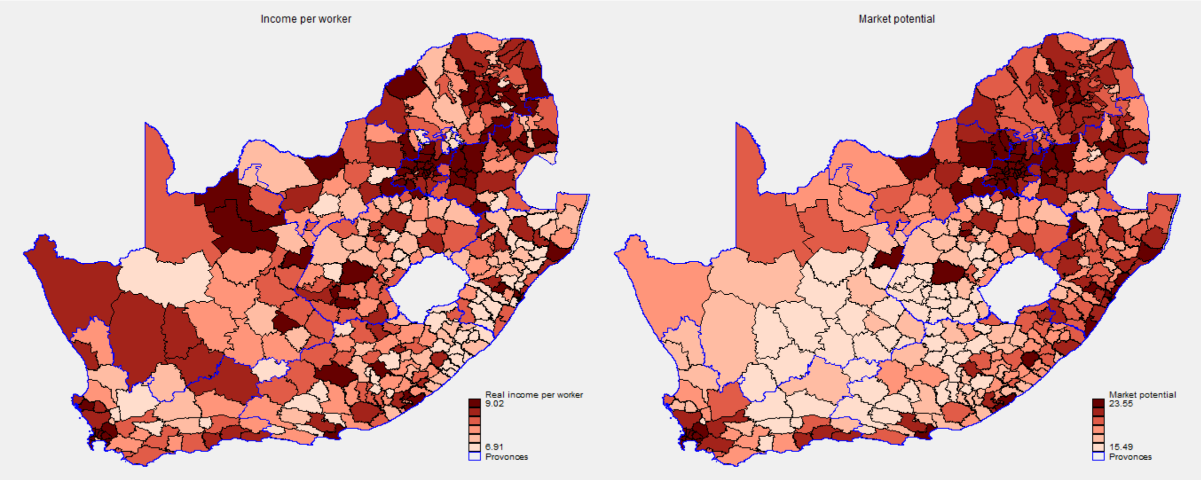
In this section, we provide initial insights concerning the relationship between regional income per worker and a measure of market potential, based on the Harris market potential index, given by distance-weighted personal income of each region. This index is not the same as the theory-based market potential function that will be estimated later. We simply use the index to gain initial insights into the potential association between the variables. We start with a visual display in Figure 5.1 of the spatial distribution of income per worker (LHS figure) and market

[85] For more details on the constriction of the distance variable, see the data section of chapter 4.

[86] While this definition of human capital is subject to debate, we check robustness of our definition by using different education cut-off levels, such as considering all workers with at least a matrix qualification. We also use the share of the working age population in each region with at least a tertiary education degree. Regardless of the definition used the importance of human capital remain evident in our analysis.

potential (RHS figure) across regions in South Africa in the year 2011[87]. On the maps, darker colours reflect higher values, while lighter colours reflect lower values. The maps highlight two key insights.

Figure 5.1: Distribution of income per worker and market potential across regions.



Source: Author's calculations based on census data aggregated to 354 magisterial districts.
Notes: Income per worker is derived by weighting total income from employed individuals with total employed individuals in each region with a positive income and aged 15-64 years. Market potential is based on the Harris (1954) market potential index, given by equation (6) in chapter 4, which shows the distance-weighted personal income for each region.

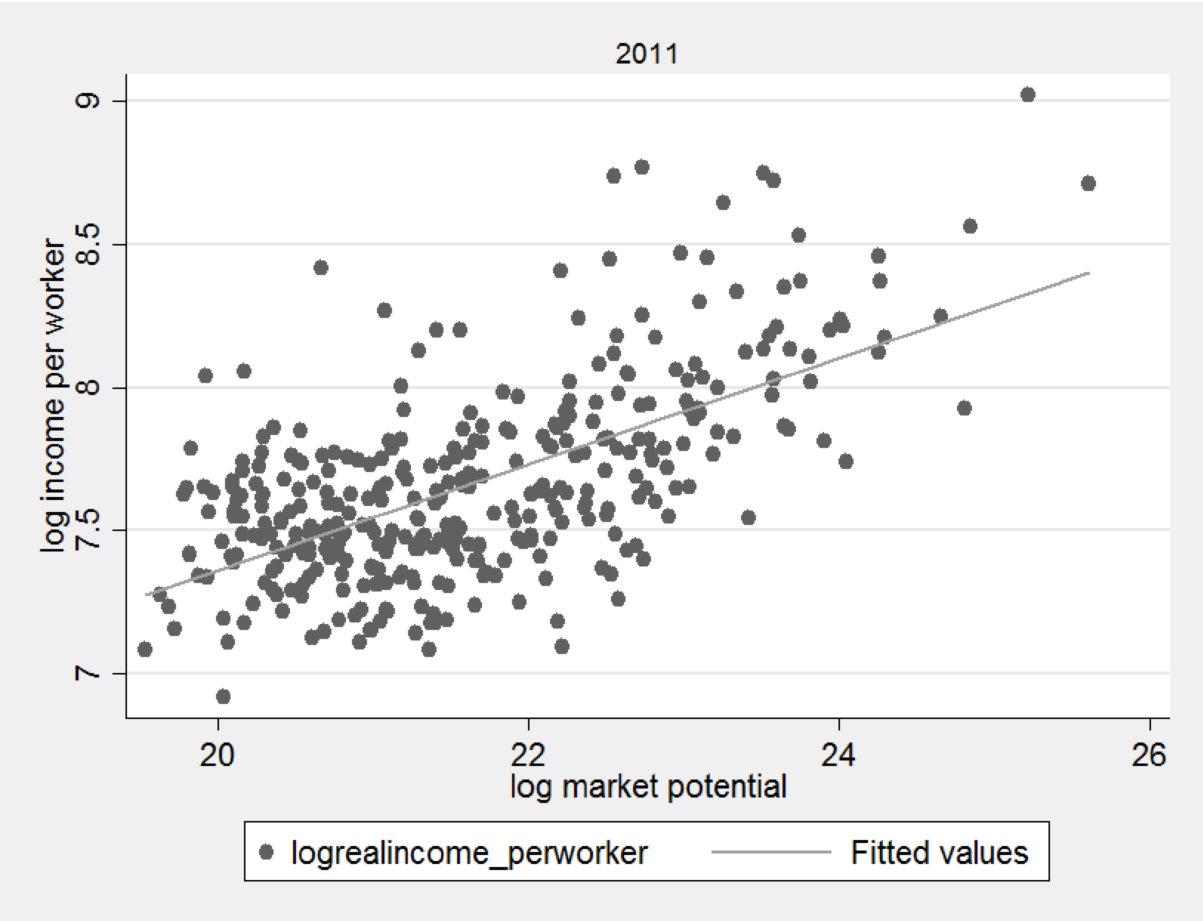
Firstly, income per worker and market potential vary significantly across regions. For example, we see high values of income per worker and market potential, of around 9 and 23 log points, respectively for regions in Gauteng, Western Cape, and some coastal cities along the east coast. In comparison, we see low values of about 7 and 15 log points, respectively for regions in some parts of Eastern Cape, KwaZulu Natal, and central parts of the country. Secondly, there is clear evidence of the significant spatial concentration of income per worker and market potential in specific locations. For instance, the bulk of regions with high values of income per worker and market potential tend to concentrate in Gauteng, Western Cape and some cities along the east coast. It is in these areas where economic activities are concentrated in South Africa. For example, looking at the 24 regions (out of 354 regions) for Gauteng only, our data shows that, while these regions account for less than 1.5% of the country's total land area, they contribute 22% to national population, 33% to national employment and 40% to total national personal income as of 2011. On the other hand, regions with low levels of income per worker and market

[87] The maps showing the spatial distribution of income per worker and market potential in 1996 and 2001 are presented in Appendix 5.1 Figure 5.2A and 5.3A respectively. The observed spatial patterns in these years mirror those observed in 2011. This suggests evidence of persistent regional disparities and spatial concentration of income per worker and market potential in South Africa.

potential tend to be clustered in the central parts of the country, and in some parts of Eastern Cape and KwaZulu Natal.

These distribution patterns suggest the existence of a positive relationship between income per worker and market potential. This relationship is confirmed by our results, shown in Figure 5.2 that plots the association between income per worker and market potential across regions in 2011. A clear positive relationship can be observed, which concurs with the underlying prediction of the NEG wage equation. Nevertheless, a closer look at Figure 5.1 and 5.2 suggests that some regions have high (low) incomes per worker, despite low (high) market potential. The regions in the northern parts of the country, some parts of North West, and KwaZulu-Natal provinces represent clear examples of these exceptions. This evidence suggests that, apart from market potential, variations in worker's incomes across regions might be explained by other factors, such as mineral resource endowments, and wildlife activities (extensive in the northern parts and North West), among other factors. This potentially provides support for the inclusion of other controls in a test of the validity of the NEG model.

Figure 5.2: Association between income per worker and market potential across regions



Source: Author's calculations based on census data aggregated to 354 magisterial districts.

Notes: Income per worker is derived by weighting total income from employed individuals with total employed individuals in each region with a positive income and aged 15-64 years. Market potential is based on the Harris (1954) market potential index, given by equation (6) in chapter 4, which shows the distance-weighted personal income for each region.

5.6. Empirical results

This section presents the results of the empirical analysis in three sections. The first section shows the estimation results for our baseline model, equation (5). The second section presents regression results testing the robustness of our baseline model to the inclusion of alternative explanations for regional disparities in workers' incomes. The last section reports estimates for various sensitivity tests carried out to explore whether our results are robust to potential bias due to reverse causation, as well as the exclusion of non-competitive sectors.

All regressions are estimated using a nonlinear least squares method, and all standard errors are corrected for heteroscedasticity[88]. From the different estimations, three sets of estimates are obtained, reduced-form, implied structural parameters, and estimates for alternative explanations. Of these estimates, the implied structural parameters are key in the validation of the Helpman-Hanson model. We provide a recap of the conditions the structural parameters need to satisfy in order for the model to be consistent in the case of South Africa, shown in Table 5.1.

Table 5.1: Helpman-Hanson model – structural parameter constraints

Structural parameter	Parameter description
$\alpha_1 > 0.$	Market potential estimate
$\sigma > 1.$	Elasticity of substitution between manufactured varieties
$0 < \mu < 1.$	Share of income devoted to manufactured varieties
$\tau > 0.$	Unit transport cost
$\sigma/(\sigma - 1) > 1.$	Market power condition reflecting imperfect competition
$\sigma(1 - \mu) < 1.$	No-black-hole condition

Notes: These structural parameters are derived from the reduced-form coefficients obtained from estimating equations (5) and (6). Thus, given α_1, α_2 and α_3 the structural parameters are obtained as follows: $\sigma = 1/\alpha_1$, $\mu = (1 - \alpha_1)/\alpha_1\alpha_2$ and $\tau = \alpha_1\alpha_3/(\alpha_1 - 1)$. From these parameters, two additional equilibrium conditions given by $\sigma/(\sigma - 1)$ – price-marginal cost ratio and $\sigma(1 - \mu)$ – no black hole condition, are also derived.

5.6.1. Baseline results of the Helpman-Hanson model

Table 5.2 presents the estimates of the baseline model, equation (5) that includes market potential as the only explanatory variable. Column (1) shows estimates for 1996, column (2)

[88] In all estimations, initial starting values for the model parameters are needed. These are extracted from the literature. To ensure robustness of our results, different starting values are used.

for 2001, and column (3) for 2011. The coefficients for α_1 to α_3 are all statistically significant, with signs consistent with theoretical expectations. The estimated coefficient of market potential (α_1) is positive and significant in all years, suggesting that market potential plays a significant role in explaining differences in income per worker across regions in South Africa. The results show that, a 10% increase in market potential is associated with a 1.13%, 0.98% and 1.7% increase in regional income per worker in 1996, 2001, and 2011, respectively. The estimate of income per worker (α_2) is positive and statistically significant, while that of distance (α_3) is negative and statistically significant in all the columns[89]. The negative distance coefficient shows that as a region's distance to consumer markets increases, its income per worker decreases. Thus, remote regions in South Africa face a market access penalty that lowers their income levels. This finding is consistent with theoretical expectations.

Table 5.2: Estimation of the Helpman-Hanson Model.

Time period	1996	2001	2011
Reduced form coefficients	(1)	(2)	(3)
Log market potential	0.113***	0.098***	0.170***
	(0.036)	(0.037)	(0.046)
Log income per worker	8.939***	10.511**	5.763***
	(3.015)	(4.183)	(1.710)
Log distance	-3.103***	-4.250***	-2.209***
	(0.774)	(1.422)	(0.377)
Implied Values			
σ.	8.823***	10.17***	5.897***
	(2.812)	(3.867)	(1.606)
μ.	0.875***	0.872***	0.850***
	(0.020)	(0.022)	(0.027)
τ.	0.397***	0.463***	0.451***
	(0.050)	(0.048)	(0.076)
$\sigma/(\sigma-1)$.	1.128	1.109	1.204
$\sigma(1-\mu)$.	1.102	1.298	0.886
Adjusted R-squared	0.491	0.477	0.416
F-statistic	114.4	108.2	84.88
Obs	354	354	354

Asterisks indicate the level of significance, where: *** p<0.01, ** p<0.05, * p<0.1 and the values in parentheses are heteroscedasticity-consistent standard errors. The estimated models includes a constant.

Next, we consider the structural parameters implied by these reduced form estimates. First, the implied values of σ are statistically significant and range between 5.9 and 10.2. The NEG theory assumes $\sigma > 1$, and therefore these values are in line with the theory. The values of σ

[89] We expect α_2 given by $\alpha_2 = (\sigma - 1)/\mu$ to be positive as $\sigma > 1$ and $0 < \mu < 1$.

suggest that firms across regions in South Africa are operating under increasing returns to scale, enjoying mark-ups (given by $\sigma/(\sigma - 1)$) of between 10.9 and 20.4 percent. These mark-ups are quite close to those of other studies from South Africa using industry and firm data. For example, Zalk (2014) finds mark-ups of between 10 and 20 percent, while Aghion, Braun, & Fedderke (2008) find a mark-up of 23.3 percent for the manufacturing sector in South Africa. While the estimates for σ are higher than the 4.9 to 7.6 range reported for the US (Hanson, 2005), they are lower than the 41.1 to 46.1 reported for Chile (Paredes, 2015). Thus, the US has stronger demand linkages and higher mark-ups than South Africa, Chile has weaker demand linkages and lower mark-ups than South Africa. This shows that the strength of demand linkages emphasized by the Helpman-Hanson model falls as one moves from developed countries towards emerging economy countries.

The implied value of μ, the share of income devoted to manufactured goods, is statistically significant and satisfies the restriction that $0 < \mu < 1$, suggested by theory, across all the years. However, with estimates of μ ranging from 0.85 to 0.88, these values indicate that only around 0.15 $(1 - \mu)$ of total household income is spent on housing services. As in other studies (Hanson, 2005; Mion, 2005), these values seem to be an overestimation of the share of income devoted to manufactured goods. According to StatsSA, about 32 percent of total household income is devoted to housing, water, electricity, gas and other fuels, with housing services taking up the largest share of the 32% (Stats SA, 2012).

The estimate of τ is statistically significant and positive ($\tau > 0$), as implied by theory. While the positive (τ) values are consistent with findings in the literature (Paredes, 2015; Mion, 2004; Pires, 2006), it is hard to compare them with other studies, due to the sensitivity of τ to the unit of analysis, and the transport cost function used, and the way distance is measured.

Finally, we consider the no black-hole condition, which holds when $\sigma (1 - \mu) < 1$. As discussed earlier, this condition is critical in the validation of the Helpman-Hanson model. While the no black-hole condition holds for 2011, it is rejected for 1996 and 2001. This suggests that, while the mechanisms emphasized by the Helpman-Hanson model play a key role in explaining the spatial distribution of worker's incomes in 2011, these mechanisms do not fully explain the distribution of worker's incomes in 1996 and 2001. The rejection of the condition for 1996 and 2001 suggests that agglomeration and dispersion forces are interacting in a way that is inconsistent with the Helpman-Hanson model. This implies that the spatial distribution of income per worker in these years might be driven by exogenous location factors.

In summary, while the baseline results reveal parameter estimates consistent with the Helpman-Hanson model, the no black hole condition is rejected for 1996 and 2001. This suggests that, while the spatial distribution of income per worker can be explained by the forces put forward by the Helpman-Hanson model (increasing returns to scale, transport costs, and consumers' love of variety), these forces are not interacting in a way that is consistent with the model. Based on this, we can conclude that the case of South Africa is not fully consistent with the Helpman-Hanson model.

5.6.2. Additional controls

A key question following from our discussion in the previous section is why the model does not fit the South African case well. A possible explanation is the presence of region-specific factors that influence the spatial distribution of income per worker across regions in a manner that is inconsistent with the predictions of the Helpman-Hanson model.

One compelling source of regional differences in income per worker is regional variation in human capital. The importance of human capital differences in driving regional wage and income disparities finds support from both theoretical (Becker, 1962; Willis, 1986; Romer, 1986; Lucas, 1988) and empirical literature (Combes, Duranton, & Gobillon, 2008; Fally, Paillacar, & Terra, 2010; Paredes, 2013; Cieślik & Rokicki, 2016).

Regional variation in local labour market conditions, such as unemployment, is another potential factor of influence. According to the wage curve theory, a negative association exists between unemployment and wage levels in a region (Blanchflower & Oswald, 1990; 2005) and this relationship has been confirmed empirically in different countries, including in South Africa (Magruder, 2012; Von Fintel, 2017).

Another notable cause of regional wage differences is variation in local amenities. According to the local amenity theory (Roback, 1982; 1988), differences in climatic conditions, natural resources, institutional quality, and cost of living all have an effect on levels of productivity, one of the main driving forces behind regional wage disparities (Maza & Villaverde, 2006). Regions with favourable amenities such as valuable natural resources, good access to waterways, favourable climatic conditions, and strong institutions and infrastructure can have higher productivity, which in turn may raise workers' incomes in these regions.

Regional incomes of workers may also be explained by historical events, such as the establishment of the apartheid system from 1948. The apartheid regime implemented several racially segregatory policies, including the homeland policy, which dispossessed about 3.5 million blacks of their land and forcefully relocated them into ten "homeland" areas, according to their ethnic groups. Apart from being already overcrowded, these "homelands" were highly marginalised and distant from major economic centres (see Figure 3.1). The advent of democracy in 1994 led to the end of the apartheid-era rule and the legal reintegration of all homeland areas into South Africa. In addition, numerous policies were implemented to promote regional economic development and address regional economic disparities created by years of apartheid-era rule. Despite this, the legacy of apartheid-era racially segregatory policies may still affect regional income distributions. This claim finds support in a growing body of research that suggests that distinct historical events have long-lasting effects that shape economic development on an ongoing basis (Nunn, 2009; Acemoglu & Dell, 2010).

The factors discussed above might create tensions with the mechanisms put forward by the NEG theory in explaining regional worker's incomes, leading to the poor empirical performance of the theory. To check whether this is the case, equation (5) is augmented to capture the effects of regional differences in human capital, local amenities, and local labour market conditions, as well as historical events as follows[90]:

$$\log(w_r) = \alpha_0 + \alpha_1 log\left[\sum_{i=1} Y_i^{\frac{1}{\alpha_1}-\alpha_2} H_i^{\frac{1}{\alpha_1}-1-\alpha_2} w_i^{\alpha_2} d_{ri}^{\alpha_3}\right] + \sum_{n=1}^{N} \beta_n X_{rn} + \varepsilon_r \qquad (6)$$

where in addition to the variables defined in equation (5), X_{rn} is a vector of regional controls, which we discuss in the data section, and β_n is the vector of corresponding coefficients. Estimating equation (6) enables us to estimate the relation between regional income per worker and market potential, after controlling for the effects of region specific factors.

The results from the empirical test of equation (6) are reported in Table 5.3. Column (1) shows estimates for 1996, column (2) for 2001, and column (3) for 2011. The main insight from these results is that the inclusion of regional controls improves the fit of the model, as shown by

[90] If these region-specific factors are constant over time, their effects, as well as effects of other unobserved regional factors, can be isolated by estimating a time-differenced version of equation (5), as done by Hanson (1998; 2005). However, apart from loss of information due to time-differencing (Pires, 2006; Cieślik & Rokicki, 2016), several empirical studies report a poor fit of the wage equation estimated in time-differences (Roos, 2001; Niebuhr, 2006). Indeed, despite our best efforts, estimating a time-differenced version of equation (5) did not provide useful results, as the model did not converge for the case of South Africa. We therefore decide not to present the results.

increasing values of the adjusted R squared in all columns. For example, the adjusted R squared increased from 0.42 in table 5.2 to 0.76 in table 5.3 for 2011. Furthermore, the estimates are now consistent with the predictions of the Helpman-Hanson model. This is confirmed by the no black hole condition ($\sigma\ (1 - \mu) < 1$) that now holds in all columns[91].

Looking at the results in more detail, we see that the estimates associated with market potential (both reduced-form and structural parameters) remain significant and consistent with theory, but important changes can be seen in these estimates. For instance, the effect of market potential becomes stronger, with the estimate (α_1), increasing from between 0.10 and 0.17 (Table 5.2) to between 0.23 and 0.34 (Table 5.3). At the same time, the coefficients associated with income per worker decrease from between 5.76 and 10.51 to between 2.70 to 4.22. The distance also decreased from between -2.21 and -4.25 to between -0.79 and -1.37. The decrease in the distance estimates highlight that failure to account for region specific factors leads to an overestimation of the effects of transport costs. The results are consistent with findings by Hanson (2005) for the US. He finds increases in market potential estimates (from 0.132 to 0.203), as well as decreases in distance estimates (from 17.91 to 6.43) after inclusion of regional controls. Overall, these results suggest that the effects of the NEG theory in explaining regional differences in income per worker will be underestimated if region specific factors are not controlled for.

