Explaining changes in post-apartheid income and earnings inequality

Ines **Janina Hundenborn**
HNDINE001

University of Cape Town

Thesis presented for the Degree of
DOCTOR OF PHILOSOPHY
in the School of Economics
Faculty of Commerce
University of Cape Town

Supervisors:
Prof. Murray Leibbrandt (University of Cape Town)
Prof. Ingrid Woolard (University of Stellenbosch)

January 2021

Keywords: Income Distribution; Inequality; South Africa; Personal Income Tax; Labour Market; Micro-Simulations

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

University of Cape Town

Declaration

This work has not been previously submitted in whole, or in part, for the award of any degree. It is my own work. Each significant contribution to, and quotation in, this dissertation from the work, or works, of other people has been attributed, and has been cited and referenced.

Date: 15 March 2021 Signature: Signed by candidate

Abstract

This doctoral thesis analyses the changes in income inequality in post-apartheid South Africa. The thesis adds to the existing literature by explaining the underlying causes of the changes in observed income inequality. As such, this thesis applies different decomposition methods to the Gini coefficient. In the analysis of household income sources, traditional static decompositions are supplemented by applying micro-simulations that allow for a dynamic decomposition of changes in income sources reported in household surveys. The results corroborate previous findings of the significant contribution of labour market incomes and governmen grants. However, the application of advanced dynamic methods highlights the effects of changes in other factors, such as investment income and the role of employed household members, which have previously received less attention.

Further study of household survey data and a unique set of tax administration data enabled a decomposition of the Gini coefficient of taxable income to investigate the effect of high earners on income inequality and the accuracy of capturing them in household surveys. This analysis highlights a significant weighting issues of high earners in the latest wave of the household survey data. Therefore, when combining the two types of data sets, a significant decrease in overall inequality of taxable income can be found between 2011 and 2014. The results ascertain the vast differences between the top and the bottom of the income distribution and concrete policies addressing both sides of the issue need to be implemented in order to overcome persisting income inequality.

Finally, the strong effects of labour market incomes on overall income inequality warrant further investigation. Therefore, changes in earnings inequality are decomposed to assess the effect of changes in the labour market. The application of micro-simulations thereby allows to decompose the changes in earnings inequality into 'price effect' and 'endowment effect' but also to assess the effect of changes in labour market participation, employment, occupational structure and unobserved characteristics. The results show that key drivers of an increase in earnings inequality between 1993 and 2012 were changes in the endowments of working age individuals. This effect was partially counteracted by the price effect. The findings show persisting discrepancies between male and female employment in the labour market and the ongoing marginalization particularly of African women which highlights the need for a revision of existing affirmative action laws and their implementation. The Na-

tional Development Plan offers several strategies for more inclusive growth for South Africa, however, government is already falling behind with its implementation. Therefore, policy makers need to re-examine the efficiency of current social spending and labour laws in order to set the right growth path for the South African economy.

The methods utilized throughout this thesis harmonise different sources of information and enable an integrated analysis of the dynamics of the South African income distribution. The static and dynamic decompositions make use of the 1993 household survey of the Project for Statistics on Living Standards and Development (PSLSD) and the 2008 and 2014 National Income Dynamics Study (NIDS). The assessment of high earners is performed by comparing tax administration data provided by the South African Revenue Service (SARS) for the 2010 and 2014 tax years with household survey data from NIDS in 2011 and 2014. Finally, the decomposition of earnings inequality is carried out using data sets from the Post-Apartheid Labour Market Series (PALMS) between 1993 and 2012.

Acknowledgements

I would like to thank my supervisors Professor Ingrid Woolard and Professor Murray Leibbrandt for their continuous advice, support, and patience. I am indeed very grateful to have had such knowledgeable and inspiring guidance for my academic and professional development.

Friends and colleagues at the Southern Africa Labour and Development Research Unit (SALDRU) and the National Income Dynamics Study (NIDS) have created a work environment that built the foundation for this research and allowed me to grow beyond the realms of this thesis. Specifically, I would like to thank Tim Brophy for his mentoring and encouragement.

I am very grateful to my supervisors for arranging financial support through a PhD scholarship from SALDRU. At different phases of the PhD, the bursary was co-funded through my work on the Programme to Support Pro-poor Policy Development (PSPPD), the UNU-WIDER "Inequality in the giants" project and NIDS. I gratefully acknowledge the support from SALDRU and each of these funders.