In line with the changes in the reduced form estimates, significant changes can also be seen in the corresponding structural parameters (σ; μ and τ). For instance, σ decreases from between 5.90 and 10.17 (Table 5.2) to between 2.98 and 4.35 (Table 5.3). This decrease shows less competition among firms, which in turn implies higher mark-ups, that rise from between 10.9% and 20.4% (Table 5.2) to between 29.8% and 50.5% (Table 5.3). Further, the share of income devoted to manufactured goods decreases, suggesting an increase in the share of income spent on housing services. This increases from between 12.5% and 15% (Table 5.2) to between 17.9% and 26.7% (Table 5.3). These values are much closer to the 32% reported by StatsSA. While the magnitude of these estimates continues to differ significantly from those reported

[91] Given the important role that apartheid-era policies played in shaping the South African spatial economy, we carried out robustness checks to see whether homelands are an anomaly to NEG theory and are potentially the reason why the model does not work. We achieved this by re-estimating equation (6) including the homeland variable as the only additional control. Our results show that the NEG theory continue to poorly explain the case of South Africa, even after controlling for homelands or other regional specific factors individually. It is only after incorporating all relevant controls that the NEG theory holds for South Africa.

for Chile (Paredes, 2015), they are in the range reported for Germany (Roos, 2001), for the US (Hanson, 2005) and for Spain (Pires, 2006).

Table 5.3: The Helpman-Hanson Model with additional controls

Year	1996	2001	2011
Log market potential	0.244***	0.230***	0.336***
	(0.071)	(0.065)	(0.100)
Log income per worker	3.779***	4.216***	2.698***
	(1.259)	(1.346)	(0.959)
Log distance	-1.099***	-1.370***	-0.785***
	(0.143)	(0.192)	(0.114)
Implied values			
σ.	4.102***	4.351***	2.979***
	(1.192)	(1.223)	(0.889)
μ.	0.821***	0.795***	0.733***
	(0.045)	(0.038)	(0.071)
τ.	0.354***	0.409***	0.397**
	(0.108)	(0.105)	(0.167)
$\sigma/(\sigma-1)$.	1.322	1.298	1.505
$\sigma(1-\mu)$.	0.735	0.893	0.794
Additional controls			
Human capital			
Skilled workers (%)	3.540***	2.345***	2.732***
	(0.345)	(0.241)	(0.166)
Local amenities			
Mineral resource endowments (%)	0.358***	0.823***	0.836***
	(0.088)	(0.154)	(0.176)
Log temperature	0.638***	0.706***	0.309***
	(0.123)	(0.146)	(0.109)
Log rainfall	-0.062**	-0.072	-0.027
	(0.031)	(0.048)	(0.040)
Local labour market condition			
Unemployment rate (%)	-0.542***	-0.219	-0.912***
	(0.097)	(0.138)	(0.162)
Historical event			
Homeland status (%)	-0.176***	-0.341***	-0.354***
	(0.040	(0.060)	(0.041)
Adjusted R-squared	0.736	0.664	0.757
F-statistic	99.413	70.774	111.13
Obs	354	354	354

Asterisks indicate the level of significance, where: *** p<0.01, ** p<0.05, * p<0.1 and the values in parentheses are heteroscedasticity- consistent errors. The estimated models include a constant.

The interplay of the key model parameters (σ and μ) now provide evidence in support of the no black hole condition for all the years. This suggests that, in South Africa, agglomeration and dispersion forces interact in a way that is consistent with the Helpman-Hanson model once we account for the effects of region specific factors. With the no black hole condition now holding for all the years, as predicted by Helpman (1998), the increase in transport costs (τ) between 1996 and 2001 suggests an increase in agglomeration of economic activities in centres locations with good access to markets, which in turn leads to an increase in regional disparities in incomes of workers. The slight decrease in transport costs between 2001 and 2011 points to a decrease in agglomeration, which in turn points to a slight decrease in regional disparities in worker's incomes. These dynamics are consistent with the findings in chapter 4, which showed evidence of an increase disparities in income per worker over the 1996-2001 period and a decrease in disparities in workers' incomes between 2001 and 2011.

Looking at the controls, the coefficient of skilled workers is positive and statistically significant, shown in all columns. On average, if the share of skilled workers increases by 1%, regional income per worker increases by between 2.3% and 3.5%. This result is consistent with existing empirical literature that finds evidence in support of the importance of skills composition in explaining regional wage disparities (Combes, Duranton, & Gobillon, 2008; Huang & Chand, 2015).

For local amenities, the estimate for mineral resource endowments is statistically significant and positive for all the years, suggesting that if mineral resources increase by 1%, incomes of workers increase by between 0.36% and 0.84%. Further, the coefficient of regional average temperature is positive and statistically significant in all the columns, suggesting that a 1% increase in regional temperature is associated with a 0.31% to 0.71% increase in income per worker. Finally, the estimate for rainfall is negative but only significant in column (1), suggesting that a 1% increase in rainfall is associated with a 0.062% decrease in income per worker in 1996.

Turning to the other controls, regional unemployment is statistically significant and negatively associated with regional income per worker, with the exception of 2001 (column 2). An increase in unemployment in a region by 1% is associated with a 0.54% to 0.91% decrease in regional workers' incomes. This result supports the wage curve theory and earlier findings in South Africa by von Fintel (2017) who finds a negative relationship between unemployment and mean wages. Finally, homeland status estimates are negative and statistically significant in all columns, suggesting that an increase of 1% in the proportion of each region in a former homeland reduces regional incomes by between 0.18% and 0.35%. This result suggests that, despite the abolishment of the apartheid-era rule and the reintegration of homeland areas into

South Africa in 1994, the legacy of apartheid-era rule continues to negatively affect incomes for regions in former homeland areas[92].

In summary, the results described in this section show that the Helpman-Hanson model alone is not sufficient to fully explain regional disparities in incomes of workers in an emerging economy like South Africa. They show that variation in regional income per worker is well explained by the mechanisms emphasised by the Helpman-Hanson model, namely increasing returns to scale, transport costs and consumers' love of variety, but only after controlling for regional specific factors unique to South Africa[93]. This implies that proper application of the Helpman-Hanson model in an emerging economy like South Africa hinges on the incorporation of other region-specific factors. Thus, the results highlight the need to extend the Helpman-Hanson model with additional explanatory factors when this is applied to emerging economies like South Africa, where several factors neglected by the NEG theory also matter in the income distributions. This finding is consistent with findings from other emerging economies which are endowed with natural resources, such as Chile (Paredes, 2015).

5.6.3. Robustness Checks

This section presents robustness and sensitivity tests of the results reported in Table 5.3 to check potential bias due to (1) reverse causality and (2) inclusion of non-competitive sectors.

Reverse causality issues

The results in Table 5.3 might be biased due to problems of reverse causality arising from two main sources. First, regional income per worker, w_r, is present on both (left and right) sides of the equation. It therefore acts as the dependent as well as the independent variable in the model. Secondly, regional income per worker, w_r, is also a component of regional total income, Y_r. The standard approach to deal with potential bias due to reverse causality is to use the instrumental variable (IV) technique. However, it is difficult to come up with reliable instrumental variables, given that most economic variables are also endogenous (Redding, 2010). In addition, the established literature acknowledges that the nonlinear form of the wage

[92] Given that apartheid-era rule provided inferior education and underdeveloped labour markets in homeland areas, it is most likely that these areas are also characterised by low levels of skilled workers, as well as high unemployment. Thus, homeland status and unemployment rate are likely to have a large negative cumulative effect on regional incomes.

[93] Acknowledging that our results might be sensitive to the inclusion of employed individuals with zero income, we also estimated equation (5) and (6) using income per worker, which includes workers with zero income. The results presented in Table 5.3A in Appendix 5, shows the estimates from the two models remain quantitatively similar to those excluding workers with zero income.

equation makes estimation incorporating instrumental variables extremely complicated, leading to non-convergence of the model (Moreno-Monroy, 2008; Paredes, 2015). An alternative way to reduce the potential bias due to reverse causality is to calculate market potential, excluding own region market potential (removing Y_r, H_r, w_r and d_{rr}). However, excluding internal market potential would introduce measurement error by considerably reducing the market potential of some of the economically more powerful locations (Breinlich, 2006), such as Gauteng.

For this reason, we choose to keep internal market potential and use two strategies to check the robustness of the results in Table 5.3 to potential bias due to reverse causality. First, we follow López-Rodríguez & Faíña (2006) and measure each region's market size using total regional population instead of total regional income. This reduces the possible correlation between market potential and the error term, as regional population is strongly correlated with regional income, but less strongly correlated with regional income per worker (see Table 5.2A in appendix 5).

The results presented in Table 5.4 show that the estimates of market potential are robust to the use of regional total population as a measure of regional market size in all the years. The estimates remain statistically significant and consistent with the underlying theory. The point estimates in Table 5.4 are not significantly different from those reported in table 5.3. Where they differ, they lie well-within the 95-percent confidence interval of the initial estimates. The importance of the additional controls also remains evident.

As a second check, we use historical data in constructing the market potential index. We take advantage of the longitudinal dataset that we created in chapter 3 using the 3 censuses and construct a market potential index based on 5, 10, and 15 year lagged data[94]. This enables us to use regional income per worker for 2001 and market potential for 1996, regional income per worker for 2011 and market potential for 2001, and regional income per worker for 2011 and market potential for 1996. This strategy should reduce the correlation between market potential and the error term significantly. The results are reported in Table 5.5.

[94] Taking longer time-lags has the advantage of reducing problems associated with shocks that are to some extent correlated over time.

Table 5.4: Sensitivity tests – Capturing market size with regional total population.

Year	1996	2001	2011
Log market potential	0.226***	0.254***	0.294***
	(0.045)	(0.048)	(0.078)
Log income per worker	4.044***	3.682***	3.126***
	(0.934)	(0.822)	(0.980)
Log distance	-1.168***	-1.327***	-0.849***
	(0.123)	(0.127)	(0.109)
Implied parameters			
σ.	4.432***	3.937***	3.407***
	(0.889)	(0.743)	(0.900)
μ.	0.849***	0.798***	0.770***
	(0.031)	(0.027)	(0.051)
τ.	0.340***	0.452***	0.353***
	(0.075)	(0.092)	(0.133)
$\sigma/(\sigma-1)$.	1.291	1.340	1.416
$\sigma(1-\mu)$.	0.670	0.797	0.784
Control variables			
Skilled workers (%)	3.524***	2.372***	2.710***
	(0.347)	(0.244)	(0.165)
Mineral resource endowment (%)	0.357***	0.802***	0.840***
	(0.088)	(0.153)	(0.176)
Log temperature	0.612***	0.669***	0.297***
	(0.120)	(0.147)	(0.105)
Log rainfall	-0.080**	-0.107**	-0.029
	(0.033)	(0.051)	(0.041)
Unemployment rate (%)	-0.554***	-0.281***	-0.908***
	(0.096)	(0.146)	(0.162)
Homeland status (%)	-0.175***	-0.335***	-0.355***
	(0.040)	(0.059)	(0.041)
Adjusted R-squared	0.739	0.669	0.757
F-statistic	100.71	72.23	111.06
Obs	354	354	354

Notes: Asterisks indicate the level of significance, where: *** p<0.01, ** p<0.05, * p<0.1 and the values in parentheses are heteroscedasticity-consistent standard errors. The estimated models include a constant. Column (1) reports estimates for 1996, (2) for 2001, and (3), for 2011 when we replace regional total income with regional total population.

While the results show some variability in the market potential estimates, the general conclusions remain as before. The estimates not only lie in the range of those reported in Table 5.3 but also retain both economic and statistical significance. This suggests that the results in Table 5.3 are robust to the use of historical data. As before, the importance of additional controls remains evident. The observed small changes in these estimates are consistent with findings by Amiti & Cameron (2007), as well as by Martinez-Galarraga et al. (2015) who also find marginal changes to initial market potential estimates, after using historical explanatory variables as well as regional population.

Table 5.5: Sensitivity tests – Use of lagged values.

Parameters	5-year lag	10-year lag	15-year lag
Log market potential	0.197***	0.207***	0.240***
	(0.049)	(0.037)	(0.042)
Log income per worker	4.618***	4.499***	3.947***
	(1.301)	(0.901)	(0.812)
Log distance	-1.571***	-1.425***	-1.479***
	(0.208)	(0.115)	(0.131)
Implied Values			
σ.	5.072***	4.830***	4.175***
	(1.260)	(0.853)	(0.732)
μ.	0.882***	0.851***	0.804***
	(0.030)	(0.023)	(0.023)
τ.	0.386***	0.372***	0.466***
	(0.081)	(0.066)	(0.081)
$\sigma/(\sigma-1)$.	1.246	1.261	1.315
$\sigma(1-\mu)$.	0.600	0.719	0.817
Control variables			
Skilled workers (%)	4.099***	3.707***	2.391***
	(0.466)	(0.343)	(0.245)
Mineral resource endowments (%)	0.485***	0.388***	0.854***
	(0.102)	(0.090)	(0.160)
Log temperature	0.590***	0.560***	0.596***
	(0.154)	(0.123)	(0.151)
Log rainfall	-0.109**	-0.083**	-0.115**
	(0.045)	(0.033)	(0.050)
Unemployment rate (%)	-0.034	-0.321	-0.348**
	(0.136)	(0.090)	(0.135)
Homeland status (%)	-0.303***	-0.170***	-0.349***
	(0.064)	(0.044)	(0.062)
Adjusted R-squared	0.595	0.719	0.664
F-statistic	58.732	101.255	78.532
Obs	354	354	354

Notes: Asterisks indicate the level of significance, where: *** p<0.01, ** p<0.05, * p<0.1 and the values in parentheses are heteroscedasticity-consistent standard errors. Estimated models include a constant. Column (1) reports estimates, where the dependent variable is regional income per worker for 2001 and independent variables, is market potential for 1996, column (2) regional income per worker for 2011 and market potential for 2001, and column (3) regional income per worker for 2011 and market potential for 1996.

Sectoral analysis

An additional concern is that the results remain biased because of inclusion of non-competitive sectors. It is the case that, while our results this far include information from all sectors of the economy, the mechanisms put forward by the NEG theory might be distorted by the inclusion of non-competitive sectors such as the public sector, which is classified under Stats SA's industry classification "community; social and personal services"[95]. The mechanisms driving demand and supply in this sector are not fully driven by market forces. For example, public sector wages are set centrally by the state. We test the robustness of the results in Table 5.3 by controlling for public sector workers.

The estimation results are presented in Table 5.6. Because sectoral data is not available in the full national population census for 2011, we present estimates based on 1996 and 2001 data only. Column (1) reports estimates for 1996, while column (2) presents estimates for 2001. The estimate for the share of workers in the public sector is positive and statistically significant in all columns. This suggests that, on average, working in the public-sector increases workers' incomes. The market potential estimates, though, remain statistically significant and highly consistent with the underlying theory. In addition, the importance of the other additional controls remains evident. In both columns, the point estimates in Table 5.6 are not significantly different from those reported in table 5.3, hence lead to similar conclusions.

However, the mechanisms put forward by the NEG theory are more appropriate for the manufacturing sector, which is the key sector driving agglomeration in most NEG models. Thus, we further test the robustness of the results in Table 5.2 and 5.3 by focusing on the manufacturing sector. The estimation results are presented in Table 5.7. Because sectoral data is missing in the full national population census for 2011, and the sector data for 1996 is implausible, we present estimates based on 2001 data only. Column (1) reports estimates for the association between income per worker and market potential, while column (2) presents estimates which include region-specific factors.

The results in column (1) show that the case of South Africa is well explained by the mechanisms emphasised by the Helpman-Hanson model, once we narrow down to the manufacturing sector. The analysis clearly shows that the interplay of the key model parameters (σ and μ) provide evidence in support of the no black hole condition, even without controlling for region specific factors. However, even if this is the case, the results in column (2) continue to show evidence of the importance of region specific factors in explaining regional disparities

[95] By non-competitive sectors we refer to sectors classified in the censuses as "community; social and personal services", which includes public administration and defence activities, education, health and social work, other community; social and personal service activities, activities of membership organisations and recreational; cultural and sporting activities.

in incomes of workers. The inclusion of other controls improves the estimates and fit of the model.

Table 5.6: The Helpman-Hanson Model controlling for public sector workers.

Year	1996	2001
Log market potential	0.221***	0.219***
	(0.055)	(0.057)
Log income per worker	4.228***	4.410***
	(1.205)	(1.288)
Log distance	-1.342***	-1.582***
	(0.149)	(0.208)
Implied values		
σ.	4.521***	4.566***
	(1.134)	(1.179)
μ.	0.833***	0.809***
	(0.032)	(0.032)
τ.	0.381***	0.444***
	(0.087)	(0.095)
$\sigma/(\sigma-1)$.	1.284	1.280
$\sigma(1-\mu)$.	0.756	0.874
Control variables		
Share of workers in public sector (%)	0.761***	1.236***
	(0.255)	(0.284)
Skilled workers (%)	3.058***	1.601***
	(0.340)	(0.250)
Mineral resource endowments (%)	0.431***	0.935***
	(0.084)	(0.152)
Log temperature	0.583***	0.580***
	(0.086)	(0.103)
Log rainfall	-0.051***	-0.078***
	(0.026)	(0.043)
Unemployment rate (%)	-0.439***	-0.522***
	(0.100)	(0.133)
Homeland status (%)	-0.207***	-0.410***
	(0.045)	(0.056)
Adjusted R-squared	0.721	0.676
F-statistic	92.443	74.768
Obs	354	354

Asterisks indicate the level of significance, where: *** p<0.01, ** p<0.05, * p<0.1 and the values in parentheses are Heteroscedasticity-consistent standard errors. The estimated models include a constant. Based on equation (9), column (1) reports estimates for 1996 and column (2) presents estimates for 2001.

The changes in the transport cost parameter, τ, which increases from 0.41 in Table 5.3 (column 2) to 0.77 in Table 5.6 (column 2), highlights that transport costs are higher for the manufacturing sector. This is to be expected, given that manufactured goods are highly tradable, and firms thus incur trade costs for shipping the goods across regions. As a result of high transport costs, the NEG theory predicts that manufacturing firms will concentrate in urban areas where demand for their products is high to save on transport costs and enjoy large-scale production (Kosfeld & Eckey, 2010).

Table 5.7: The Helpman-Hanson Model– Manufacturing sector analysis – 2001.

Variables	Without controls	With controls
Log market potential	0.109**	0.150***
	(0.047)	(0.047)
Log income per worker	9.014**	6.418***
	(4.077)	(2.163)
Log distance	-6.014**	-4.327***
	(2.478)	(1.157)
Implied Values		
σ.	9.195**	6.645***
	(3.976)	(2.085)
μ.	0.909***	0.880***
	(0.031)	(0.030)
τ.	0.734***	0.767***
	(0.061	(0.091)
$\sigma/(\sigma-1)$.	1.122	1.177
$\sigma(1-\mu)$.	0.836	0.800
Control variables		
Skilled workers (%)		0.660**
		(0.310)
Mineral resource endowment (%)		0.986***
		(0.330)
Log temperature		0.721***
		(0.214)
Log rainfall		-0.142
		(0.094)
Unemployment rate (%)		-0.432*
		(0.235)
Homeland status (%)		-0.347***
		(0.095)
Adjusted R-squared	0.543	0.625
F-statistic	140.88	59.87
Obs	354	354

Asterisks indicate the level of significance, where: *** p<0.01, ** p<0.05, * p<0.1 and the values in parentheses are Heteroscedasticity-consistent standard errors. The estimated models include a constant. Based on equation (9), column (1) reports estimates for all sectors, while column (2) focuses on the manufacturing sector and column (3) on the services sector. All estimates are based on 2001 data.

5.7. Conclusion

This chapter sets out to empirically test whether the prediction of a wage equation derived from the new economic geography (NEG) theory is consistent with regional wage disparities in South Africa. The bulk of existing regional studies in South Africa estimate reduced-form equations incorporating variables inspired by the NEG theory. A major contribution of our research covered in this chapter is the estimation of a structural wage equation based on the Helpman-Hanson model derived directly from the NEG theory. The study reveals several important findings.

Firstly, the analysis provides evidence of a highly significant and positive relationship between regional income per worker and market potential, over the period 1996 – 2011, as postulated by the NEG theory. Regional incomes of workers in South Africa rise with increasing market potential of the region. In addition, the study finds evidence of highly significant and theoretically consistent structural parameters (σ, μ and τ) of the Helpman-Hanson model. This evidence which seems to confirm the importance of market potential in the determination of regional income per worker. However, a condition critical in the validation of the Helpman-Hanson model, the no black hole condition, $\sigma(1-\mu) < 1$, is rejected for 1996 and 2001. This suggests that income per worker in South Africa is not fully explained by the mechanisms emphasised by the Helpman-Hanson model. This evidence is consistent with findings from other emerging economies, such as Argentina and Chile (Alvarado & Atienza, 2014; Paredes, 2015).

Secondly, the results indicate that regional disparities in income per worker in South Africa are well-explained by the Helpman-Hanson model only after controlling for region-specific factors. The inclusion of these factors improves the fit of the model, leading to more precise market potential estimates (both reduced-form and structural parameters) that lie in the range reported by other studies (Hanson, 2005, Pires, 2006). The results reveal that the effect of market potential become stronger, while the no black hole condition now holds for all the years. This suggests that, once the effects of region-specific factors are isolated, the mechanisms emphasised by the Helpman-Hanson model play an important role in explaining dispersions in income per worker across regions in South Africa. Based on these findings, we can conclude that neglecting region-specific factors can seriously bias market potential estimates.

Thirdly, the study shows evidence of increasing importance of market potential over time, suggesting growing demand linkages across regions over time. The revealed market potential estimates, however, show that the strength of demand linkages emphasised by the Helpman-Hanson model are weak for South Africa compared to developed economies like the U.S and Japan, but strong compared to other emerging economies, like Chile and Ecuador (see Hanson, 2005; Kiso, 2005; Alvarado & Atienza, 2014; Paredes, 2015).

Finally, further analysis shows that the results are robust to a set of sensitivity tests carried out to check for potential bias due to reverse causality. Restricting estimates to the manufacturing sector provides clear evidence in support of the Helpman-Hanson model, even without controlling for region-specific factors. Even in this case, the inclusion of other controls improves the estimates and fit of the model.

From this analysis, we can, therefore, conclude that, although the mechanisms put forward by the Helpman-Hanson model are appropriate, there are not sufficient to fully explain the observed disparities in income per worker across regions in an emerging economy like South Africa. It is the interplay of increasing returns to scale, transport costs and demand patterns, as postulated by the NEG theory that, together with region-specific factors that fully explain the distribution of income per worker across regions in South Africa. Thus, the prediction of the NEG wage equation holds for the case of South Africa only after inclusion of alternative explanatory factors. This finding is consistent with results in chapter 3 which suggested the co-existence of the NEG theory (positive autocorrelation) and standard economic theory (negative autocorrelation) features in the distributions of income per worker in South African.

The findings of this study have important empirical, theoretical and policy implications. From an empirical perspective, we show that the Helpman-Hanson model is not sufficient in explaining regional disparities in incomes of workers in emerging economy countries like South Africa. Its proper application hinges on the incorporation of other region-specific factors unique to emerging economies. From a theoretical perspective, the evidence points to a need to extend the NEG theory to incorporate explanatory factors unique to emerging economies. Finally, from a policy perspective, the results highlight the need for policy initiatives that improves the underlying conditions of lagging-periphery regions, especially in former homeland areas which were highly marginalised during Apartheid. It further highlights the need for policy initiatives that foster human capital accumulation and address the problem of high unemployment. Despite these policy options, the findings also suggest that regional disparities are likely to remain a feature of the South African spatial economy because of the large agglomeration benefits arising from greater access to markets. These benefits can create possible tensions with policy initiatives aimed at spreading economic activities to peripheral regions and addressing regional imbalances, as economic agents are reluctant to relocate from large urban areas like Johannesburg and Cape Town where access to markets and wages are high.

Chapter 6

6. General Conclusions and Policy Relevance

6.1. Summary of key insights

Regional wage disparities are known to be large in both developed and developing countries and are often a source of public concern. Globally, growing regional wage disparities are known to act as a barrier to greater well-being, social cohesion and economic growth. Moreover, empirical evidence reveals that low productivity, social dissatisfaction, negative externalities and high levels of regional inflation, are also partly a consequence of regional income and wage disparities. Considering these negative consequences, the issue of regional wage disparities should concern researchers and policymakers. However, there is little systematic analysis of the spatial patterns that characterise the distribution of wages across regions in many developing countries, particularly in Africa. Correspondingly, there is insufficient understanding of the convergence dynamics and the causes of regional wage disparities in most African countries. As a result, policy debates tend to take place in something of an information vacuum.

The primary objective of this thesis is to provide new empirical evidence on regional wage disparities by examining the spatial patterns, the convergence dynamics, and the causes of regional wage disparities in South Africa using 1996, 2001 and 2011 full population census data. The thesis contributes to a currently small body of empirical evidence on regional wage disparities in the context of Africa. The study also contributes to a deeper understanding of the South African spatial economy. The empirical analysis is conducted in three inter-related chapters.

The first empirical chapter (Chapter 3) discusses and construct a regional level dataset using the full population census data. This data is then used to examine the spatial patterns that characterise the distribution of wages across regions in South Africa. An objective was to assess the consistency of these patterns with predictions from alternative theories. Using exploratory spatial data analysis (ESDA) techniques, we highlight two important features of the distribution of wages across regions in South Africa.

Firstly, the analysis reveals significant and persistent disparities in the distribution of wages across regions. In line with the international literature, the results show that these disparities are characterised by positive spatial autocorrelation, which suggests that regions with high (low) wages have neighbouring regions that are characterised by high (low) wages, as well as negative spatial autocorrelation, which suggests that regions with high (low) wages have neighbouring regions that are characterised by low (high) wages. These spatial patterns show that regions tend to be interdependent in their development processes, such that wages in a given region are not only influenced by economic conditions in that region, but also by wages and economic conditions in neighbouring regions.

Secondly, the revealed spatial patterns suggest the co-existence of the new economic geography (NEG) theory (positive autocorrelation) and standard economic theory (negative autocorrelation) features in South Africa. None of the theories fully match all the empirical features found in the South African regional wage distribution. However, with the spatial patterns dominated by positive spatial autocorrelation, the NEG theory seems to offer a more plausible explanation of the spatial features observed in the distribution of wages in South Africa. Overall, the data and analysis discussed in this chapter provide an important input into the subsequent two chapters of the thesis.

The second empirical chapter (Chapter 4) presents a detailed analysis of the convergence dynamics of regional wages in post-apartheid South Africa. The aim of the study is to find out whether wages have converged or diverged across regions in South Africa between 1996 and 2011. Using both descriptive (kernel density estimator and the notion of σ-convergence) and econometric (notion of β-convergence) methods, the analysis shows the consistency of the findings across the three measures of convergence employed.

The descriptive analysis based on the kernel density estimator and the notion of σ-convergence provides evidence of increasing regional wage disparities from 1996 to 2001 and decreasing regional wage disparities from 2001 to 2011. This points to a process of regional wage divergence over the 1996-2001 period, which was followed by a process of regional wage convergence between 2001 and 2011. Secondly, the study finds that, despite evidence of regional wage convergence in more recent years, regional wage disparities remain in South Africa. The extent of these disparities is much greater than observed in other countries, both developed and developing. Thirdly, the cross section econometric analysis based on the notion of unconditional β-convergence provides no evidence of unconditional convergence among regions over the 1996-2001 period. However, it reveals evidence of unconditional convergence across regions from 2001 to 2011, with the unconditional convergence rate estimated at 3.7%

per year. This means that it could take 19 years to reduce the wage gap between rich and poor regions by half.

Furthermore, in Chapter 4 we also test for evidence of conditional β-convergence by controlling for region specific factors. We find robust evidence of conditional convergence over the 1996-2001 period, as well as the 2001-2011 period. The conditional convergence rate is estimated at 6.1% per year between 1996 and 2001 and 13% per year between 2001 and 2011, after controlling for homeland status and initial differences in human capital, unemployment, market potential, population density and share of workers in the agricultural sector. Thus, for the 2001-2011 period if differences in regional specific factors had been eliminated, it would take 5 years as opposed 19 years to reduce the wage gap between rich and poor regions by half. This suggests that region specific factors, some of which can be influenced directly by policy, are constraining regional wage convergence. The analysis further shows that the findings from the econometric analysis are robust to exclusion of outlier regions, use of an alternative sample of workers and measure of regional disparity, income per capita, as well as use of panel data models.

The final chapter (Chapter 5) empirically tests whether the prediction of a wage equation derived from the new economic geography (NEG) theory is consistent with regional wage disparities in South Africa. A structural wage equation based on the Helpman-Hanson model derived directly from the NEG theory is estimated to test the coherence of the NEG theory with South African data. The chapter further extends the Helpman-Hanson model to include other potential explanatory factors concerning regional wages. The study reveals new insights on the applicability of the NEG theory in emerging economies.

The analysis provides evidence of a highly significant and positive relationship between regional wages and market potential, as postulated by the NEG theory, over the period 1996 – 2011. Regional wages in South Africa, therefore, rise with the increasing market potential of the region. In addition, the study finds evidence of highly significant and theoretically consistent structural parameters (σ, μ and τ) of the Helpman-Hanson model. In spite of this evidence, a key condition critical in the validation of the Helpman-Hanson model, the no black hole condition, $\sigma(1-\mu) < 1$, is rejected for 1996 and 2001. This suggests that the case of South Africa is not fully explained by the mechanisms emphasised by the Helpman-Hanson model. The results show that regional wage disparities in South Africa are well explained by NEG forces such as access to markets, but only after controlling for regional specific factors such as human capital, mineral resource endowments, local climatic conditions, local unemployment, and homeland status. The findings of this study imply that NEG theory alone is not sufficient to explain regional wage disparities in South Africa. Its proper application hinges on the incorporation of other regional specific factors. This finding is consistent with an emerging economy that is characterised by moderate industrial and transport sectors, on the one hand, and a strong primary sector driven by natural resource exploitation, on the other hand.

In sum, the findings of this thesis provide important and new insights concerning regional wage disparities in South Africa. The results show that while regional wage disparities have decreased recently, the rate of convergence (unconditional convergence rate of 3.7% per year) is very low and regional wage disparities remain high in South Africa. The thesis shows that differences in access to markets and regional factor endowments exacerbate and constrain regional wage disparities in South Africa.

6.2. Policy implications of findings

The findings of this thesis have the potential to help guide policy efforts to address regional wage disparities and overall regional economic development in South Africa. In more specific terms, the findings in chapter three reveal that regional economies are interdependent. This implies that political and policy decisions of local government authorities affect not only their own regions but neighbouring regions as well. Accordingly, it is necessary for policy initiatives aimed at promoting regional economic development and addressing regional wage disparities to take into account regional interdependence, as the effect of those initiatives will spill over to neighbouring regions. As a result, central government needs to play an active role in coordinating regional development initiatives such as infrastructure development, to maximise resource use and avoid duplication of interventions by neighbouring regions.

Furthermore, evidence from this thesis of a spatially heterogeneous economy characterised by four spatial cluster regimes (H-H, L-L, H-L and L-H), highlights the need for implementation of geographically targeted interventions to address location-specific needs. This is because these regimes contain regions with different needs and challenges which in turn require targeted policy interventions. Of these regimes, there is a need for regional policy to pay close attention to the L-L and H-L spatial regimes, as these regimes can act as a development barrier not only for regions in these geographical locations but also for the nation as a whole. For example, the L-L geographical cluster might lead to a development trap, as poor regions surrounded by other poor regions are more likely to find it difficult to attract economic activities to propel them out of poverty. On the other hand, the H-L geographical cluster might also lead to a development trap, as more prosperous regions surrounded by poor regions can act as local growth poles, absorbing economic activity from the nearby poor regions. Thus, active policy is needed to positively influence economic outcomes of regions in these regimes to enable them to escape the poverty trap and ensure full utilisation of the country’s resources.

The evidence in chapter four of regional wage convergence over the 2001-2011 period offers an optimistic and encouraging message to policymakers concerned with promoting regional economic development and addressing regional wage disparities in South Africa. However, a key message from the chapter is that, while regional wage disparities are converging over time in South Africa, they remain high compared to other countries. Convergence is very slow and is influenced by regional specific factors such as human capital, access to markets, population density, mineral resource endowments, local climatic conditions, local unemployment and homeland status. Accordingly, the thesis lends strong support to policy initiatives aimed at improving the underlying conditions of lagging regions, especially for regions in former homeland areas. To enhance regional wage convergence, the thesis supports policy measures aimed at promoting human capital accumulation, greater access to markets, as well as addressing the problems of high unemployment, population density and low productivity in the agricultural sector.

Nevertheless, even with these policies, as shown in Chapter 5, regional wage disparities will remain a feature of the South African spatial economy as these are, in part, driven by economic forces associated with new economic geography. It was shown that economic forces associated with greater access to markets such as transport costs, increasing returns to scale and consumer’s love of variety promote agglomeration of economic activities, leading to regional wage disparities. This is because economic agents are reluctant to relocate from large urban areas like Johannesburg and Cape Town with good access to markets, despite high congestion costs in these areas. This has important implications for urban planning as the benefits associated with greater access to markets in urban areas is likely to promote rural to urban migration. Apart from increasing competition in the urban labour market, such migration can place increasing pressure on local government's ability to respond to social service needs of urban populations. To ease the migrant pressure, there is need for formation of primary and secondary cities to act as local growth poles, absorbing some of the rural migrants.

6.3. Suggestions for future research

Future research could build on the important contribution made by this thesis to advancing the broad research agenda on regional wage disparities in the context of African economies. For this study, regional income per worker was used to proxy regional wage per worker, but the scope of future research will depend largely on the availability of reliable regional wage data.

Based on empirical evidence from this thesis, the first extension is the study of the spatial patterns that characterise the distribution of other labour market outcomes to fully understand the South African spatial economy. In addition, future research utilising data covering longer time periods may be more informative, given that the regional wage convergence should be looked at from a long-term perspective. Panel data models can also be employed in the study of regional wage convergence dynamics to appropriately address the problem of omitted variable bias.

To ensure proper applicability of the NEG theory in emerging economies like South Africa, the study proposes future research that extends the NEG theory to incorporate additional explanatory factors unique to emerging economies, as done by Paillacar (2006) for Brazil. Moreover, to fully understand the mechanisms emphasised by the NEG theory, we suggest future research that tests other insights derived from the theory, such as the relationship between market potential and factor inflows (labour and capital migration). Given the important role that distance play in the NEG theory as a proxy for transport costs, the thesis also recommends further analysis using a better measure of transport costs, such as travel time distance which is more informative for policy.

Finally, while this thesis used magisterial districts as the unit of analysis, there is a need for further research examining spatial wage disparities within magisterial districts, as we see substantial disparities within magisterial districts.

References

Acemoglu, D., & Dell, M. (2010). Productivity differences between and within countries. *American Economic Journal: 2*(1), 169–188.

Acemoglu, D., Johnson, S., & Robinson, J. A. (2002). Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution. *The Quarterly Journal of Economics*, *117*(4), 1231–1294.

African Development Bank. (2012). Briefing Note 5: Income Inequality in Africa. *Briefing Notes for AfDB's Long-Term Strategy*.

African Economic Outlook. (2015). Regional Development and Spatial Inclusion. *African Economic Outlook. Org (Http://www.africaneconomicoutlook.org/).*

Aghion, P., Braun, M., & Fedderke, J. (2008). Competition and productivity growth in South Africa. *Economics of Transition*, *16*(4), 741–768.

Alexiadis, S. (2010). Convergence in agriculture: Evidence from the European Regions. *Agricultural Economics Review*, *11*(2), 84–96.

Allison, P. D. (2001). *Missing Data* (Vol. 136). SAGE Publications.

Altay, H., & Çelebioğlu, F. (2012). Spatial Analysis of Concentration in Production and Trade: An Exploratory Spatial Data Analysis for Emerging Markets. *Journal of Alanya Faculty of Business*, *4*(2), 125–140.

Alvarado, R., & Atienza, M. (2014). The role of market access and human capital in regional wage disparities: Empirical evidence for Ecuador. *Working Paper No. 50. Universidad Catolica Del Norte, Chile, Department of Economics*.

Amaral, P. V, Lemos, M., Simões, R., & Chein, F. (2010). Regional imbalances and market potential in Brazil. *Spatial Economic Analysis*, *5*(4), 463–482.

Amiti, M., & Cameron, L. (2007). Economic geography and wages. *The Review of Economics and Statistics*, *89*(1), 15–29.

Anselin, L. (1988). Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity. *Geographical Analysis*, *20*(1), 1–17.

Anselin, L. (1992). Spatial data analysis with GIS: an introduction to application in the social sciences. *Technical Report 92-10. National Center for Geographic Information and Analysis University of California Santa Barbara*.

Anselin, L. (1995). Local indicators of spatial association—LISA. *Geographical Analysis*, *27*(2), 93–115.

Anselin, L. (1997). The Moran scatterplot as an ESDA tool to assess local instability in spatial association. In . *In Fischer M, Scholten H, Unwin D (eds) Spatial analytical perspectives* *on GIS in environmental and socio-economic sciences. London, Taylor and Francis:111–25*.

Anselin, L., & Bao, S. (1997). Exploratory spatial data analysis linking SpaceStat and ArcView. *Recent Developments in Spatial Analysis: Springer Berlin Heidelberg*, 35–59.

Anselin, L., & Bera, A. K. (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics. *Statistics Textbooks and Monographs*, *155*, 237–290.

Anselin, L., & Florax, R. J. G. M. (1995). Small sample properties of tests for spatial dependence in regression models: Some further results. In *New directions in spatial econometrics* (pp. 21–74). Springer.

Anselin, L., Le Gallo, J., & Jayet, H. (2008). Spatial panel econometrics. In *The econometrics of panel data* (pp. 625–660). Springer.

Anselin, L., Sridharan, S., & Gholston, S. (2007). Using exploratory spatial data analysis to leverage social indicator databases: the discovery of interesting patterns. *Social Indicators Research*, *82*(2), 287–309.

Ardington, C., Lam, D., Leibbrandt, M., & Welch, M. (2006). The sensitivity to key data imputations of recent estimates of income poverty and inequality in South Africa. *Economic Modelling*, *23*(5), 822–835.

Artelaris, P., Arvanitidis, P., & Petrakos, G. (2011). Convergence patterns in the world economy: exploring the nonlinearity hypothesis. *Journal of Economic*, *38*(3), 236–252.

Asongu, S. (2014). African development: beyond income convergence. *South African Journal of Economics*, *82*(3), 334–353.

Bai, C.-E., Ma, H., & Pan, W. (2012). Spatial spillover and regional economic growth in China. *China Economic Review*, *23*(4), 982–990.

Baldwin, R., Forslid, R., & Martin, P. (2005). *Economic geography and public policy*. Princeton, NJ, Princeton University Press.

Barro, R. (2015). Convergence and modernisation. *The Economic Journal*, *125*(585), 911–942.

Barro, R. J., & Sala-i-Martin, X. (1992). Convergence. *Journal of Political Economy*, *100*(2), 223–251.

Barro, R. J., & Sala-i-Martin, X. (2004). *Economic Growth*. *Cambridge, Massachusettes* ((2nd edn). The MIT Press.

Barro, R., & Sala-i-Martin, X. (1991). Convergence across states and regions. *Brookings Papers on Economic Activity*, *1*, 107–182.

Bastos, P., & Bottan, N. (2014). Overcoming the Tyranny of History: Evidence from Post-Apartheid South Africa. *World Bank, Washington, DC Available Online: Http://beta.*

Udep. Edu. pe/empresas/files/2014/07/3E_1_BastosBOTTAN_2. Pdf (Accessed 7 September 2014).

Baumol, W. (1986). Productivity Growth, Convergence, and Welfare: What the Long-Run Data Show. *The American Economic Review*, *76*(5), 1072–1085.

Becker, G. S. (1962). Investment in human capital: A theoretical analysis. *Journal of Political Economy*, *70*(5), 9–49.

Becker, S. O., Boeckh, K., Hainz, C., & Woessmann, L. (2016). The Empire Is Dead, Long Live the Empire! Long-Run Persistence of Trust and Corruption in the Bureaucracy. *The Economic Journal*, *126*(590), 40–74.

Behar, A., & Edwards, L. (2006). Trade liberalization and labour demand within South African Manufacturing firms. *Journal for Studies in Economics and*, *30*(2), 1–20.

Blanchflower, D. G., & Oswald, A. J. (1995). An introduction to the wage curve. *The Journal of Economic Perspectives*, *9*(3), 153–167.

Blanchflower, D. G., & Oswald, A. J. (2005). The wage curve reloaded. *Working Paper No. 11338. National Bureau of Economic Research (NBER).*

Blanchflower, D., & Oswald, A. (1990). The Wage Curve. *Scandinavian Journal of*, *92*(2), 215–235.

Bosker, E. (2008). *The empirical relevance of geographical economics*. PhD thesis, Utrecht University.

Bosker, M., & Garretsen, H. (2010). Trade costs in empirical New Economic Geography. *Papers in Regional Science*, *89*(3), 485–511.

Bosker, M., & Garretsen, H. (2012). Economic geography and economic development in Sub-Saharan Africa. *The World Bank Economic Review*, *26*(3), 443–485.

Bosker, M., & Krugell, W. (2008). Regional Income Evolution in South Africa, *48*(3), 493–523.

Brakman, S., Garretsen, H., & Marrewijk, C. Van. (2009). *The new introduction to geographical economics* (Second). Cambridge University Press.

Brakman, S., Garretsen, H., & Schramm, M. (2004). The spatial distribution of wages: estimating the Helpman-Hanson model for Germany. *Journal of Regional Science*, *44*(3), 437–466.

Breau, S., & Rigby, D. (2009). International trade and wage inequality in Canada. *Journal of Economic Geography*, *10*(1), 55–86.

Breau, S., & Saillant, R. (2016). Regional income disparities in Canada: exploring the geographical dimensions of an old debate. *Regional Studies, Regional Science*, *3*(1), 464–

482.

Breetzke, G. (2008). Exploratory spatial data analysis (ESDA) of violent, economic and sexual offenders in the city of Tshwane, South Africa. *Acta Criminologica*, *1*, 131–149.

Breinlich, H. (2006). The spatial income structure in the European Union—what role for Economic Geography? *Journal of Economic Geography*, *6*(5), 593–617.

Breinlich, H., Ottaviano, G., & Temple, J. (2014). Regional Growth and Regional Decline. *Elsevier*, *2*, 683–779.

Bukenya, J. O., Davis, C., Banerjee, S., & Gyawali, B. (2011). Analysis of regional disparities and wage convergence in Alabama. *African Journal of Agricultural Research*, *6*(2), 363–375.

Burger, P. (2015). Wages, Productivity and Labour's Declining Income Share in Post-Apartheid South Africa. *South African Journal of Economics*, *83*(2), 159–173.

Burger, R., & Yu, D. (2007). Wage trends in post-apartheid South Africa: Constructing an earnings series from household survey data. *DPRU Working Paper 07/117. Development Policy Research Unit.*

Candelaria, C., Daly, M., & Hale, G. (2015). Persistence of Regional Wage Differences in China. *Pacific Economic Review*, *20*(3), 365–387.

Card, D. (1995). The Wage Curve: A review. *Journal of Economic Literature*, *33*(2), 785–799.

Carlino, G. A., & Mills, L. (1996). Testing neoclassical convergence in regional incomes and earnings. *Regional Science and Urban Economics*, *26*(6), 565–590.

Carmignani, F. (2006). The road to regional integration in Africa: Macroeconomic convergence and performance in COMESA. *Journal of African Economies*, *15*(2), 212–250.

Celebioglu, F., & Dall'erba, S. (2010). Spatial disparities across the regions of Turkey: an exploratory spatial data analysis. *The Annals of Regional Science*, *45*(2), 379–400.

Charles, A., Darne, O., & Hoarau, J. (2012). Convergence of real per capita GDP within COMESA countries: A panel unit root evidence. *The Annals of Regional Science*, *49*(1), 53–71.

Chen, Y., Chang, H.-L., & Su, C.-W. (2016). Does real wage converge in China? *Journal of Economic Interaction and Coordination*, *11*(1), 77–93.

Chenery, H. B. (1960). Patterns of industrial growth. *The American Economic Review*, *50*(4), 624–654.

Cherodian, R., & Thirlwall, A. (2015). Regional disparities in per capita income in India: convergence or divergence? *Journal of Post Keynesian Economics*, *37*(3), 384–407.

Cheruiyot, K., Wray, C., & Katumba, S. (2015). Spatial statistical analysis of dissatisfaction with the performance of local government in the Gauteng City-Region, South Africa. *South African Journal of Geomatics*, *4*(3), 224–239.

Cieślik, A., & Rokicki, B. (2016). European integration and spatial wage structure in Poland. *Tijdschrift Voor Economische En Sociale*, *107*(4), 435–453.

Collins, L. M., Schafer, J. L., & Kam, C.-M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods*, *6*(4), 330.