Without the reassurances and assistance from friends and family, this thesis would not have been possible.

Contents

1 Introduction **1**

1.1 Motivation for the Thesis . . . 2

1.1.1 Macro-Economic Context . . . 3

1.2 Structure of the Thesis . . . 5

2 Drivers of Inequality in South Africa **9**

2.1 Introduction . . . 10

2.2 Descriptive Statistics . . . 11

2.3 Gini Decompositions . . . 16

2.3.1 Static Approach . . . 16

2.3.2 Micro-Simulations . . . 21

2.4 Conclusion . . . 32

Appendices **35**

2.A Extended Methodology . . . 35

2.B Additional Figures and Tables . . . 37

3 The Effect of Top Incomes on Inequality **39**

3.1 Introduction . . . 40

3.2 Background . . . 43

3.2.1 The Tax System in South Africa . . . 43

3.3 Data Sources . . . 43

3.4 Distributional Analysis . . . 45

3.4.1 Overview of the Household Survey Data . . . 45

3.4.2 Overview of Data on Taxpayers . . . 49

3.4.3 Comparison of Household Survey Data with Tax Administration Data 52

3.5 Measuring Inequality . . . 55

3.5.1 Inequality within the NIDS and PIT Data . . . 56

3.5.2 Inequality across the NIDS and PIT Data . . . 57

3.5.3 Combining Household Survey Data and Tax Administration Data . . . 61

3.6 Conclusion . . . 66

Appendix **70**
3.A Additional Tables and Figures 70

4 Evolution of Earnings Inequality in the Post-Apartheid Era **72**
4.1 Introduction 73
4.2 Literature Review 74
4.2.1 Literature on the South African Labour Market 75
4.2.2 Literature on Micro-Simulations 76
4.3 Data Issues 78
4.3.1 Summary Statistics 79
4.4 Methodology 87
4.4.1 Simulating Labour Market Outcomes 88
4.4.2 Counterfactual Micro-Simulations 99
4.5 Results 105
4.6 Conclusion 110

Appendix **115**
4.A Counterfactual Results 115

5 Conclusion **118**
5.1 Key Findings and Results 119
5.2 Contributions to Understanding Income Inequality in South Africa 121

Data Sources **124**

Literature **124**

List of Tables

2.1 Household Income Components in Per Capita Terms . . . 13
2.2 Household Composition from 1993 to 2014 . . . 15
2.3 Static Decomposition of the Gini Index by Income Sources . . . 18
2.4 Dynamic Decompositions of Per Capita Income Sources 1993 to 2008 and 2008 to 2014 . . . 25
2.5 Dynamic Decompositions - 1993 to 2008 and 2008 to 2014 Including Household Composition and Different Ranking Variables . . . 27
2.6 Dynamic Decompositions of Per Capita Income Sources 1993 to 2014 . . . 37
2.7 Dynamic Decompositions - 1993 to 2014 Including Household Composition and Different Ranking Variables . . . 38

3.1 Taxable Income - Percentiles in the NIDS Data . . . 46
3.2 Zero Earners in the NIDS Data (Percentages) . . . 48
3.3 Income Brackets in the PIT Data . . . 49
3.4 Demographics in the PIT Data . . . 51
3.5 High Earners' Income Distribution . . . 52
3.6 Gini Coefficients at Different Thresholds . . . 56
3.7 Gini Coefficients Combined at Different Thresholds . . . 65
3.8 NIDS Attrition Rate by Deciles . . . 70
3.9 Statutory Tax Rates in South Africa . . . 70
3.10 Comparison of the Information in the PIT and NIDS Data sets . . . 71

4.1 Individual Characteristics . . . 81
4.2 Household Level Characteristics . . . 83
4.3 Participation Rates of Working Age Population . . . 84
4.4 Employment Rates Conditional on Participation . . . 85
4.5 Distributional Statistics of Real Earnings . . . 86
4.6 Bivariate Probit Results with Selection Correction . . . 91
4.7 Results of Wage Regression (in logs) . . . 96
4.8 Ordered Logistic Regression of Occupational Choice . . . 98
4.9 Counterfactual Estimations . . . 106
4.10 Counterfactual Gini Simulations . . . 116

4.11 Decomposition of Average Effects for Women 1993 - 2012 117
4.12 Decomposition of Average Effects for Men 1993 - 2012 117