Collins, W. J. (1999). Labor Mobility, Market Integration, and Wage Convergence in Late 19th Century India. *Explorations in Economic History*, *36*(3), 246–277.

Combes, P., Duranton, G., & Gobillon, L. (2008). Spatial wage disparities: Sorting matters! *Journal of Urban Economics*, *63*(2), 723–742.

Coombes, M. (1995). Dealing with census geography: principles, practices and possibilities. *Census Users Handbook GeoInformation International, Cambridge*, 111–132.

Cronje, M., & Budlender, D. (2004). Comparing Census 1996 With Census 2001: An Operational Perspective. *Southern African Journal of Demography*, *9*(1), 67–90.

Crozet, M. (2004). Do migrants follow market potentials? An estimation of a new economic geography model. *Journal of Economic Geography*, *4*(4), 439–458.

Crozet, M., & Koenig, P. (2005). The Cohesion vs Growth Tradeoff - Evidence from EU Regions (1980-2000). In *ERSA conference papers. No ersa05p716. European Regional Science Association (ERSA).*

Cuñado, J., & Gracia, F. De. (2006). Real convergence in Africa in the second-half of the 20th century. *Journal of Economics and Business*, *58*(2), 153–167.

Dall'erba, S. (2005). Distribution of regional income and regional funds in Europe 1989--1999: an exploratory spatial data analysis. *The Annals of Regional Science*, *39*(1), 121–148.

Davern, M., Rodin, H., Beebe, T. J., & Call, K. T. (2005). The effect of income question design in health surveys on family income, poverty and eligibility estimates. *Health Services Research*, *40*(5p1), 1534–1552.

Dawkins, C. (2003). Regional development theory: conceptual foundations, classic works, and recent developments. *Journal of Planning Literature*, *18*(2), 131–172.

De Arcangelis, G., & Mion, G. (2002). Spatial Externalities and Empirical Analysis: The case of Italy. *Working Paper No 66. Sapienza University of Rome, CIDEI.*

De Bruyne, K. (2010). Explaining the Location of Economic Activity. Is there a Spatial Employment Structure in Belgium? *International Journal of Economic Issues*, *3*(2), 199–222.

de Sousa, F. L. (2010). Regional Manufacturing Wages: Dancing to the Tune of Trade Shocks. *Discussion Paper 0046. Spatial Economics Research Centre, LSE*.

Deller, S. (2009). Wages, rent, unemployment and amenities. *Journal of Regional Analysis and Policy*, *39*(2), 141–154.

Drukker, D., Peng, H., Prucha, I., & Raciborski, R. (2013). Creating and managing spatial-weighting matrices with the spmat command. *Stata Journal*, *13*(2), 242–286.

Durlauf, S., Johnson, P., & Temple, J. (2005). Growth econometrics. *Handbook of Economic Growth*, *1*, 555–677.

Elhorst, J. P. (2003). Specification and estimation of spatial panel data models. *International Regional Science Review*, *26*(3), 244–268.

Enflo, K., Lundh, C., & Prado, S. (2014). The role of migration in regional wage convergence: Evidence from Sweden 1860–1940. *Explorations in Economic History*, *52*, 93–110.

Estanislau, P., Staduto, J., & Parré, J. L. (2013). Wage Convergence of Agricultural Workers of Brazil: 1992-2009. In *ERSA conference papers. No. ersa13p611. European Regional Science Association*.

Ezcurra, R., Iraizoz, B., Pascual, P., & Rapun, M. (2008). Spatial disparities in the European agriculture: a regional analysis. *Applied Economics*, *40*(13), 1669–1684.

Ezcurra, R., Pascual, P., & Rapún, M. (2007). The spatial distribution of income inequality in the European Union. *Environment and Planning*, *39*(4), 869–890.

Fallah, B. N., Partridge, M. D., & Olfert, M. R. (2011). New economic geography and US metropolitan wage inequality. *Journal of Economic Geography*, *11*(5), 865–895.

Fally, T., Paillacar, R., & Terra, C. (2010). Economic geography and wages in Brazil: Evidence from micro-data. *Journal of Development Economics*, *91*(1), 155–168.

Fedderke, J., & Wollnik, A. (2007). The Spatial Distribution of Manufacturing in South Africa 1970-1996, its Determinants and Policy Implications. *ERSA Working Paper 53, Economic Research Southern Africa, Cape Town.*

Ferens, E. (2015). Evaluation of regional wage convergence in Poland. *Acta Scientiarum Polonorum. Oeconomia*, *14*(4), 25–36.

Fingleton, B., López-Bazo, E., & Lopez-Bazo, E. (2006). Empirical growth models with spatial effects. *Papers in Regional Science*, *85*(2), 177–198.

Finn, A., & Leibbrandt, M. (2013). Mobility and Inequality in the First Three Waves of NIDS. *SALDRU Working Paper Number 120/ NIDS Discussion Paper 2013/2. University of Cape Town*.

Frame, E., De Lannoy, A., Koka, P., & Leibbrandt, M. (2016). Multidimensional Youth

Poverty: Estimating the Youth MPI in South Africa at ward level. *Working Paper Series, Number 189. Southern Africa Labour and Development Research Unit, University of Cape Town*.

Fujita, M. (2007). Towards the new economic geography in the brain power society. *Regional Science and Urban Economics*, *37*(4), 482–490.

Fujita, M., Krugman, P., & Venables, A. (1999). *The spatial economy: cities, regions and international trade*. Cambridge, Massachusetts London, England: MIT press.

Fujita, M., & Mori, T. (2005). Frontiers of the new economic geography. *Papers in Regional Science*, *84*(3), 377–405.

Fujita, M., & Thisse, J.-F. (2009). New Economic Geography: An appraisal on the occasion of Paul Krugman's 2008 Nobel Prize in Economic Sciences. *Regional Science and Urban Economics*, *39*(2), 109–119.

Gallo, J. Le, & Ertur, C. (2003). Exploratory spatial data analysis of the distribution of regional per capita GDP in Europe, 1980− 1995. *Papers in Regional Science*, *82*(2), 175–201.

Geary, R. C. (1954). The contiguity ratio and statistical mapping. *The Incorporated Statistician*, *5*(3), 115–146.

Gelb, S. (2004). An overview of the South African economy. *State of the Nation: South Africa, 2005*, 367–400.

Getis, A., & Ord, J. (1992). The analysis of spatial association by use of distance statistics. *Geographical Analysis*, *24*(3), 189–206.

Gezici, F., & Hewings, G. (2004). Regional convergence and the economic performance of peripheral areas in Turkey. *Review of Urban & Regional*, *16*(2), 113–132.

Ghosh, M. (2008). Economic reforms, growth and regional divergence in India. *Margin: The Journal of Applied Economic Research*, *2*(3), 265–285.

Goodchild, M. F., Anselin, L., & Deichmann, U. (1993). A Framework for the Areal Interpolation of Socioeconomic Data. *Environment and Planning A*, *25*(3), 383–397.

Goodchild, M. F. and Lam, N. S. (1980) 'Lam. 1', *Areal interpolation: a variant of the traditional spatial problem. Geo-Processing*, 1, pp. 297–312.

Goschin, Z. (2007). Spatial and sectoral analysis of productivity-wage dissimilarities in Romania. *Romanian Journal of Regional Science*, *1*(1), 33–44.

Graves, P., Arthur, M., & Sexton, R. (1999). Amenities and the labor earnings function. *Journal of Labor Research*, *20*(3), 367–376.

Gregory, I., & Ell, P. (2006). Error-sensitive historical GIS: Identifying areal interpolation errors in time-series data. *International Journal of Geographical*, *20*(2), 135–152.

Gregory, I. N., Marti-Henneberg, J., & Tapiador, F. J. (2010). Modelling long-term pan-European population change from 1870 to 2000 by using geographical information systems. *Journal of the Royal*, *173*(1), 31–50.

Gries, T., & Naude, W. (2008). Trade and endogenous formation of regions in a developing country. *Review of Development Economics*, *12*(2), 248–275.

Guillain, R., & Gallo, J. Le. (2010). Agglomeration and dispersion of economic activities in and around Paris: an exploratory spatial data analysis. *Environment and Planning B: Planning and Design*, *37*(6), 961–981.

Hakizimana, J., & Geyer, H. (2014). Socio-economic inequality in South Africa according to different disparity indices. *ERSA Conference Papers. No. ersa14p133. European Regional Science Association*.

Hanson, G. (1998). *Market Potential, Increasing Returns, and Geographic Concentration. NBER Working Paper No. 6249 , University of Michigan.*

Hanson, G. (2003). What has happened to wages in Mexico since NAFTA? *Working Paper No 9563. National Bureau of Economic Research*.

Hanson, G. (2005). Market potential, increasing returns and geographic concentration. *Journal of International Economics*, *67*(1), 1–24.

Harris, C. C. D. (1954). The, Market as a Factor in the Localization of Industry in the United States. *Annals of the Association of American Geographers*, *44*(4), 315–348.

Harris, I., Jones, P. D., Osborn, T. J., & Lister, D. H. (2014). Updated high-resolution grids of monthly climatic observations--the CRU TS3. 10 Dataset. *International Journal of Climatology*, *34*(3), 623–642.

Harris, J. R., & Todaro, M. P. (1970). Migration, unemployment and development: a two-sector analysis. *The American Economic Review*, *60*(1), 126–142.

Head, K., & Mayer, T. (2004). The empirics of agglomeration and trade. *Handbook of Regional and Urban Economics*, *4*, 2609–2669.

Heckscher, E. F. (1919). The effect of foreign trade on the distribution of income,Economisk Tidskrift. *In E. F. Heckscher and B. Ohlin (1991), Heckscher–Ohlin Trade Theory. Cambridge, MA: MIT Press; Translated, Edited and Introduced by H. Flam and M. J. Flanders*.

Helpman, E. (1998). The size of regions. *Topics in Public Economics: Theoretical and Applied*, 33–54.

Henderson, V. (2003). The urbanization process and economic growth: The so-what question. *Journal of Economic Growth*, *8*(1), 47–71.

Hering, L., & Poncet, S. (2009). The impact of economic geography on wages: Disentangling the channels of influence. *China Economic Review*, *20*(1), 1–14.

Hering, L., & Poncet, S. (2010). Market access and individual wages: Evidence from China. *The Review of Economics and Statistics*, *92*(1), 145–159.

Hiropoulos, A., & Porter, J. (2014). Visualising property crime in Gauteng: Applying GIS to crime pattern theory. *SA Crime Quarterly*, *47*, 17–28.

Hofer, H., & Wörgötter, A. (1997). Regional Per Capita Income Convergence in Austria. *Regional Studies*, *31*(1), 1–12.

Holl, A. (2012). Market potential and firm-level productivity in Spain. *Journal of Economic Geography*, *12*(6), 1191–1215.

Huang, Q., & Chand, S. (2015). Spatial spillovers of regional wages: Evidence from Chinese provinces. *China Economic Review*, *32*, 97–109.

International Labour Office. (2015). Global Wage Report 2014/15: Wages and income inequality. *International Labour Office – Geneva*.

Islam, N. (2003). What have we learnt from the convergence debate? *Journal of Economic Surveys*, *17*(3), 309–362.

Jeanty, P. (2014). SPWMATRIX: Stata module to generate, import, and export spatial weights. *Statistical Software Components. The Kinder Institute for Urban Research/Hobby Center for the Study of Texas, Rice University.*

Juessen, F. (2009). A distribution dynamics approach to regional GDP convergence in unified Germany. *Empirical Economics*, *37*(3), 627–652.

Khan, F. (2014). Economic Convergence in the African Continent: Closing the Gap. *South African Journal of Economics*, *83*(2), 354–370.

Khomiakova, T. (2008). Spatial analysis of regional divergence in india: income and economic structure perspectives. *The International Journal of Economic Policy Studies*, *3*(7), 2008.

Kim, S. (2008). Spatial inequality and economic development: Theories, facts, and policies. *World Bank Publications*, 133–166.

Kingdon, G., & Knight, J. (2006). How flexible are wages in response to local unemployment in South Africa? *Industrial and Labour Relations Review*, *59*(3), 471–495.

Kiso, T. (2005). Does new economic geography explain the spatial distribution of wages in Japan. *Tokyo: SIRJE*.

Knaap, T. (2006). Trade, location, and wages in the United States. *Regional Science and Urban Economics*, *36*(5), 595–612.

Kok, P. (2002). Migration data from South African censuses: evaluation and suggestions.

HSRC Library: Shelf Number 2006.

Kosfeld, R., & Eckey, H. (2010). Market access, regional price level and wage disparities: the German case. *Jahrbuch Für Regionalwissenschaft*, *30*(2), 105–128.

Krugell, W., Koekemoer, G., & Allison, J. (2005). Convergence or divergence of South African cities and towns? Evidence from kernel density estimates. *Biennial Conference of the Economic Society of South Africa: Development Perspectives: Is Africa Different*.

Krugell, W., & Rankin, N. (2012). Agglomeration and Firm-Level Efficiency in South Africa. *Urban Forum*, *23*(3), 299–318.

Krugman, P. (1991). Increasing Returns and Economic Geography. *The Journal of Political Economy*, *99*(3), 483–499.

Krugman, P., & Venables, A. J. (1995). Globalization and the Inequality of Nations. *The Quarterly Journal of Economics*, *110*(4), 857–880.

Lall, S., & Yilmaz, S. (2001). Regional economic convergence: Do policy instruments make a difference? *The Annals of Regional Science*, *35*(1), 153–166.

Larraz, B., Navarrete, M., & Pavía, J. (2016). Wage Concentration in Spain: A Spatial Analysis. *Applications of Spatial Statistics. Edited by Ming-Chih, H. InTech*.

Laurini, M., Andrade, E., & Pereira, P. V. (2005). Income convergence clubs for Brazilian municipalities: a non-parametric analysis. *Applied Economics*, *37*(18), 2099–2118.

Lee, N., Sissons, P., & Jones, K. (2016). The geography of wage inequality in British cities. *Regional Studies*, *50*(10), 1714–1727.

Leibbrandt, M., Finn, A., & Woolard, I. (2012). Describing and decomposing post-apartheid income inequality in South Africa. *Development Southern Africa*, *29*(1), 19–34.

Leibbrandt, M., Poswell, L., Naidoo, P., Welch, M., & Woolard, I. (2005). Measuring Recent Changes in South African Inequality and Poverty using 1996 and 2001 Census Data. *In: Bhorat, H. and R. Kanbur (Eds.) Poverty and Policy in Post-Apartheid South Africa, Pretoria, HSRC Press*, 1–51.

LeSage, J. P. (1999). The theory and practice of spatial econometrics. *University of Toledo. Toledo, Ohio*, *28*, 11.

Lewis, W. A. (1954). Economic development with unlimited supplies of labour. *The Manchester School*, *22*(2), 139–191.

Lindley, J., & Machin, S. (2013). Wage inequality in the Labour years. *Oxford Review of Economic Policy*, *29*(1), 165–177.

Lindley, J., & Machin, S. (2014). Spatial changes in labour market inequality. *Journal of Urban Economics*, *79*, 121–138.

Liu, Y., & Yamauchi, F. (2014). Population density, migration, and the returns to human capital and land: Insights from Indonesia. *Food Policy*, *48*, 182–193.

Longhi, S., Nijkamp, P., & Poot, J. (2006). Spatial heterogeneity and the wage curve revisited. *Journal of Regional Science*, *46*(4), 707–731.

López-Rodríguez, J., & Faíña, J. A. (2006). Does distance matter for determining regional income in the European Union? An approach through the market potential concept. *Applied Economics Letters*, *13*(6), 385–390.

Lucas, R. (1988). On the mechanics of economic development. *Journal of Monetary Economics*, *22*(1), 3–42.

Maantay, J. A., Maroko, A. R., & Porter-Morgan, H. (2008). Research Note—A New Method for Mapping Population and Understanding the Spatial Dynamics of Disease in Urban Areas: Asthma in the Bronx, New York. *Urban Geography*, *29*(7), 724–738. https://doi.org/10.2747/0272-3638.29.7.724

Magrini, S. (2004). Regional (di) convergence. *Handbook of Regional and Urban Economics*, *4*, 2741–2796.

Magruder, J. (2012). High Unemployment Yet Few Small Firms: The Role of Centralized Bargaining in South Africa. *American Economic Journal: Applied Economics*, *4*(3), 138–166.

Mankiw, N., Romer, D., & Weil, D. (1992). A Contribution to the Empirics of Economic Growth. *The Quarterly Journal of Economics*, *107*(2), 407–437.

Martin, D., Dorling, D., & Mitchell, R. (2002). Linking censuses through time: problems and solutions. *Area*.

Martinez-Galarraga, J. (2014). Market potential estimates in history: a survey of methods and an application to Spain, 1867-1930. *Working Paper No. 0051. European Historical Economics Society (EHES).*

Martínez-Galarraga, J., Tirado, D. A., & González-Val, R. (2015). Market potential and regional economic growth in Spain (1860-1930). *European Review of Economic History*, *19*(4), 335–358.

Matthee, M., & Naudé, W. (2008). The determinants of regional manufactured exports from a developing country. *International Regional Science Review*, *31*(4), 343–358.

Maza, A., Hierro, M. M., & Villaverde, J. J. (2012). Income distribution dynamics across European regions: Re-examining the role of space. *Economic Modelling*, *29*(6), 2632–2640.

Maza, A., & Villaverde, J. (2006). A territorial analysis of wage convergence/differentials in

Spain. *Revue D'économie Régionale et Urbaine*, (4), 615–630.

Mazol, A. (2016). Spatial wage inequality in Belarus. *Working Paper No. 35. Belarusian Economic Research and Outreach Center (BEROC).*

McCoskey, S. (2002). Convergence in Sub-Saharan Africa: a nonstationary panel data approach. *Applied Economics, 34*(7), 819–829.

Messner, S., & Anselin, L. (2004). Spatial analyses of homicide with areal data. *In Spatially Integrated Social Science, 127–44, Edited by M. Goodchild and D. Janelle.New York: Oxford University Press.*

Mion, G. (2004). Spatial externalities and empirical analysis: the case of Italy. *Journal of Urban Economics, 56*(1), 97–118.

Mion, G., & Naticchioni, P. (2009). The spatial sorting and matching of skills and firms. *Canadian Journal of Economics/Revue*, *42*(1), 28–55.

Moazzami, B. (1997). Regional wage convergence in Canada: an error-correction approach. *Canadian Journal of Regional Science*, *20*, 341–350.

Monastiriotis, V. (2009). Examining the consistency of spatial association patterns across socio-economic indicators: an application to the Greek regions. *Empirical Economics, 37*(1), 25–49.

Monastiriotis, V. (2014). Regional growth and national development: transition in Central and Eastern Europe and the regional Kuznets curve in the East and the West. *Spatial Economic Analysis*, *9*(2), 142–161.

Moncarz, P. (2007). Regional employment and wages. The effects of transport costs and market potential. An application for Argentina. *Revista de Economía Y Estadística*, *45*(1), 75–108.

Moreno-Monroy, A. (2008). The dynamics of spatial agglomeration in China: an empirical assessment. *Working Paper 08-06, Economics Program.*

Moreno-Monroy, A. (2011). Market access and the heterogeneous effect of shocks on wages: evidence from Chinese cities. *Papers in Regional Science, 90*(1), 9–25.

Naudé, W., & Gries, T. (2009). Explaining regional export performance in a developing country: The role of geography and relative factor endowments. *Regional Studies, 43*(7), 967–979.

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

Naudé, W., & Krugell, W. (2005). The geographical economy of South Africa. *Journal of Development Perspectives Special Edition on Geographical Economics, 1*(1), 85–128.

Naudé, W., & Krugell, W. (2006). Sub-national growth rate differentials in South Africa: an econometric analysis. *Papers in Regional Science*, *85*(3), 443–457.

Naudé, W., Krugell, W., & Matthee, M. (2010). Globalization and Local Economic Growth in South Africa. In *In B. Dallago & C. Guglielmetti (Eds.), Local economies and global competitiveness (pp. 45–73). London: Palgrave MacMillan.* Springer.

Naz, A., Ahmad, N., & Naveed, A. (2017). Wage Convergence across European Regions: Do International Borders Matter? *Journal of Economic Integration*, *32*(1), 35–64.

Nel, E., & Rogerson, C. (2009). Re-thinking spatial inequalities in South Africa: Lessons from international experience. *Urban Forum*, *20*(2), 141–155.

Niebuhr, A. (2006). Market access and regional disparities. *The Annals of Regional Science*, *40*(2), 313–334.

Niebuhr, A., Granato, N., Haas, A., & Hamann, S. (2012). Does labour mobility reduce disparities between regional labour markets in Germany? *Regional Studies*, *46*(7), 841–585.

Noble, M., & Wright, G. (2013). Using indicators of multiple deprivation to demonstrate the spatial legacy of apartheid in South Africa. *Social Indicators Research*, *112*(1), 187–201.

Noble, M., Zembe, W., & Wright, G. (2014). Poverty may have declined, but deprivation and poverty are still worst in the former homelands. *Southern African Social Policy Research Institute*.

Ntuli, M., & Kwenda, P. (2014). Labour unions and wage inequality among African men in South Africa. *Development Southern Africa*, *31*(2), 322–346.

Nunn, N. (2009). The importance of history for economic development. *Annu. Rev. Econ.*, *1*(1), 65–92.

O'Rourke, K. H., & Williamson, J. G. (1999). Globalization and history: The evolution of a 19th century atlantic economy. *MIT Press*, *6*, 100–105.

Ohlin, B. (1933). *International and interregional trade*. *Harvard Economic Studies, Cambridge, MA.*

Ord, J., & Getis, A. (1995). Local spatial autocorrelation statistics: distributional issues and an application. *Geographical Analysis*, *27*(4), 286–306.

Oshchepkov, A. (2007). Are Interregional Wage Differentials in Russia Compensative? *Discussion Paper No750.DIW Berlin, German Institute for Economic Research*.

Oshchepkov, A. (2015). *Compensating wage differentials across Russian regions*. *In: Mussida C, Pastore BK (eds) Geographical labor market imbalances. recent explanations and cures. Iss AIEL series in labour economics. Springer, Berlin. pp 65-105*.

Ottaviano, G. I. P. P., & Pinelli, D. (2006). Market potential and productivity: Evidence from Finnish regions. *Regional Science and Urban Economics*, *36*(5), 636–657.

Overman, H., Redding, S., & Venables, A. (2003). The Economic Geography of Trade Production and Income: A Survey of Empirics. *In: Kwan-Choi, E., Harrigan, J. (Eds.), Handbook of International Trade. Basil Blackwell, Oxford, Pp. 353 – 387*.

Paillacar, R. (2006). Market potential and worker heterogeneity as determinants of Brazilian wages. *University of Paris. Unpublished Manuscript.*

Paluzie, E., Pons, J., & Tirado, D. (2009). A test of the market potential equation in Spain. *Applied Economics*, *41*(12), 1487–1493.

Paredes, D. (2013). The role of human capital, market potential and natural amenities in understanding spatial wage disparities in Chile. *Spatial Economic Analysis, 8*(2), 154–175.

Paredes, D. (2015). Can NEG explain the spatial distribution of wages of Chile? *Tijdschrift Voor Economische En Sociale Geografie*, *106*(1), 65–77.

Paredes, D., & Iturra, V. (2012). Market access and wages: A spatially heterogeneous approach. *Economics Letters*, *116*(3), 349–353.

Partridge, M. D., Rickman, D. S., Ali, K., & Olfert, M. R. (2010). Recent spatial growth dynamics in wages and housing costs: Proximity to urban production externalities and consumer amenities. *Regional Science and Urban Economics*, *40*(6), 440–452.

Patacchini, E. (2008). Local analysis of economic disparities in Italy: a spatial statistics approach. *Statistical Methods and Applications*, *17*(1), 85–112.

Patacchini, E., & Rice, P. (2007). Geography and economic performance: exploratory spatial data analysis for Great Britain. *Regional Studies*, *41*(4), 489–508.

Pires, A. (2006). Estimating Krugman’s economic geography model for the Spanish regions. *Spanish Economic Review*, *6*(2), 83–112.

Pons, J., Paluzie, E., Silvestre, J., & Tirado, D. A. (2007). Testing the new economic geography: migrations and industrial agglomerations in Spain. *Journal of Regional Science*, *47*(2), 289–313.

Posel, D., & Casale, D. (2005). “Who Replies in Brackets and what are the Implications for Earnings Estimates?: An Analysis of Earnings Data from South Afria.” *Working Paper No. 07,Economic Research Southern Africa*.

Quah, D. (1993). Empirical cross-section dynamics in economic growth. *European Economic Review*, *37*(2–3), 426–434.

Quah, D. (1996a). Convergence empirics across economies with (some) capital mobility.

Journal of Economic Growth, *1*(1), 95–124.

Quah, D. (1996b). Empirics for economic growth and convergence. *European Economic Review*, *40*(6), 1353–1375.

Quah, D. (1996c). Twin Peaks: Growth and Convergence in Models of Distribution Dynamics. *The Economic Journal*, *106*(437), 1045–1055.

Ramos, R., Nicodemo, C., & Sanromá, E. (2015). A spatial panel wage curve for Spain. *In Spatial and Resource Sciences*, *8*(2), 25–139.

Ravallion, M., & Chen, S. (2012). Monitoring inequality. *Blog Post on Let's Talk Development, World Bank. Http://blogs. Worldbank. Org/developmenttalk/monitoringinequality*.

Raymond, M. R., & Roberts, D. M. (1987). A comparison of methods for treating incomplete data in selection research. *Educational and Psychological Measurement*, *47*(1), 13–26.

Redding, S. (2005). Spatial income inequality. *Swedish Economic Policy Review*, *12*(1), 29–55.

Redding, S. (2010). The empirics of new economic geography. *Journal of Regional Science*, *50*(1), 297–311.

Redding, S. J. (2013). Economic Geography: A a review of the theoretical and empirical literature. *In Palgrave Handbook of International Trade. Palgrave Macmillan UK*, 497–531.

Redding, S., & Venables, A. J. (2004). Economic geography and international inequality. *Journal of International Economics*, *62*(1), 53–82.

Reibel, M., & Agrawal, A. (2007). Areal interpolation of population counts using pre-classified land cover data. *Population Research and Policy Review*, *26*(5–6), 619–633.

Rey, S., & Montouri, B. (1999). US regional income convergence: a spatial econometric perspective. *Regional Studies*, *33*(2), 143–156.

Roback, J. (1982). Wages, rents, and the quality of life. *Journal of Political Economy, 90*(6), 1257–1278.

Roback, J. (1988). Wages, rents, and amenities: differences among workers and regions. *Economic Inquiry*, *26*(1), 23–41.

Rodríguez-Pose, A., & Tselios, V. (2011). Mapping the European regional educational distribution. *European Urban and Regional …*, *18*(4), 354–374.

Romer, P. (1986). Increasing returns and long-run growth. *Journal of Political Economy, 94*(5), 1002–1037.

Roos, M. (2001). Wages and market potential in Germany. *Jahrbuch Für Regionalwissenschaft*, *21*, 171–195.

Rosés, J. R., & Sánchez-Alonso, B. (2004). Regional wage convergence in Spain 1850--1930. *Explorations in Economic History*, *41*(4), 404–425.

Roth, P. L. (1994). Missing data: A conceptual review for applied psychologists. *Personnel Psychology, 47*(3), 537–560.

Rubin, D. B. (1987). Multiple imputation for nonresponse in surveys. *New York: Wiley*.

Rusche, K. (2010). Quality of life in the regions: an exploratory spatial data analysis for West German labor markets. *Jahrbuch Für Regionalwissenschaft*, *30*(1), 1–22.

Rybczynski, T. (1955). Factor endowment and relative commodity prices. *Economica*, *22*(88), 336–341.

Sakamoto, H., & Islam, N. (2008). Convergence across Chinese provinces: an analysis using Markov transition matrix. *China Economic Review*, *19*(1), 66–79.

Samuelson, P. (1949). International factor-price equalisation once again. *The Economic Journal*, *59*(234), 181–197.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. New York: Chapman & Hall.

Schafer, J. L. (1999). Multiple imputation: a primer. *Statistical Methods in Medical Research*, *8*(1), 3–15.

Shimeles, A., & Nabassaga, T. (2015). Why is inequality high in Africa? *Working Paper No. 246. African Development Bank*, *23*.

Silverman, B. (1986). Density estimation for statistics and data analysis. *London: Chapman and Hall.*

Šlander, S., & Ogorevc, M. (2010). Labour Cost Convergence in the EU: Spatial Econometrics Approach. *Privredna Kretanja I Ekonomska Politika*, *20*(122), 27–52.

Solarin, S., & Sahu, P. (2013). Convergence or divergence in CFA franc countries: A time series analysis. *The Journal of Applied Business and*, *14*(2), 112–127.

Solow, R. (1956). A Contribution to the Theory of Economic Growth. *The Quarterly Journal of Economics*, *70*(1), 65–95.

Statistics South Africa. (2012). Census 2011 Census in brief. *Report No. 03-01-41, Published by Statistics South Africa, Pretoria*.

Statistics South Africa, & Sciences Research Council Human. (2007). Using the 2001 Census Approaches to analysing data, A collaboration between Statistics South Africa (Stats SA) and the Human Sciences Research Council (HSRC).

Swan, T. (1956). Economic growth and capital accumulation. *Economic Record*, *32*(2), 334–361.

Tavernier, E., & Temel, T. (1997). National and regional analysis of convergence of real wages in the US agricultural sector. *Journal of Regional Analysis*, *27*(1), 63–74.

Tobler, W. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, *46*, 234–240.

Turgut, M. B. (2014). Regional economic activity in Turkey: A new economic geography approach. *Discussion Paper No. 2014/5, Turkish Economic Association*.

Vakulenko, E. (2016). Does migration lead to regional convergence in Russia? *International Journal of Economic Policy in Emerging Economies*, *9*(1), 1–25.

Venables, A. (1996). Equilibrium locations of vertically linked industries. *International Economic Review*, *37*(2), 341–359.

von Fintel, D. (2007). Dealing with Earnings Bracket Responses in Household Surveys–How Sharp are Midpoint Imputations? *South African Journal of Economics*, *75*(2), 293–312.

Von Fintel, D. (2014). Spatial heterogeneity, generational change and childhood socioeconomic status: microeconometric solutions to South African labour market questions. *Doctoral Dissertation: Stellenbosch University*.

Von Fintel, D. (2015). Wage flexibility in a high unemployment regime: spatial heterogeneity and the size of local labour markets. REDI3x3 Working Paper 8. Research Project on Employment, Income Distribution and Inclusive Growth, Cape Town.

Von Fintel, D. (2017). Institutional wage-setting, labour demand and labour supply: Causal estimates from a South African pseudo-panel. *Development Southern Africa*, *34*(1), 1–16.

Webber, D. (2001). Convergence of labour's factor reward between regions of the EU. *Applied Economics Letters*, *8*(5), 355–357.

Weir-Smith, G. (2015). *Unemployment in South Africa: In Search of a Spatial Model*. Doctoral Dissertation. University of KwaZulu-Natal, Durban.

Weir-Smith and Ahmed. (2013). Unemployment in South Africa: Building a Spatio-temporal Understanding. *South African Journal of Geomatics*, *2*(3), 218–230.

Williamson, J. G. (1998). Real wages and relative factor prices in the Third world 1820-1940: Latin America. *Working Paper No. 1853. Harvard Institute of Economic Research, Harvard University*.

Willis, R. J. (1986). Wage determinants: A survey and reinterpretation of human capital earnings functions. *Handbook of Labor Economics, Orley Ashenfelter and David Card, Eds. New York: North-Holland*, *1*, 525–602.

Wittenberg, M. (2014). Analysis of employment, real wage, and productivity trends in South Africa since 1994. *Series No. 994847703402676. International Labour Organization*.

Wittenberg, M. (2016). Wages and Wage Inequality in South Africa 1994–2011: Part 1–Wage

Measurement and Trends. *South African Journal of Economics*, *0*(0), 1–21.

Wittenberg, M. (2017). Wages and Wage Inequality in South Africa 1994-2011: Part 2 - Inequality Measurement and Trends. *South African Journal of Economics*, *85*(2), 298–318.

Wittenberg, M., & Pirouz, F. (2013). The measurement of earnings in the post-Apartheid period: An overview. *Working Paper No 108. Southern Africa Labour and Development Research Unit, University of Cape Town*.

World Bank. (2009). Reshaping Economic Geography: World development report. *The World Bank, Washington DC*.

Young, A., Higgins, M., & Levy, D. (2008). Sigma convergence versus beta convergence: Evidence from US county-level data. *Journal of Money, Credit and*, *40*(5), 1083–1093.

Zalk, N. (2014). Markups in South African Manufacturing-Are they high and what can they tell us? In *TIPS. Manufacturing Led Growth for Employment and Equality. Conference papers. Pretoria*.

Zaman, G., & Goschin, Z. (2014). Economic crisis and wage divergence: empirical evidence from Romania. *Prague Economic Papers*, *23*(4), 493–513.

Appendices

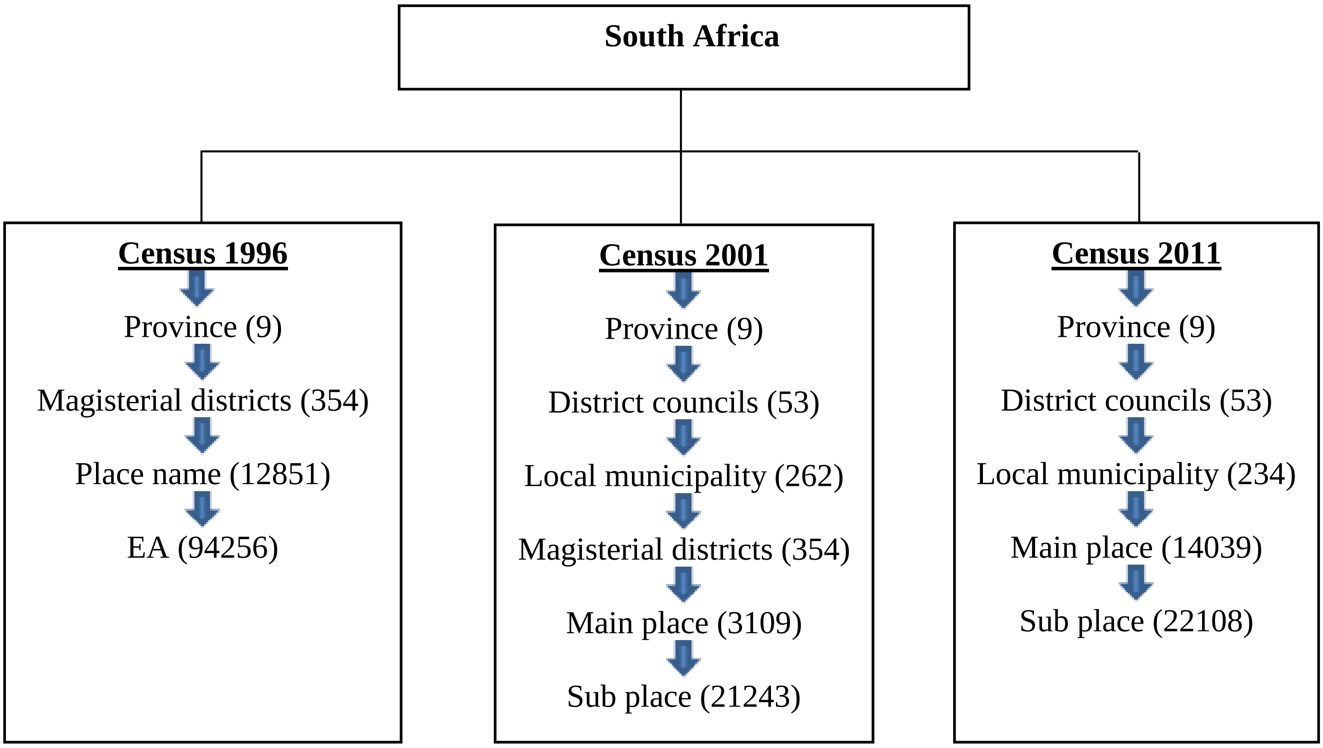
Appendix for Chapter 3

Appendix 3.1: Creating a geographically consistent database from censuses data.

Space and time are the two fundamental organising principles for any population census information (Coombes, 1995), as they enable comparison of economic outcomes across different geographical units over time. To allow such comparison geographical units at which economic data is aggregated to, need to be comparable over time. In other words, the geographical boundaries should not change over time. This is a big challenge, as administrative boundaries at which economic data is collected in many countries change over time in line with government policies. This is the case for South Africa, where re-demarcation of its administrative boundaries has been and continues to be part of government's regional policy initiatives aimed at re-dressing huge spatial imbalances created by years of apartheid-era policies. While the greatest advantage of the censuses is the availability of information at different geographical levels, as shown in Figure 3.1A, a major challenge is the incomparability of the various geographical units across the censuses over time.

For example, re-demarcation of the administrative boundaries between 2001 and 2011 saw the number of municipal units decreasing from 262 in 2001 to 234 2011, while main places and sub places increased from 3109 and 21243 units in 2001 to 14039 and 22108 units in 2011 respectively. Figure 3.2A provides an illustration of the changes in municipal boundaries between 2001 and 2011, where the grey boxed areas represent municipalities in 2001 that were merged with other municipalities in 2011. Furthermore, while magisterial districts were available in 1996 and 2001, they were dropped completely in 2011. The problem of changing boundaries over time, which is referred to in the literature as the modifiable areal unit problem (MAUP), compromises the time dimension principle for the collected census data and poses a challenge for longitudinal comparison of census data. Stats SA acknowledges that the inconsistencies of these geographical units over time complicate the work of researchers and academics who wish to do comparative studies using censuses data (Statistics South Africa & Sciences Research Council Human, 2007).

Figure 3.1A: Geographical hierarchy for various censuses in South Africa.

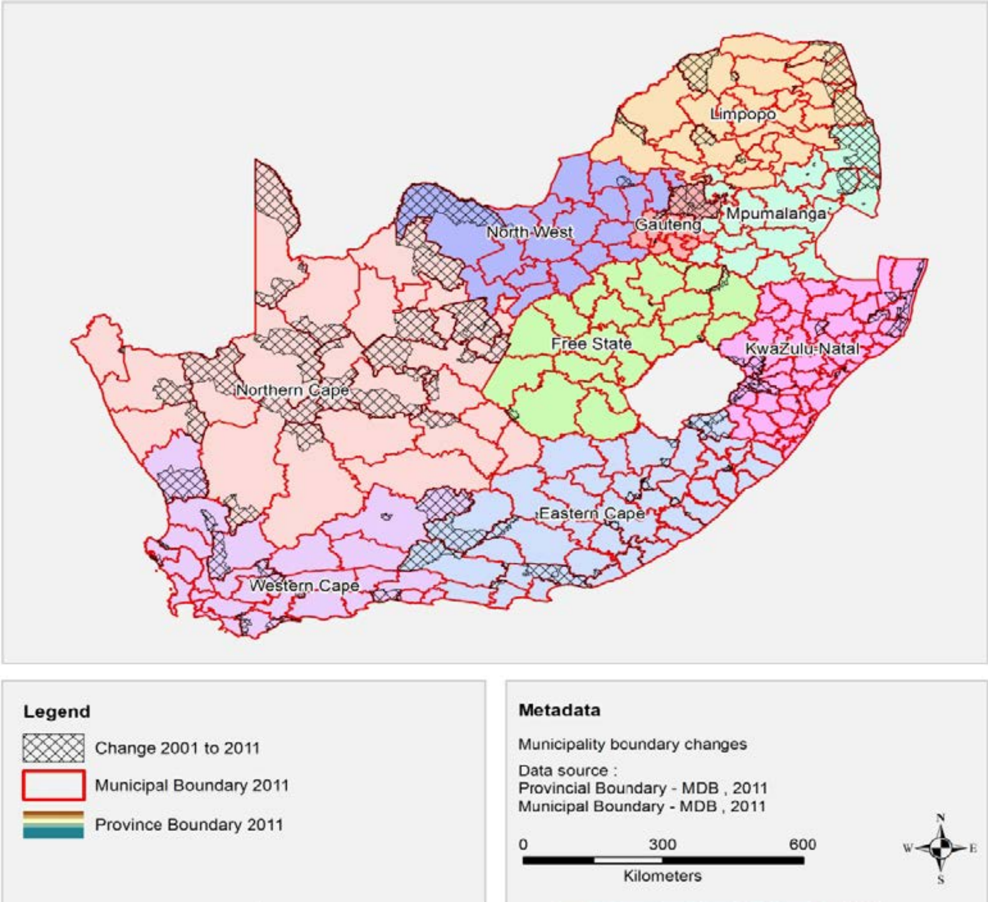


Source: Own construction based on the geographical information in the censuses.

To resolve the problem of gradually changing boundaries of the geographical units at which census data is collected at in South Africa, an innovation of this chapter is the construction of a spatially disaggregated and geographically consistent database from the 1996, 2001 and 2011 censuses data using areal interpolation techniques. These techniques allow us to transfer population (attributes) data from one set of administrative units that were used to publish the data onto an overlapping but incompatible set of geographical units (Gregory, Marti-Henneberg, & Tapiador, 2010)[96]. The first set of geographical units for which the population values are known are often referred to as the source zone units, while the second set for which the population values need to be transferred to are known as the target zone units (Reibel and Agrawal, 2007). While a number of areal interpolation techniques have been proposed in the literature, the most commonly used, which we also apply in this chapter, is the areal-weighting technique (Goodchild & Lam, 1980; Gregory, Marti-Henneberg, & Tapiador, 2010).

[96] We use the word population to generally refer to a given socio-economic variable of interest from the censuses.

Figure 3.2A: Municipal boundary changes between 2001 and 2011.



Source: 2011 Census in brief (Statistics South Africa, 2012)

Based on the assumption that population is homogeneously distributed across the source zones (Goodchild, Anselin, & Deichmann, 1993), areal-weighting involves transfer of population data from the source zone units to the target zone units based on the proportion of the source zone's area contained within the common area between the source and target zone units (Reibel & Agrawal, 2007). Over the years, a growing number of studies have used areal-weighting interpolation to create geographically consistent databases, among them Goodchild & Lam (1980), Martin, Dorling, & Mitchell. (2002) and Gregory & Ell (2006)[97]. In the case of South Africa, areal-weighting interpolation has been applied by Weir-Smith (2013), as well as Bastos & Bottan (2014), who used it to create consistent geographical units based on the South African population censuses data. We follow this literature and use areal-weighting interpolation to

[97] While, Martin et al. (2002) linked the census data of 1971, 1981 and 1991 for England, Scotland and Wales by aggregating the 1991 data to 1981 Enumeration Districts using a point in polygon technique, Gregory and Ell (2006) interpolated British census data from 1851 to 1930 to the parish boundaries of 1951.

create a geographically consistent database over time using 1996, 2001 and 2011 full population censuses.

To achieve this, we need to choose the geographical units to use as the target and source zone units from the spatial hierarchy for the various censuses. We chose 1996/2001 magisterial district units as the target zone unit, hence our geographical unit of analysis. While von Fintel (2014) suggests that district councils are the most appropriate measure of the local labour market, in this thesis we argue that with labour market outcomes highly dispersed across regions within a given district council, the local labour market should be narrowly defined to account for these differences[98]. Apart from evidence of significant dispersion in various labour market outcomes across magisterial districts, magisterial districts do not vary in geographical size as much as the other possible geographical units (such as municipal units), and are sufficiently large to allow the identification of spatial patterns and trends in various socio-economic outcomes (Kok, 2002).

Furthermore, magisterial districts closely define the location of cities and towns in South Africa. It is in these cities and towns where most economic activities take place and it is our belief that magisterial districts correspond with an optimal functional economic entity at which the local labour market can be optimally measured[99]. As the source zone unit from which to transfer 2011 population data to the magisterial district level, we use 2011 sub-place units. Our choice of using sub-place units is motivated by existing literature which acknowledges that using the smallest feasible geographical unit as the source zone unit minimises the error associated with the highly restrictive assumption underlying the areal-weighting interpolation technique of homogeneously distributed population across the source zone units (Maantay, Maroko, & Porter-Morgan, 2008).

Having defined the source (sub-place) and target (magisterial district) zone units, we use the ArcGIS union overlay tool to overlay and intersect the incompatible sub-place and magisterial district zonal units to get a set of union zone units common to both sub-place and magisterial district zone units. Using the resulting union zone areas, we fractionally assign the population values of each sub place zone unit to its corresponding union zone unit based on the proportion

[98] The factor that the sampling framework of the Labour Force Survey which Von Fintel's (2014) used is designed to be representative at the district council level rather than magisterial districts level might have influenced his results.
[99] The appropriateness of magisterial districts has seen it being used in most sub-national studies in South Africa (see, Naude and Krugell, 2003; 2005; Krugell, 2005; Bosker and Krugell, 2008; Magruder, 2012 among others).

of each sub place zone area occupied by the union zone unit. This allocation process can be summarised as follows (Goodchild & Lam, 1980):

$$y_i = \sum_r \frac{A_{ir}}{A_r} y_r \tag{1A}$$

where y_i gives the predicted population value of each union zone unit, y_r is the population value for each sub-place zone unit, A_r is the corresponding area of the sub-place zone and A_{ir} is the area of the union between the sub-place and magisterial district zone units. Given that the sub-place zone units can be split into many union zones of the sub-place-magisterial district zones, the predicted population values of each union zone unit (y_i) are summed up and aggregated to their corresponding magisterial district units to complete the interpolation process. At this stage, we get a geographically consistent database across the three censuses geo-referenced to 354 magisterial district units.

Apart from mapping 2011 sub-place population data to 1996/2001 magisterial district boundaries, we also utilise the ArcGIS union overlay tool to construct a homeland status variable. We use this variable in chapters 4 and 5 of this thesis as a control for the effects of a key historical event, namely the apartheid-era system. We create this variable by mapping magisterial district boundaries to homeland area boundaries to derive the union zone units common to magisterial district and homeland area units. From the resulting mapping, we use areal-weighting to derive a variable based on the area of each magisterial district that falls in each homeland area using the following equation:

$$y_h = \sum_r \frac{A_{ir}}{A_r} \tag{2A}$$

where y_h is the proportion of the area of each magisterial district that falls into the union area common between magisterial district and homeland area; A_r is the corresponding area of each magisterial district and A_{ir} is the area of the union between magisterial district and homeland area units. Given the fragmentation of the homeland areas, each magisterial district can be split into many homeland areas. Thus, we sum the fractional ratios of each magisterial district so that each magisterial district will have a ratio between 0 and 1, with 1 highlighting that it falls completely in a given homeland area and 0 otherwise.

A brief look at the results from the areal-weighting interpolation process based on equation (1A)[100], shows that having started with 354 magisterial district units and 22108 sub-place units, an overlay and intersection of these two zone units produced 279366 union zone units, with an areal-weighting ratio of between 0 and 1. We provide a summary of the distribution of the areal weighting ratio across the 27366 union zone units in Table 3.1A. From the areal-weighting process, the ideal ratio is 1, which shows that a given sub-place falls completely in a given magisterial district.

Of the 27366 union zones, 18277 union zones which account for 12.37% of the country's total area and a massive 71.47% of the country's total employment have a ratio of 1. These percentages highlight a key stylised fact, that of the significant spatial concentration of economic activities in a few places in South Africa, as about 12% of the country's area accounts for more than 70% of the country's employment. If we are to consider that union zones with a ratio of at least 0.7 lead to a reasonable prediction of the population data, we see that 21556 union zones have a ratio of between 0.7 and 1. These union zones account for 51.07% of the country's total area and a massive 92.02 percent of the country's total employment. On the other end, 3828 union zones which account for 3.94% of the country's total area and 0.85% of the country's total employment have a ratio of between 0 and 0.09.

Taken together, these union zones account for 1219067 square km of South Africa's total land area of 1219602 square km as of 2001 (99. 95 percent). Further, the union zone areas account of 13156028 of South Africa's total employment of 13179825 in 2011 (99.83 percent)[101]. Thus, the areal-weighting interpolation process fails to account for 535 square km of the land and 23797 employment. This translates to prediction errors of less than 1% (0.044% for total area and 0.181% for total employment), which can be explained largely by lost area due to non-union zone units. Our results show that 644 sub-place portions accounting for 728 square km of South Africa's total area (0.0597%) and 20170 of South Africa's employment (0.00153%)

[100] While we do not provide a discussion of the results of the homeland status variable, we see that out of the 354 magisterial districts, 188 magisterial districts have a portion of their area falling in a given former homeland area. However, of these districts, only 5 districts fall completely in a given former homeland area (have a ratio of 1), while 80 districts have at least 75 percent of their total area falling in a given former homeland area. On the other hand, 166 magisterial districts do not belong to any former homeland area. This distribution shows the fragmented nature of former homeland areas.

[101] While we focus on employment only in this discussion, the same conclusions reached for employment also holds for other population variables.

did not intersect with a given magisterial district unit. These portions contain information lost, hence the error (non-union error) associated with the interpolation process[102].

Table 3.1A: Distribution of the areal-weighting ratio across the union zone units.

Area-weighted ratio	Obs	Obs %	Area accounted (Sq km)	Area %	Employment	Employment %
1	18277	66.79	150915	12.37	9419892	71.47
0.9 – 0.99	2386	8.72	202068	16.57	1996351	15.15
0.8 – 0.89	613	2.24	142766	11.71	462094	3.51
0.7 – 0.79	280	1.02	127062	10.42	249916	1.90
0.6 – 0.69	258	0.94	62716	5.14	210778	1.60
0.5 – 0.59	218	0.8	102490	8.4	166225	1.26
0.4 – 0.49	227	0.83	85839	7.04	144240	1.09
0.3 – 0.39	274	1	111745	9.16	155130	1.18
0.2 – 0.29	329	1.2	104453	8.56	117088	0.89
0.1 – 0.19	676	2.47	80940	6.64	122806	0.93
0 – 0.09	3828	13.99	48074	3.94	111508	0.85
Total predicted values	27366	100	1219068	99.95	13156028	99,82
Actual values			1219602		13179825	

Notes: Total predicted values are the estimates we obtain from the areal-weighting interpolation process, while the actual values are the actual population values provided in 2011 census data.

To quantify the extent of this error we present the mean, minimum, maximum, mean absolute percentage error (MAPE) and the root mean squared error (RMSE) statistics of the error distribution of the predicted values for employment in Table 3.2A. The more accurate our areal-weighting interpolation process, the lesser the error would be recorded and this is reflected by error statistics closer to zero.

Table 3.2A: Analysis of the prediction error distribution

Variable	Mean	Min	Max	MAPE	RMSE
Total employment	-1.85	-2630	136	0.24	22.74

Notes: While the other statistics are straightforward, the mean absolute percentage error (MAPE) and the root mean square error (RMSE) are obtained as follows: $MAPE = \sum_{i=1}^{R} \left| P_i - \overline{P}_i) / P_i \right| / n$; $RMSE = \sqrt{\sum_{i=1}^{R} (P_i - \overline{P}_i)^2 / n}$, where P_i is the employment value of zone i, $\overline{P}_i$ is the predicted employment value of zone i and n is the number of union zones in study area.

From the table, the overall error distribution is reasonably symmetrical with a mean close to zero, showing that on average there is underprediction of employment of about 1.85. However,

[102] Acknowledging the lost information due to non-union sub place zones, we remain with a very small unaccounted area of 193 square km and employment of 3627.

it is clear that this mean figure mask significantly deviates from zero across regions. While the maximum error value is 136, the minimum error value is -2630. A look at the zone units with negative errors shows that the bulk of units registering high negative error values are located in major coastal cities and towns in the Eastern Cape, Western Cape and KwaZulu-Natal, while the remaining few are found in national parks in Limpopo and Mpumalanga[103]. Further, the MAPE, which expresses the accuracy of the prediction as a percentage of the error, shows that on average the predicted employment values are off the true values by 0.24 percent. On the other hand, the RMSE, which is the standard deviation of the error distribution, is also of a reasonable size and close to zero compared to findings from other studies in the literature (see Reibel & Agrawal, 2007).

Overall, while interpolation processes will inevitably contain a certain degree of error (Gregory & Ell, 2006), the areal-weighting interpolation process carried out in this chapter goes a long way in providing a solution to the problem of incompatible geographical units in the population censuses in South Africa. Furthermore, the marginal information lost due to non-union sub-places, as well as the little-unaccounted information gives great confidence in the validity of our areal-weighting interpolation results.

Appendix 3.2: Construction of the spatial weight matrix for South Africa.

A spatial weight matrix is a representation of the spatial structure of geographically referenced data. It is a quantification of the spatial relationships that exist among the data attributes in the study area. Because the spatial weights matrix imposes a spatial structure on the data, there is a need for a spatial weight matrix that best reflects how the features in the study area interact with each other. As described in section 3.5, this study utilises a distance-based spatial weight matrix, where the spatial arrangement of the data is captured by an inverse distance function with a distance cut off threshold. We construct this matrix as follows. First, we converted the magisterial district GIS shapefile into Stata format using a Stata user-written command "shp2dta" (Crow, 2006)[104]. Then, using the geographical centres (centroids - latitude and longitude) for each region derived using ArcGIS, we use a Stata user-written command

[103] This evidence seems to suggest that the boundary of South Africa in 2001 (magisterial districts) along the coastal areas was pegged more inland compared to 2011 (sub-place), a fact which seems to explain the increase in South Africa's total area which increased from 1219602 square km in 2001 to 1220813 square km in 2011. The positive errors are mainly a result of rounding off the area weighted ratio.

[104] The shapefile is provided by Stats SA and comes together with the census data. This shapefile is also used to draw maps to show the spatial distribution of regional income per worker.

"spatwmat" developed by Jeanty (2014)[105] to generate a $354\ x\ 354$ inverse-distance spatial weight matrix, with the following structure when we consider only three regions:

$$W = \begin{bmatrix} 0 & w_{12} & w_{13} \\ w_{21} & 0 & w_{23} \\ w_{31} & w_{32} & 0 \end{bmatrix} \tag{3A}$$

where w_{ir} is the inverse distance between two regions i and r, such that when $d_{ir} \leq 205$ kms, $w_{ir} = 1$ and 0 otherwise or when i=r. The weights are standardised ($w^*_{ir} = w_{ir}\ /\ \sum_{r=1}^{n} w_{ir}$), so that the spatial relations of a region with another regions depend on the region's location relative to other regions.

To ensure the robustness of our results, we take different distance cut-off points and decay functions for the inverse distance weight matrix. Furthermore, we construct and utilise two additional spatial weight matrices, the binary contiguity and k-nearest neighbour weight matrices. Under binary contiguity weight matrix, the spatial structure of the data is conditioned on regions sharing a common boundary (Anselin, 1988; LeSage, 1999)[106]. Thus, regions i and r are spatially related if they share a common border and this holds true in equation (3A) when $w_{ir} = 1$ and are not spatially related if they do not share a common border, thus $w_{ir} = 0$. Under the k-nearest neighbours, in capturing the spatial structure of the data, each region is assigned the same number of neighbours (Gallo & Ertur, 2003). The matrix is based on distance and its form is given by equation (1), where $d_{ir} \leq D_i$ (instead of $d_{ir} \leq D_{max}$), with D_i the critical cut-off distance defined for each region i, above which spatial relations are assumed to be negligible (Dall'erba, 2005). Using the converted magisterial district GIS shapefile, we use Stata user-written command "spatwmat" to create the binary contiguity matrix and "swmatrix" to create the k-nearest neighbour weight matrix.

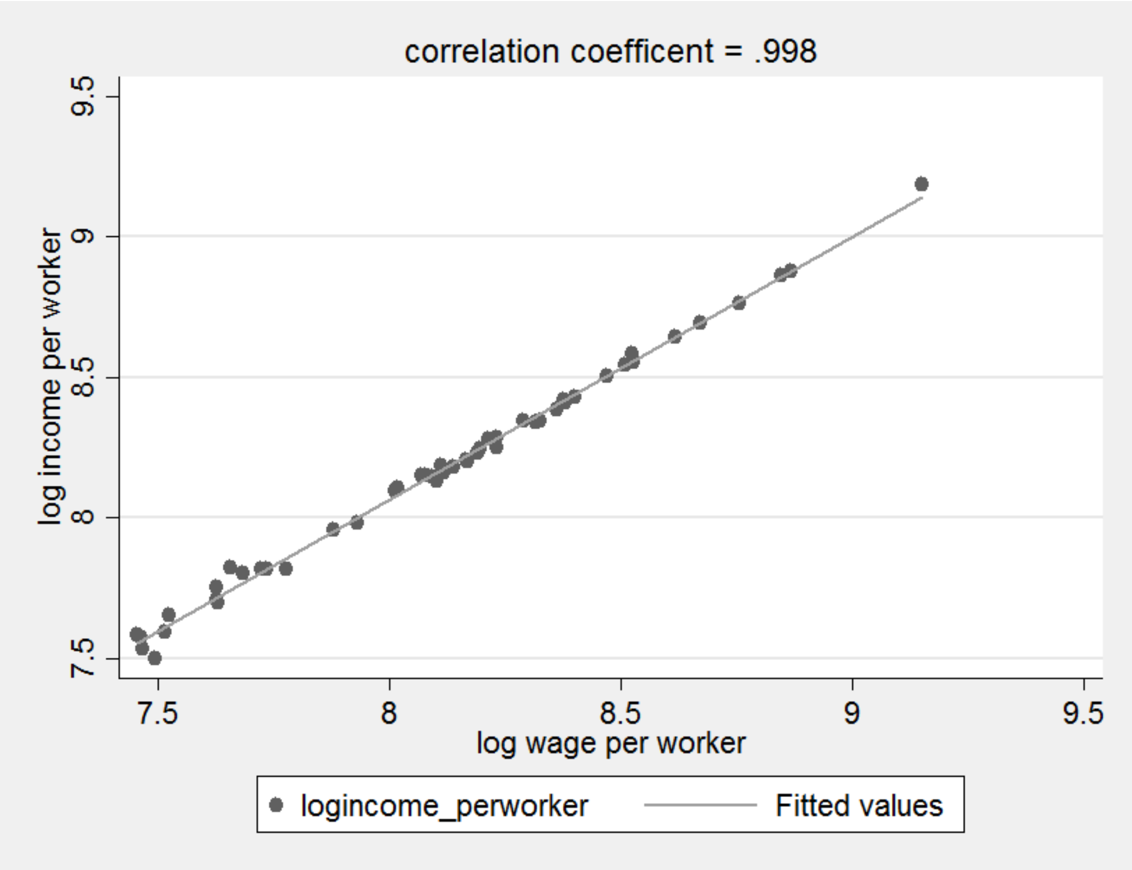
---

[105] In calculating distance between two points the explanation is built in the spwmatrix command. The distance computed from the latitudes and longitudes of the units under analysis is given by $d_{ir} = \sqrt{\sum_{t=1}^{q}(x_t[i] - x_t[j])^2}$, where $x_t[i]$ and $x_t[r]$ are the longitude and latitude points of region i and region r (Drukker, Peng, Prucha, & Raciborski, 2013). This formula gives the euclidean (straight line) distance and to get the spherical (great circle) distance, the haversine formula is used which is given by: $d_{ir} = r \times c,$ where r is the mean radius of the earth and is set at 6371 km (3959 miles) which is considered to be the earth's mean radius. c is the great circle distance in radians given by: $c = 2acrsin\{\min(1, \sqrt{a})\}$.

[106] The binary contiguity spatial weight matrix takes many forms among them rook, and queen weight matrices (see, (LeSage, 1999) for more details). While rook contiguity matrix defines regions that share a common side of their boundaries as neighbours, under queen contiguity matrix regions are considered to be neighbours if they share a common border on either side.

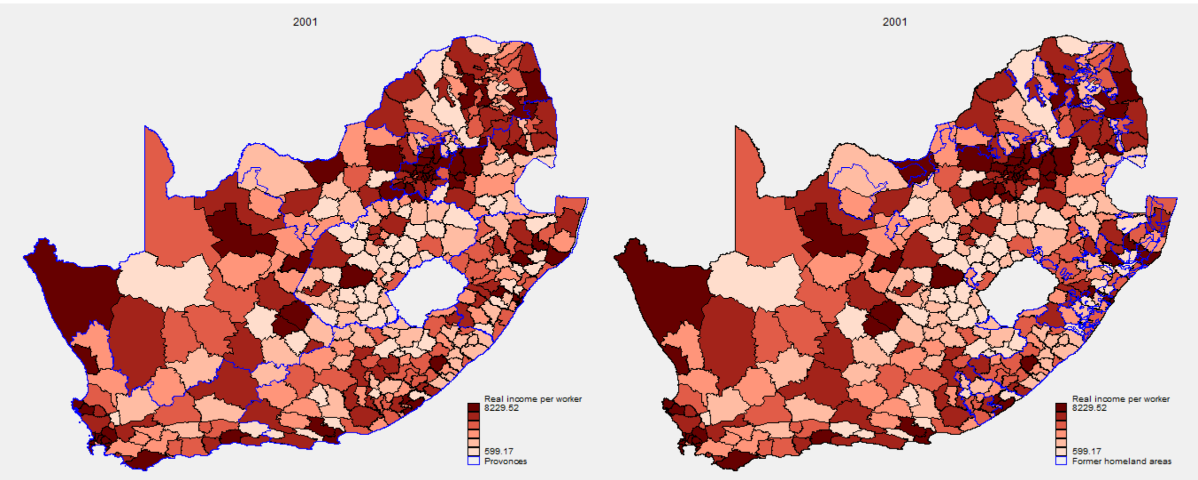
Appendix 3.3: Additional Figures and Tables.

Figure 3.3A: Association between income and wage per worker across districts.



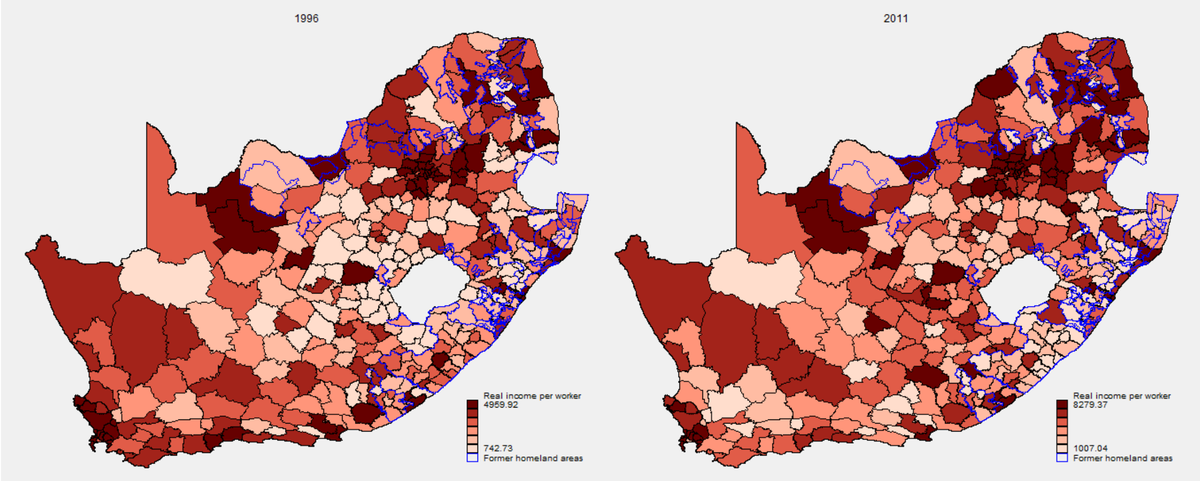
Notes: Calculation using income and wage data from NIDS Wave 2 for a sample of 53 district councils.

Figure 3.4A: Spatial distribution of income per worker across regions 2001



Notes: The maps are based on a sample of 354 magisterial districts in South Africa using 2001 census data. The blue lines trace out the boundaries of provinces in South Africa.

Figure 3.5A: Spatial distribution of income per worker across regions 1996 and 2011



Notes: The maps are based on a sample of 354 magisterial districts in South Africa using 2001 census data. The blue lines trace out the boundaries of former homeland areas in South Africa.

Table 3.3A: Association between regional income and wage per worker in South Africa.

	(1)	(2)
Log wage per worker	0.940***	0.936***
	(0.009)	(0.010)
% Human capital		-0.505
		(0.415)
Rural dummy		0.007
		(0.007)
% Manufacturing sector workers		-0.047
		(0.069)
% Agricultural sector workers		-0.084**
		(0.032)
% Mining sector workers		-0.069
		(0.072)
% Participation rate		-0.016
		(0.035)
% Unemployment rate		0.003
		(0.034)
Constant	0.545***	0.599***
	(0.072)	(0.087)
Observations	53	53
R-squared	0.995	0.997
F-test	11137	1836

Notes: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1. Using income per worker as the dependent variable, column (1) reports the estimates of the association between regional income and wage per worker, while column (2) adds regional specific factors. Human capital is the share of each region's population with at least a tertiary degree.

Table 3.4A: Monthly personal income brackets in the various censuses.

1996 census	2001/2011 census
No income	No income
R1 - R200	R1 - R400
R201 - R500	R401 - R800
R501 - R1000	R801 - R1600
R1001 - R1500	R1601 - R3200
R1501 - R2500	R3201 - R6400
R2501 - R3500	R6401 - R12 800
R3501 - R4500	R12 801 - R25 600
R4501 - R6000	R25 601 - R51 200
R6001 - R8000	R51 201 - R102 400
R8001 - R11000	R102 401 - R204 800
R11001 - R16000	R204 801+
R16001- R30000	
R30001+	

Notes: Brackets for income in the censuses.

Table 3.5A: Proportion of individuals with missing and zero income

	All Individuals			Employed Individuals		
Year	Missing	Zero	Missing & Zero	Missing	Zero	Missing & Zero
1996	10.1%	60.6%	70.7%	3.8%	1.2%	5%
2001	0%	68.2%	68.2%	0%	2.2%	2.2%
2011	7.9%	41.2%	49.1%	4.7%	8.5%	13.3%

Note: All Individuals includes all the people interviewed in the censuses.

Table 3.6A: Global Moran's I statistics including workers with zero income

Variables	1996	2001	2011
Moran's I statistic (I)	0.310	0.275	0.224
Moran's I statistic expected value - E(I)	-0.003	-0.003	-0.003
sd(I)	0.016	0.016	0.016
Z	20.14	17.87	14.571
p-value*	0.0000	0.0000	0.0000

Notes: The analysis in this table is a re-estimation of Table 3.5 results using income per worker which include workers with zero income.

Table 3.7A: Logistic regression model predicting missingness

VARIABLES	1996 Missing dummy	2001 Missing dummy	2011 Missing dummy
Age	-0.005***	-0.025***	-0.032***
	(0.001)	(0.001)	(0.000)
Education	-0.002	-0.053***	-0.085***
	(0.002)	(0.002)	(0.001)
Gender: Female	0.107***	0.251***	0.380***
	(0.011)	(0.015)	(0.006)
Race: Coloured	0.565***	-0.214***	0.127***
	(0.021)	(0.031)	(0.011)
Indian	0.435***	-0.293***	0.193***
	(0.030)	(0.046)	(0.015)
White	1.146***	0.174***	0.378***
	(0.015)	(0.023)	(0.009)
Location: Urban	-0.013	-0.211***	-0.157***
	(0.016)	(0.020)	(0.008)
Province: Eastern Cape	0.114***	0.327***	0.272***
	(0.024)	(0.034)	(0.013)
Northern Cape	-0.372***	-0.017	-0.261***
	(0.045)	(0.056)	(0.023)
Free State	-0.257***	0.014	-0.143***
	(0.031)	(0.040)	(0.016)
Kwazulu-Natal	0.217***	0.260***	0.203***
	(0.023)	(0.032)	(0.012)
North West	-0.021	-0.197***	-0.089***
	(0.029)	(0.040)	(0.015)
Gauteng	0.176***	0.026	0.076***
	(0.019)	(0.029)	(0.010)
Mpumalanga	0.331***	-0.190***	-0.118***
	(0.028)	(0.041)	(0.015)
Limpopo	0.327***	0.010	-0.075***
	(0.031)	(0.040)	(0.015)
Constant	-3.425***	-2.397***	-0.030*
	(0.033)	(0.045)	(0.018)
Observations	722,718	767,180	1,073,587

Note: Missing dummy is an indicator variable taking a value of 1 if a worker has missing income information and 0 otherwise. Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 3.8A: Regions confirming significant local spatial autocorrelation, 1996.

Region name	Region code	Quadrant	Spatial autocorrelation	Region name	Region code	Quadrant	Spatial autocorrelation
Bellville	101	H-H	Positive	Mt Fletcher	261	L-L	Positive
Goodwood	102	H-H	Positive	Colesberg	308	L-L	Positive
Cape	103	H-H	Positive	Hanover	310	L-L	Positive
Simonstown	104	H-H	Positive	Boshof	401	L-L	Positive
Wynberg	105	H-H	Positive	Jacobsdal	402	L-L	Positive
Kuilsrivier	107	H-H	Positive	Petrusburg	405	L-L	Positive
Paarl	108	H-H	Positive	Bultfontein	410	L-L	Positive
Stellenbosch	109	H-H	Positive	Hoopstad	413	L-L	Positive
Somerset West	110	H-H	Positive	Theunissen	417	L-L	Positive
Strand	111	H-H	Positive	Ventersburg	418	L-L	Positive
Wellington	112	H-H	Positive	Wesselsbron	421	L-L	Positive
Bredasdorp	113	H-H	Positive	Ficksburg	423	L-L	Positive
Hermanus	115	H-H	Positive	Fouriesburg	424	L-L	Positive
Worcester	130	H-H	Positive	Senekal	429	L-L	Positive
Hopefield	131	H-H	Positive	Brandfort	431	L-L	Positive
Malmesbury	132	H-H	Positive	Clocolan	432	L-L	Positive
Vredenburg	134	H-H	Positive	Dewetsdorp	433	L-L	Positive
Moorreesburg	135	H-H	Positive	Edenburg	434	L-L	Positive
Sasolburg	450	H-H	Positive	Excelsior	435	L-L	Positive
Durban	501	H-H	Positive	Marquard	438	L-L	Positive
Chatswoth	504	H-H	Positive	Philippolis	439	L-L	Positive
Potchefstroom	614	H-H	Positive	Reddersburg	440	L-L	Positive
Rustenburg	616	H-H	Positive	Trompsburg	441	L-L	Positive
Brits	617	H-H	Positive	Wepener	442	L-L	Positive
Pretoria	701	H-H	Positive	Winburg	443	L-L	Positive
Soshanguve	702	H-H	Positive	Botshabelo	444	L-L	Positive
Wonderboom	703	H-H	Positive	Smithfield	446	L-L	Positive

Johannesburg	704	H-H	Positive	Rouxville	448	L-L	Positive
Randburg	705	H-H	Positive	Zastron	449	L-L	Positive
Alberton	706	H-H	Positive	Underberg	513	L-L	Positive
Benoni	707	H-H	Positive	Polela	514	L-L	Positive
Boksburg	708	H-H	Positive	Kranskop	516	L-L	Positive
Germiston	709	H-H	Positive	Weenen	524	L-L	Positive
Kempton Park	710	H-H	Positive	Ngotshe	531	L-L	Positive
Brakpan	711	H-H	Positive	Paulpietersburg	532	L-L	Positive
Heidelberg	712	H-H	Positive	Schweizer0Reneke	609	L-L	Positive
Nigel	713	H-H	Positive	Wolmaransstad	610	L-L	Positive
Springs	714	H-H	Positive	Wakkerstroom	808	L-L	Positive
Krugersdorp	715	H-H	Positive	Ceres	126	L-H	Negative
Randfontein	717	H-H	Positive	Tulbagh	129	L-H	Negative
Roodepoort	718	H-H	Positive	Koppies	414	L-H	Negative
Westonaria	719	H-H	Positive	Vredefort	419	L-H	Negative
Bronkhorstspruit	720	H-H	Positive	Frankfort	425	L-H	Negative
Vereeniging	722	H-H	Positive	Reitz	428	L-H	Negative
Vanderbijlpark	723	H-H	Positive	Vrede	430	L-H	Negative
Standerton	806	H-H	Positive	Ventersdorp	613	L-H	Negative
Kriel	809	H-H	Positive	Temba	619	L-H	Negative
Hoëveldrif	811	H-H	Positive	Mkobola	828	L-H	Negative
Delmas	812	H-H	Positive	Waterberg	907	L-H	Negative
Middelburg	815	H-H	Positive	Venterstad	205	H-L	Negative
Witbank	817	H-H	Positive	Kimberley	321	H-L	Negative
Lady Grey	203	L-L	Positive	Welkom	408	H-L	Negative
Steynsburg	204	L-L	Positive	Bloemfontein	445	H-L	Negative
Barkley0East	207	L-L	Positive	Pietermaritzburg	507	H-L	Negative
Elliot	208	L-L	Positive	Lower Umfolozi	536	H-L	Negative
Maclear	210	L-L	Positive	Nelspruit	821	H-L	Negative
Maluti	259	L-L	Positive				

Table 3.9A: Regions confirming significant local spatial autocorrelation, 2001.

Region name	Region code	Quadrant	Spatial autocorrelation	Region name	Region code	Quadrant	Spatial autocorrelation
Bellville	101	H-H	Positive	Sterkspruit	276	L-L	Positive
Goodwood	102	H-H	Positive	Petrusburg	405	L-L	Positive
Cape	103	H-H	Positive	Bultfontein	410	L-L	Positive
Simonstown	104	H-H	Positive	Hoopstad	413	L-L	Positive
Wynberg	105	H-H	Positive	Theunissen	417	L-L	Positive
Kuilsrivier	107	H-H	Positive	Ventersburg	418	L-L	Positive
Paarl	108	H-H	Positive	Wesselsbron	421	L-L	Positive
Stellenbosch	109	H-H	Positive	Ficksburg	423	L-L	Positive
Somerset West	110	H-H	Positive	Fouriesburg	424	L-L	Positive
Strand	111	H-H	Positive	Senekal	429	L-L	Positive
Bredasdorp	113	H-H	Positive	Brandfort	431	L-L	Positive
Hopefield	131	H-H	Positive	Clocolan	432	L-L	Positive
Malmesbury	132	H-H	Positive	Dewetsdorp	433	L-L	Positive
Vredenburg	134	H-H	Positive	Edenburg	434	L-L	Positive
Moorreesburg	135	H-H	Positive	Excelsior	435	L-L	Positive
Sasolburg	450	H-H	Positive	Jagersfontein	436	L-L	Positive
Potchefstroom	614	H-H	Positive	Marquard	438	L-L	Positive
Rustenburg	616	H-H	Positive	Philippolis	439	L-L	Positive
Brits	617	H-H	Positive	Reddersburg	440	L-L	Positive
Pretoria	701	H-H	Positive	Trompsburg	441	L-L	Positive
Wonderboom	703	H-H	Positive	Wepener	442	L-L	Positive
Johannesburg	704	H-H	Positive	Winburg	443	L-L	Positive
Randburg	705	H-H	Positive	Botshabelo	444	L-L	Positive
Alberton	706	H-H	Positive	Smithfield	446	L-L	Positive
Benoni	707	H-H	Positive	Bethulie	447	L-L	Positive

Boksburg	708	H-H	Positive	Rouxville	448	L-L	Positive
Germiston	709	H-H	Positive	Zastron	449	L-L	Positive
Kempton Park	710	H-H	Positive	Thaba Nchu	451	L-L	Positive
Brakpan	711	H-H	Positive	Schweizer-Reneke	609	L-L	Positive
Heidelberg (GT)	712	H-H	Positive	Wakkerstroom	808	L-L	Positive
Nigel	713	H-H	Positive	Ceres	126	L-H	Negative
Springs	714	H-H	Positive	Tulbagh	129	L-H	Negative
Krugersdorp	715	H-H	Positive	Clanwilliam	136	L-H	Negative
Oberholzer	716	H-H	Positive	Heilbron	411	L-H	Negative
Randfontein	717	H-H	Positive	Koppies	414	L-H	Negative
Roodepoort	718	H-H	Positive	Frankfort	425	L-H	Negative
Westonaria	719	H-H	Positive	Vrede	430	L-H	Negative
Bronkhorstspruit	720	H-H	Positive	Ventersdorp	613	L-H	Negative
Cullinan	721	H-H	Positive	Temba	619	L-H	Negative
Vereeniging	722	H-H	Positive	Balfour	810	L-H	Negative
Vanderbijlpark	723	H-H	Positive	Groblersdal	814	L-H	Negative
Kriel	809	H-H	Positive	Mkobola	828	L-H	Negative
Highveld Ridge	811	H-H	Positive	KwaMhlanga	830	L-H	Negative
Middelburg (MP)	815	H-H	Positive	Waterberg	907	L-H	Negative
Witbank	817	H-H	Positive	Molteno	214	H-L	Negative
Lady Grey	203	L-L	Positive	Port Elizabeth	240	H-L	Negative
Venterstad	205	L-L	Positive	Kimberley	321	H-L	Negative
Barkly-East	207	L-L	Positive	Welkom	408	H-L	Negative
Elliot	208	L-L	Positive	Bloemfontein	445	H-L	Negative
Indwe	209	L-L	Positive	Nelspruit	821	H-L	Negative
Maclear	210	L-L	Positive	Pietersburg	904	H-L	Negative
Wodehouse	211	L-L	Positive				

Table 3.10A: Regions confirming significant local spatial autocorrelation, 2011.

Region name	Region code	Quadrant	Spatial autocorrelation	Region name	Region code	Quadrant	Spatial autocorrelation
Bellville	101	H-H	Positive	Mount Frere	262	L-L	Positive
Goodwood	102	H-H	Positive	Mqanduli	263	L-L	Positive
Cape	103	H-H	Positive	Ngqueleni	264	L-L	Positive
Simonstown	104	H-H	Positive	Port St Johns	266	L-L	Positive
Wynberg	105	H-H	Positive	Qumbu	267	L-L	Positive
Kuilsrivier	107	H-H	Positive	Tabankulu	269	L-L	Positive
Paarl	108	H-H	Positive	Tsolo	270	L-L	Positive
Stellenbosch	109	H-H	Positive	Willowvale	273	L-L	Positive
Somerset West	110	H-H	Positive	Lady Frere	275	L-L	Positive
Strand	111	H-H	Positive	Umzimkulu	277	L-L	Positive
Hopefield	131	H-H	Positive	Ntabethemba	278	L-L	Positive
Malmesbury	132	H-H	Positive	Clocolan	432	L-L	Positive
Vredenburg	134	H-H	Positive	Marquard	438	L-L	Positive
Sasolburg	450	H-H	Positive	Richmond	506	L-L	Positive
Potchefstroom	614	H-H	Positive	Ixopo	509	L-L	Positive
Rustenburg	616	H-H	Positive	Alfred	510	L-L	Positive
Brits	617	H-H	Positive	Underberg	513	L-L	Positive
Pretoria	701	H-H	Positive	Polela	514	L-L	Positive
Wonderboom	703	H-H	Positive	Impendle	515	L-L	Positive
Johannesburg	704	H-H	Positive	Kranskop	516	L-L	Positive
Randburg	705	H-H	Positive	New Hanover	518	L-L	Positive
Alberton	706	H-H	Positive	Bergville	521	L-L	Positive
Benoni	707	H-H	Positive	Weenen	524	L-L	Positive
Boksburg	708	H-H	Positive	Babanango	530	L-L	Positive
Germiston	709	H-H	Positive	Ngotshe	531	L-L	Positive
Kempton Park	710	H-H	Positive	Paulpietersburg	532	L-L	Positive
Brakpan	711	H-H	Positive	Eshowe	534	L-L	Positive

Heidelberg (GT)	712	H-H	Positive
Nigel	713	H-H	Positive
Springs	714	H-H	Positive
Krugersdorp	715	H-H	Positive
Oberholzer	716	H-H	Positive
Randfontein	717	H-H	Positive
Roodepoort	718	H-H	Positive
Westonaria	719	H-H	Positive
Bronkhorstspruit	720	H-H	Positive
Cullinan	721	H-H	Positive
Vereeniging	722	H-H	Positive
Vanderbijlpark	723	H-H	Positive
Bethal	802	H-H	Positive
Standerton	806	H-H	Positive
Kriel	809	H-H	Positive
Highveld Ridge	811	H-H	Positive
Middelburg (MP)	815	H-H	Positive
Witbank	817	H-H	Positive
Thabazimbi	909	H-H	Positive
Pearston	233	L-L	Positive
Jansenville	234	L-L	Positive
Joubertina	239	L-L	Positive
Hewu	244	L-L	Positive
Bizana	250	L-L	Positive
Elliotdale	252	L-L	Positive
Flagstaff	254	L-L	Positive
Libode	257	L-L	Positive
Lusikisiki	258	L-L	Positive
Maluti	259	L-L	Positive
Mount Ayliff	260	L-L	Positive
Mthonjaneni	537	L-L	Positive
Mapumulo	544	L-L	Positive
Nkandla	545	L-L	Positive
Nqutu	546	L-L	Positive
Msinga	547	L-L	Positive
Nongoma	549	L-L	Positive
Mitchell's Plain	106	L-H	Negative
Wellington	112	L-H	Negative
Caledon	114	L-H	Negative
Ceres	126	L-H	Negative
Tulbagh	129	L-H	Negative
Clanwilliam	136	L-H	Negative
Koppies	414	L-H	Negative
Vredefort	419	L-H	Negative
Frankfort	425	L-H	Negative
Vrede	430	L-H	Negative
Ventersdorp	613	L-H	Negative
Temba	619	L-H	Negative
Soweto	724	L-H	Negative
Mkobola	828	L-H	Negative
Mbibana	829	L-H	Negative
KwaMhlanga	830	L-H	Negative
Queenstown	215	H-L	Negative
King William's Town	220	H-L	Negative
Cradock	229	H-L	Negative
Port Elizabeth	240	H-L	Negative
Umtata	272	H-L	Negative
Bloemfontein	445	H-L	Negative
Pinetown	503	H-L	Negative
Pietermaritzburg	507	H-L	Negative

Mount Fletcher	261	L-L	Positive	Vryheid	533	H-L	Negative
				Lower Umfolozi	536	H-L	Negative

Appendix for Chapter 4

Table 4.1A: Summary Statistics of key variables (1996-2011).

Variable	Obs	Mean	Std. Dev.	Min	Max
Growth of income per worker	708	0.014	0.042	-0.078	0.108
Income per worker	1062	7.481	0.387	6.340	9.023
Market potential	1062	20.66	1.360	18.02	25.61
Distance	125316	601.4	362.5	2.5	1795.5
Total population	1062	11.07	1.255	8.130	14.17
Average temperature	1062	2.861	0.131	2.244	3.160
Average rainfall	1062	4.006	0.431	1.960	4.825
Skilled workers	1062	0.053	0.039	0.004	0.312
Unemployment rate	1062	0.374	0.168	0.0279	0.841
Homeland status	1062	0.253	0.396	0	1
Share of mining workers	708	0.039	0.098	0	0.838
Share of manufacturing workers	708	0.076	0.063	0.001	0.352
Share of agricultural workers	708	0.220	0.178	0.003	0.795

Note: The summary statistics are for data for the entire sample period (1996-2011) pooled together so that for each variable we have the average value for 1996, 2001 and 2011 data. Of these variables growth of income per worker and income per worker are in real terms. Income per worker, market potential, total personal income, total population, total housing stocks, average temperature and rainfall are in logs. Distance is in levels for a 354 x 354 matrix. Skilled workers, share of mining workers, share of agricultural workers, share of manufacturing workers, and homeland status variables are shares, unemployment rate is a proportion.

Table 4.2A: Empirical studies on regional wage convergence for developed countries.

Authors	Data	Method	Main findings
Distributional dynamic studies			
Maza & Villaverde (2006)	Spanish provinces, 1995-2001	Kernel density	Convergence
Webber (2001)	57 regions in EU countries, 1980-1994	Markov chain	Convergence
***σ*-convergence studies**			
Carlino & Mills (1996)	8 regions of United States, 1929 – 1990	Coefficient of variation	Convergence: σ decreased from 0.33 to 0.12.
Bukenya, Davis, Banerjee, & Gyawali (2011)	67 Alabama counties in the United States, 1969 – 2008	Coefficient of variation	Convergence: σ decreased from 0.22 to 0.16 However, the trends of σ changes across different sub periods and show evidence of convergence and divergence.
Ferens (2015)	16 regions of Poland, 2005 – 2013	Coefficient of variation	Convergence: σ decreased from 0.12 to 0.11
Maza & Villaverde (2006)	Spanish provinces, 1995-2001	Coefficient of variation	Convergence: σ decreased from 0.095 to 0.083
Niebuhr, Granato, Haas, & Hamann (2012)	439 German counties, 1996-2005	Coefficient of variation	Neither convergence nor divergence: σ is stable over time.
Enflo, Lundh, & Prado. (2014)	24 Swedish regions, 1860 – 1940	Coefficient of variation	Convergence: σ decreased from 0.22 to 0.14
Rosés & Sánchez-Alonso (2004)	48 Spanish provinces, 1850 – 1914	Coefficient of variation	Mixed evidence of divergence and convergence according to worker type for entire period: σ increased from 0.25 to 0.32 for agrarian and from 0.15 to 0.18, while it decreased from 0.21 to 0.16 for skilled industrial workers. However, the trends of σ changes for all workers across different sub periods.
***β*-convergence studies**			
Carlino & Mills (1996)	8 regions of United States, 1929 – 1990	Cross section Time series	Unconditional convergence based on cross section and less conclusive evidence of convergence based on time series.

Bukenya, Davis, Banerjee, & Gyawali (2011)	67 Alabama counties in the United States, 1969 – 2008	Time series	No evidence in support of the convergence hypothesis
Moazzami (1997)	10 provinces and 2 regions of Canada, 1960-1994	Time series	Mixed trends, with evidence for strong convergence in six of the regions and evidence of weak convergence in 5 of the regions.
Ferens (2015)	16 regions of Poland, 2005 – 2013	Cross section	Convergence: unconditional convergence rate of 1.3% per year.
Maza & Villaverde (2006)	Spanish provinces, 1995-2001	Cross section Spatial cross section	Convergence: unconditional convergence rate of 4.1% per year and conditional convergence rate of 5.3%
Rosés & Sánchez-Alonso (2004)	48 Spanish provinces, 1850 – 1914	Cross section	Convergence: unconditional convergence rate of 4.1% per year and conditional convergence rate of 5.3%
Tavernier & Temel (1997)	3130 counties of United States, 1978 – 1992	Cross section	Convergence: unconditional convergence rate of between 6% and 7.1% per year and conditional convergence rate of between 7.1% and 7.8% per year
Enflo, Lundh, & Prado. (2014)	24 Swedish regions, 1860 – 1940	Panel	Convergence: unconditional convergence rate of 13.5% per year and conditional convergence rate of 27%
Naz, Ahmad, & Naveed (2017)	203 regions in EU countries, 1996 – 2006	Cross section Panel	Convergence for regions within the same country but no evidence of convergence for border regions.
Šlander & Ogorevc (2010)	115-210 regions in EU countries, 1996 – 2006	Cross section Spatial cross section	Convergence: unconditional convergence rate of between 1.8% and 3% per year and conditional convergence rate of between 1.6% and 2.1% per year

Table 4.3A: Empirical studies on regional wage convergence for developing countries.

Authors	**Data**	Method	Main findings
***σ*-convergence studies**			
Huang & Chand (2015)	31 provinces of China, 1992 - 2010	Coefficient of variation	Divergence over the entire period: σ increased from 0.18 to 0.27. Changing trends in different sub-periods, with divergence between 1992 to 2001 as σ increased from 0.18 to 0.31 and convergence between 2001 and 2010 as σ decreased from 0.31 to 0.27.
Candelaria, Daly, & Hale (2015)	28 provinces of China, 1993 – 2011	Coefficient of variation	Divergence for entire study period: σ increased from 0.14 to 0.23 Different trends across the sub-periods: σ increased from 0.14 to 0.26 between 1990 and 1999 and decreased from 0.26 to 0.23 between 2000 and 2011.
Estanislau, Staduto, & Parré (2013)	27 Brazilian states, 1992 – 2009	Coefficient of variation	Convergence: σ decreased from 0.37 to 0.28 for permanent workers and from 0.46 to 0.23 for temporary workers.
Zaman & Goschin (2014)	41 counties of Romania, 1991 – 2010	Coefficient of variation	Divergence: σ increased from 0.07 to 0.15 for the whole period, but its trend changed for different sub-periods.
Goschin (2015)	Romanian counties, 1991 – 2013	Coefficient of variation	Divergence: σ increased from 0.06 to 0.16
Vakulenko (2016)	77 Russian regions, 1995- 2010	Gini coefficient	Convergence: σ decreased from 0.14 to 0.13
Collins (1999)		Coefficient of variation	Mixed evidence of divergence and convergence according to worker type for the entire study period: σ increased from 0.25 to 0.32 for agrarian and from 0.15 to 0.18, while it decreased from 0.21 to 0.16 for skilled industrial workers. However, the trends of σ changes for all workers in different sub periods and show evidence of divergence and convergence.
***β*-convergence studies**			

Chen, Chang, & Su (2016)	31 provinces of China, 1978 – 2010	Time series	Mixed trends, with evidence for strong convergence in eastern and western regions and evidence of weak convergence in northeastern and central regions.
Huang & Chand (2015)	31 provinces of China, 1992 – 2010	Cross section	Divergence for the entire period: unconditional divergence at a rate of 2.8% per year. Changing patterns in different sub-periods: unconditional divergence at a rate of 5.3% per year and unconditional convergence at a rate of 2.7% per year.
Estanislau, Staduto, & Parré (2013)	27 Brazilian states, 1992 – 2009	Spatial cross section	Convergence: unconditional convergence rate of 3.2% per year and conditional convergence rate of 4.1% for permanent workers. Unconditional convergence rate of 3.4% per year and conditional convergence rate of 3.5% for temporary workers.
Zaman & Goschin (2014)	41 counties of Romania, 1991 – 2010	Cross section	No evidence in support of unconditional convergence of the entire period, but unconditional convergence at a rate of between 2% per year and 4.2% per year for different sub-periods. Conditional convergence at a rate of between 4% per year and 5.4% per year for different sub-periods.
Vakulenko (2016)	77 Russian regions, 1995- 2010	Panel Spatial panel	Conditional convergence

Table 4.4A: Conditional β-convergence test, stepwise approach, 1996-2011.

VARIABLES	(1)	(2)	(3)	(4)	(6)	(7)	(8)	(9)	(10)
Log initial income per worker	-0.012***	-0.010***	-0.009***	-0.010***	-0.012***	-0.023***	-0.015***	-0.009***	-0.006**
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)
Log market potential	0.001								
	(0.001)								
Initial share of mining workers		0.006*							
		(0.004)							
Log initial temperature			-0.016***						
			(0.006)						
Log initial rainfall				-0.001					
				(0.001)					
Homeland status					-0.015***				
					(0.002)				
Initial human capital						0.563***			
						(0.107)			
Initial unemployment rate							-0.029***		
							(0.004)		
Initial share of manu workers								-0.011	
								(0.013)	
Initial share of agric workers									0.013**
									(0.005)
Constant	0.113***	0.110***	0.148***	0.112***	0.130***	0.208***	0.157***	0.101***	0.077***
	(0.020)	(0.019)	(0.024)	(0.020)	(0.017)	(0.027)	(0.019)	(0.020)	(0.022)
Observations	354	354	354	354	354	354	354	354	354
R-squared	0.054	0.055	0.073	0.053	0.215	0.141	0.192	0.054	0.073
F-test	9.382	10.39	12.94	9.334	52.41	23.03	33.15	9.446	12.12
Convergence rate (%)	1.27	1.09	0.997	1.07	1.33	2.78	1.65	0.96	0.67
Half-life (years)	55	64	70	65	52	25	42	72	103

Asterisks indicate the level of significance, where: *** p<0.01, ** p<0.05, * p<0.1 and the values in parentheses are heteroscedasticity consistent standard errors. The dependent variable is average annual growth in income per worker.

Table 4.5A: Conditional β-convergence test, stepwise approach, 1996-2001.

Variables	(1)	(2)	(3)	(4)	(6)	(7)	(8)	(9)	(10)
Initial income per worker	-0.004	0.011*	0.011*	0.011*	0.011*	-0.004	0.015**	0.009	-0.000
	(0.009)	(0.006)	(0.006)	(0.006)	(0.006)	(0.009)	(0.007)	(0.007)	(0.007)
Log market potential	0.005***								
	(0.002)								
Mineral resource endowments		0.014							
		(0.011)							
Log temperature			0.007						
			(0.015)						
Log rainfall				0.006					
				(0.004)					
Share of area in homelands					-0.001				
					(0.005)				
Skilled workers						0.684***			
						(0.213)			
Unemployment rate							0.020		
							(0.012)		
Share of manufacturing workers								0.022	
								(0.036)	
Share of agricultural workers									-0.042***
									(0.014)
Constant	0.104*	0.067	0.047	0.041	0.065	0.185**	0.031	0.078	0.168***
	(0.055)	(0.053)	(0.064)	(0.054)	(0.054)	(0.074)	(0.057)	(0.056)	(0.062)
Observations	354	354	354	354	354	354	354	354	354
R-squared	0.028	0.011	0.010	0.014	0.009	0.027	0.018	0.010	0.039
F-test	5.353	2.186	1.692	2.720	1.617	11.29	2.834	1.707	6.062

Asterisks indicate the level of significance, where: *** $p<0.01$, ** $p<0.05$, * $p<0.1$ and the values in parentheses are Heteroscedasticity and autocorrelation consistent (HAC) standard errors. The dependent variable is average annual growth in income per worker.

VARIABLES	(1)	(2)	(3)	(4)	(6)	(7)	(8)	(9)	(10)
Initial income per worker	-0.038***	-0.031***	-0.030***	-0.031***	-0.033***	-0.055***	-0.035***	-0.032***	-0.026***
	(0.005)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)
Initial market potential	0.003**								
	(0.001)								
Initial mineral resources		0.018**							
		(0.008)							
Initial temperature			-0.027***						
			(0.009)						
Initial rainfall				0.000					
				(0.002)					
Share of homeland area					-0.023***				
					(0.002)				
Initial human_capital1						1.014***			
						(0.149)			
Initial unemployment rate							-0.051***		
							(0.008)		
Initial share of manuf workers								0.021	
								(0.017)	
Initial share of agric workers									0.020**
									(0.008)
Constant	0.255***	0.241***	0.308***	0.237***	0.266***	0.445***	0.302***	0.247***	0.192***
	(0.029)	(0.028)	(0.037)	(0.029)	(0.025)	(0.038)	(0.028)	(0.029)	(0.032)
Observations	354	354	354	354	354	354	354	354	354
R-squared	0.259	0.249	0.267	0.245	0.379	0.387	0.370	0.248	0.261
F-test	53.09	51.24	57.64	51.89	92.80	91.20	76.78	51.03	58.60
Convergence rate	4.76	3.72	3.57	3.66	4.04	8.02	4.36	3.84	3.02
Half-life (years)	15	19	19	19	17	7	16	18	23

Table 4.6A: Conditional β-convergence test, stepwise approach, 2001-2011.

Table 4.7A: Conditional β-convergence test for a restricted sample.

	Unskilled workers			Skilled workers		
VARIABLES (Initial values)	1996-2011	1996-2001	2001-2011	1996-2011	1996-2001	2001-2011
Log initial income per worker	-0.051***	-0.077***	-0.089***	-0.038***	-0.108***	-0.089***
	(0.005)	(0.013)	(0.005)	(0.004)	(0.018)	(0.003)
Log market potential	0.001	0.002	0.003	0.003***	0.019***	0.007***
	(0.001)	(0.004)	(0.002)	(0.001)	(0.005)	(0.002)
Log population density	0.001	0.006*	0.003*	0.000	-0.005	-0.000
	(0.001)	(0.003)	(0.002)	(0.001)	(0.004)	(0.001)
Unemployment rate (%)	-0.023**	0.003	-0.033*	-0.017**	-0.066*	-0.041***
	(0.011)	(0.029)	(0.018)	(0.008)	(0.040)	(0.013)
Share agricultural workers	-0.041***	-0.100***	-0.064***	-0.015*	-0.073*	-0.015
	(0.012)	(0.032)	(0.018)	(0.008)	(0.041)	(0.013)
Share manufacturing workers	-0.022	0.059	-0.017	-0.015	0.031	-0.012
	(0.025)	(0.065)	(0.037)	(0.017)	(0.086)	(0.027)
Log average rainfall	0.005	0.003	0.004	0.003	0.016	0.009**
	(0.003)	(0.008)	(0.006)	(0.002)	(0.010)	(0.004)
Log average temperature	0.007	0.024	0.009	-0.003	0.020	-0.013
	(0.010)	(0.025)	(0.014)	(0.007)	(0.033)	(0.010)
Share mining workers	0.005	0.038	0.044**	-0.014*	-0.029	-0.013
	(0.012)	(0.030)	(0.022)	(0.008)	(0.039)	(0.016)
Share of area in homelands	-0.000	0.024*	-0.000	-0.010***	-0.067***	-0.022***
	(0.005)	(0.013)	(0.007)	(0.004)	(0.018)	(0.005)
Constant	0.300***	0.384***	0.523***	0.320***	0.579***	0.712***
	(0.044)	(0.114)	(0.067)	(0.038)	(0.189)	(0.048)
Observations	354	354	354	354	354	354
R-squared	0.363	0.177	0.553	0.280	0.153	0.701
F-test	19.55	7.394	42.46	13.37	6.208	80.52
Convergence rate (%)	10	10	22	6	16	23
Half-life (years)	7	7	3	12	4	3

Notes: Asterisks indicate the level of significance, where: *** p<0.01, ** p<0.05, * p<0.1

Table 4.8A: Convergence test for income per worker including workers with zero income.

	Unconditional			Conditional		
VARIABLES	1996-2011	1996-2001	2001-2011	1996-2011	1996-2001	2001-2011
Log real income per worker	-0.009***	0.009	-0.027***	-0.039***	-0.058***	-0.065***
	(0.002)	(0.006)	(0.003)	(0.004)	(0.012)	(0.005)
Skilled workers %)				0.325***	0.785***	0.322***
				(0.063)	(0.185)	(0.051)
Log market potential				0.003***	0.006**	0.005***
				(0.001)	(0.002)	(0.001)
Log population density				-0.002***	-0.002	-0.004***
				(0.001)	(0.002)	(0.001)
Unemployment rate (%)				-0.010	0.077***	-0.024**
				(0.008)	(0.022)	(0.011)
Share of agricultural workers (%)				-0.044***	-0.088***	-0.043***
				(0.007)	(0.021)	(0.012)
Share of manufacturing workers (%)				-0.028**	0.008	0.001
				(0.014)	(0.040)	(0.020)
Log average rainfall				0.008***	-0.002	0.015***
				(0.002)	(0.005)	(0.003)
Log average temperature				0.005	0.019	0.002
				(0.005)	(0.015)	(0.007)
Share of mining workers (%)				-0.008	0.010	0.011
				(0.007)	(0.020)	(0.013)
Share of area in homelands				-0.016***	-0.035***	-0.017***
				(0.003)	(0.008)	(0.004)
Constant	0.092***	-0.037	0.220***	0.235***	0.327***	0.351***
	(0.016)	(0.041)	(0.021)	(0.031)	(0.091)	(0.043)
Observations	354	354	354	354	354	354
R-squared	0.048	0.008	0.204	0.389	0.166	0.536
F-test	17.73	2.70	90.36	19.78	6.18	35.87
Convergence rate (%)	0.99	Diverge	3.10	5.93	6.87	10.47
Half-life (years)	70	Diverge	22	12	10	7

Standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1

Appendix for Chapter 5

Table 5.1A: Summary Statistics of key variables (1996-2011).

Variable	Obs	Mean	Std. Dev.	Min	Max
Income per worker	1062	7.481	0.387	6.340	9.023
Market potential	1062	20.66	1.360	18.02	25.61
Total personal income	1062	17.57	1.628	14.10	23.09
Total housing stocks	1062	11.006	1.267	8.084	14.59
Distance	125316	601.4	362.5	2.5	1795.5
Total population	1062	11.07	1.255	8.130	14.17
Average temperature	1062	2.861	0.131	2.244	3.160
Average rainfall	1062	4.006	0.431	1.960	4.825
Share skilled workers	1062	0.053	0.039	0.004	0.312
Unemployment rate	1062	0.374	0.168	0.0279	0.841
Homeland status	1062	0.253	0.396	0	1
Share of mining workers	708	0.039	0.098	0	0.838

Note: The summary statistics are for data for the entire sample period (1996-2011) pooled together so that for each variable we have the average value for 1996, 2001 and 2011 data. Of these variables income per worker, market potential, total personal income, total population, total housing stocks, average temperature and rainfall are in logs. Distance is in levels for a 354 x 354 matrix. The share of skilled workers, share of mining workers, homeland status variables are shares and unemployment rate is a proportion.

Table 5.2A: Correlation coefficients.

	Income per worker	Total personal income	Total population
Income per worker	1.0000		
Total personal income	0.7387	1.0000	
Total population	0.4025	0.8236	1.0000

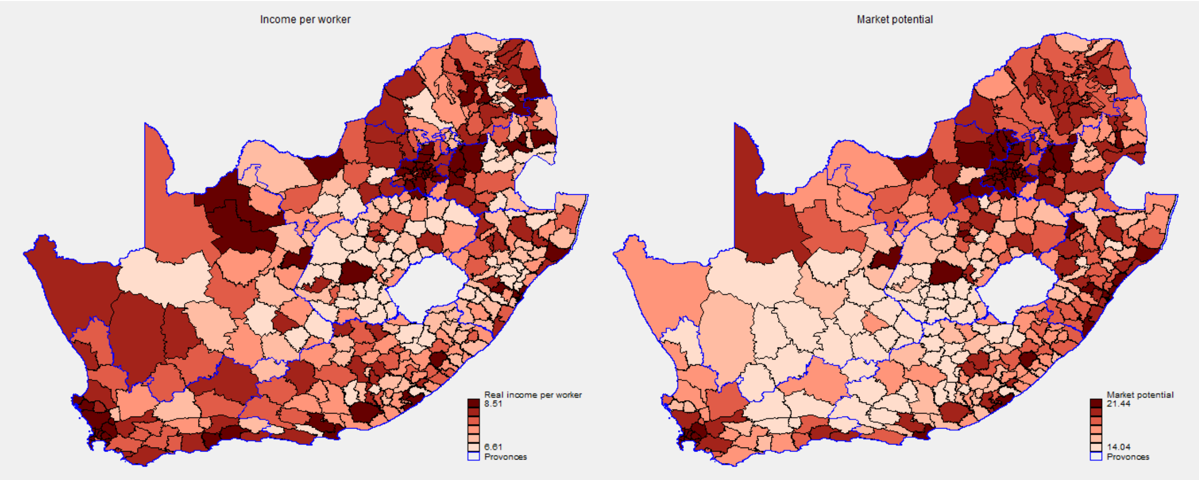
Note: Variables are in levels and are for the entire sample period (1996–2011). Thus, the correlation coefficient is an average value for 1996, 2001 and 2011 data.

Table 5.3A: The Helpman-Hanson Model including workers with zero income

	Without Controls			With Controls		
Year	1996	2001	2011	1996	2001	2011
Log market potential	0.111***	0.101***	0.170***	0.221***	0.238***	0.354***
	(0.027)	(0.031)	(0.032)	(0.055)	(0.064)	(0.085)
Log income per worker	9.116***	10.250***	5.754***	4.254***	4.021***	2.512***
	(2.357)	(3.311)	(1.176)	(1.208)	(1.229)	(0.730)
Log distance	-3.142***	-3.977***	-2.192***	-1.332***	-1.388***	-0.763***
	(0.617)	(1.093)	(0.267)	(0.144)	(0.172)	(0.091)
Implied Values						
σ.	8.987***	9.938***	5.890***	4.522***	4.198***	2.826***
	(2.206)	(3.103)	(1.106)	(1.133)	(1.124)	(0.682)
μ.	0.876***	0.872***	0.850***	0.828***	0.795***	0.727***
	(0.016)	(0.018)	(0.019)	(0.033)	(0.037)	(0.062)
τ.	0.393***	0.445***	0.448***	0.378***	0.434***	0.418***
	(0.036)	(0.039)	(0.054)	(0.088)	(0.107)	(0.148)
$\sigma/(\sigma-1)$.	1.125	1.112	1.204	1.284	1.313	1.548
$\sigma(1-\mu)$.	1.113	1.272	0.884	0.778	0.859	0.771
Control variables						
Skilled workers				3.736***	2.419***	2.814***
				(0.247)	(0.193)	(0.141)
Mineral resource endowments				0.381***	0.897***	0.851***
				(0.084)	(0.137)	(0.116)
Log temperature				0.578***	0.596***	0.280***
				(0.088)	(0.134)	(0.101)
Log rainfall				-0.063**	-0.092**	-0.043
				(0.026)	(0.044)	(0.034)
Unemployment rate				-0.354***	-0.322**	-0.954***
				(0.092)	(0.129)	(0.133)
Homeland status				-0.170***	-0.350***	-0.361***
				(0.044)	(0.053)	(0.037)
Adjusted R-squared	0.492	0.477	0.413	0.717	0.674	0.758
F-statistic	115.164	108.175	83.747	100.569	82.170	123.560
Obs	354	354	354	354	354	354

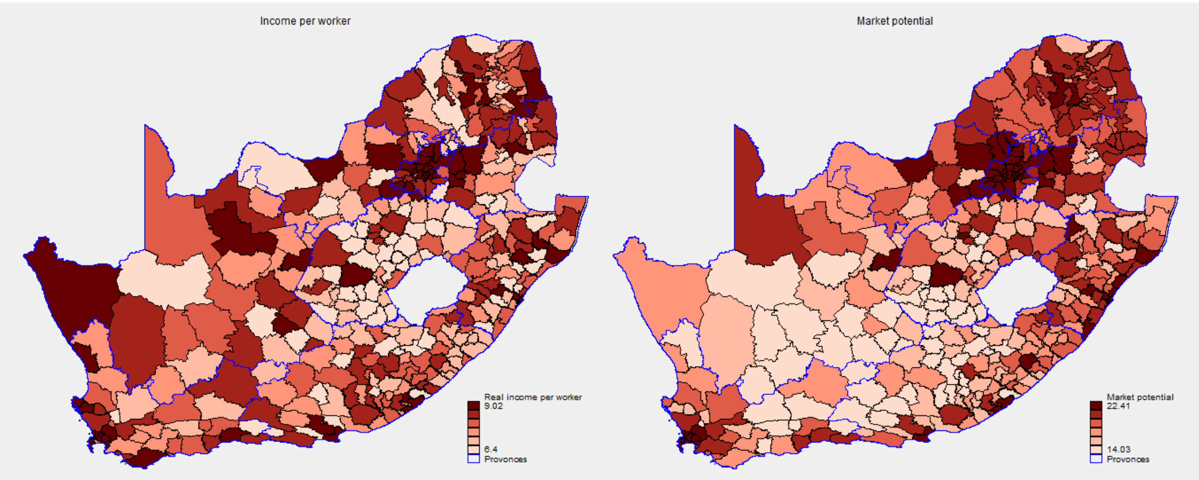
Asterisks indicate the level of significance, where: *** p<0.01, ** p<0.05, * p<0.1 and the values in parentheses are heteroscedasticity- consistent errors. Estimated models include a constant.

Figure 5.1A: Spatial distribution of income per worker & market potential 1996



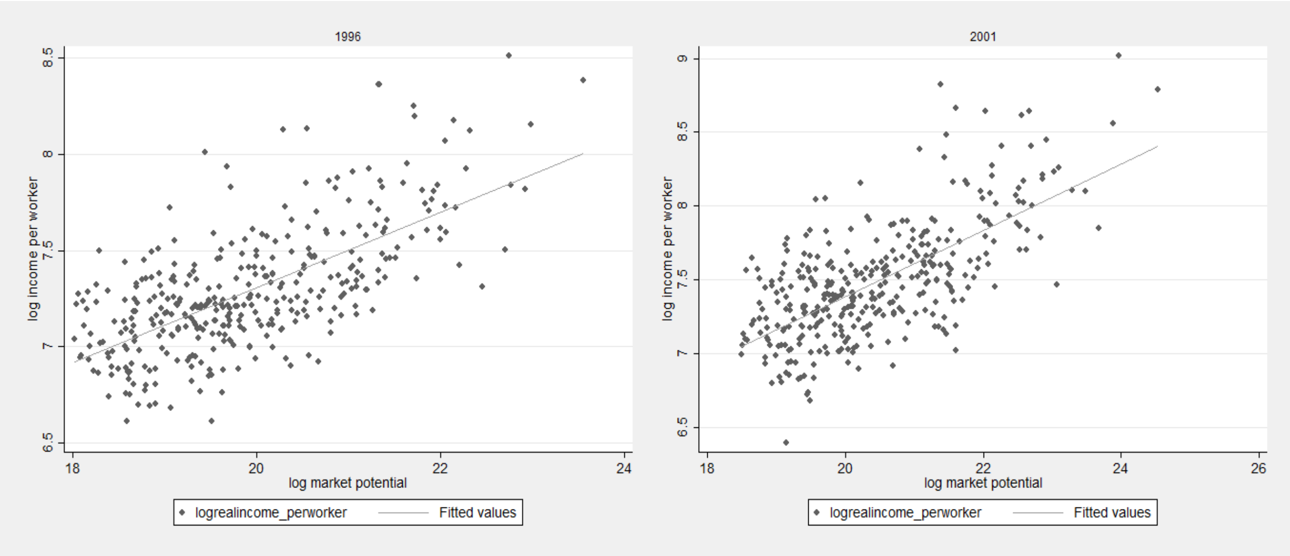
Source: Author's calculations based on census data aggregated to 354 magisterial districts.
Notes: Income per worker is derived by weighting total income from employed individuals with total employed individuals in each region with a positive income and aged 15-64 years. Market potential is based on the Harris (1954) market potential index given by equation (6) in chapter 4, which shows the distance-weighted personal income for each region.

Figure 5.2A: Spatial distribution of income per worker & market potential 2001



Source: Author's calculations based on census data aggregated to 354 magisterial districts.
Notes: Income per worker is derived by weighting total income from employed individuals with total employed individuals in each region with a positive income and aged 15-64 years. Market potential is based on the Harris (1954) market potential index given by equation (6) in chapter 4, which shows the distance-weighted personal income for each region.

Figure 5.3A: Association between regional income per worker and market potential



Source: Author's calculations based on census data aggregated to 354 magisterial districts.
Notes: Income per worker is derived by weighting total income from employed individuals with total employed individuals in each region with a positive income and aged 15-64 years. Market potential is based on the Harris (1954) market potential index given by equation (6) in chapter 4, which shows the distance-weighted personal income for each region.