List of Figures

1.1 Real Annual Growth Rates of total GDP in South Africa since 1990 4

2.1 Box 2 in Azevedo et al. (2013) 37

3.1 Difference in Estimated Mean Taxable Income by Income Bracket between NIDS and PIT 53
3.2 2011 Taxable Income in NIDS vs. PIT 58
3.3 2011 Taxable Income in NIDS vs. PIT above the Filing Threshold 59
3.4 2014 Taxable Income in NIDS vs. PIT 60
3.5 2014 Taxable Income in NIDS vs. PIT above the Filing Threshold 60

4.1 Decomposition of Average Effects for Women 1993 - 2012 107
4.2 Decomposition of Average Effects for Men 1993 - 2012 109

Abbreviations

LFS	Labour Force Survey
NDP	National Development Plan
NIDS	National Income Dynamics Study
OB	Oaxaca-Blinder
OHS	October Household Survey
PALMS	Post-Apartheid Labour Market Series
PIT	Personal Income Tax
PSLSD	Project for Statistics on Living Standards and Development
SALDRU	Southern Africa Labour and Development Research Unit
SARS	South African Revenue Service
SDG	Sustainable Development Goals
Stats SA	Statistics South Africa
RIF	Re-centred Influence Function
QLFS	Quarterly Labour Force Survey

Chapter 1

Introduction

1.1 Motivation for the Thesis

The keen interest in explaining income inequality in a country with the highest level of income and earnings inequality in the world is inevitable. Investigating the causes and consequences of such high levels of inequality has therefore always been a key concern for scholars and policy makers. However, this thesis is motivated not only by these high levels of income inequality but also driven by the curiosity about what has led to the changes observed in income inequality in the last twenty-five years. Therefore, the principal questions that motivate this thesis are "What are the drivers of inequality in South Africa for household income and individual earnings, and what has been preventing South Africa from growing into a more equal and inclusive society?"

Democratic institutions and affirmative action laws have been implemented in an attempt to redress the effects of the legislated discrimination under the apartheid regime. However, the economic gains made by the country have failed to improve the welfare of the majority of South Africans. More so, the perception of income inequality seems to have worsened for many since 1993 (Alexander, 2010). This is particularly alarming seeing the goals of eliminating poverty and reducing inequality that the South African government has set out in the National Development Plan (National Planning Committee, 2013). The National Development Plan (NDP) aims to improve the lives of South Africans by introducing policies to reduce poverty such that the proportion of households with a monthly income below R419 per person (in 2009 prices) should be reduced from 39% to zero. Additionally, the NDP plans to reduce the Gini coefficient from 0.69 to 0.6 and eliminate gender inequality. These goals should be achieved by 2030 mainly through inclusive economic growth and structural reforms to public services, education and infrastructure as well as other measures that aim at social cohesion. The NDP was drawn up on the basis of the United Nations' Millennium Development Goals. Their successor, the Sustainable Development Goals (SDGs), formulate relevant goals similar to those in the NDP. As such, SDG 5 aims to "Achieve gender equality and empower all women and girls", SDG 8 aims to "Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all"and SDG 10 wants to "Reduce inequality within and among countries". The fact that income inequality features in different aspects in the goals set out not only by the national government but also by the United Nations emphasizes the significance of this issue and explains why further research into this topic is warranted.

The focus of researchers studying the high prevailing levels of inequality, poverty and unemployment thus far has been on investigating and decomposing income inequality at a certain point in time, particularly in the case of household income inequality (for example Leibbrandt et al., 2012; Hoogeveen and Özler, 2006 or Leibbrandt, Levinsohn and McCrary,

2010). With regards to individual incomes, some research has been done to analyse income and earnings inequality that has implemented more dynamic methods such as the recentered influence function (see for example Finn and Leibbrandt, 2018; Kwenda and Ntuli, 2018) or unconditional quantile regressions (for example Ntuli and Wittenberg, 2013). These studies, which have been essential to the understanding of the current levels of income and earnings inequality, have either fallen short in explaining household income dynamics over time or, when explaining the dynamics of earnings in particular, have failed to account for one of South Africa's most prevailing problems, structural unemployment.

Beyond the traditional analyses of both, household income inequality and individual incomes, it matters how the different parts of the income distributions are accounted for. Research has shown that extremely high-income earners may be under-represented in household surveys or understate their incomes, however, their contributions can have a significant effect on measuring income inequality (Atkinson, 2007). The overall remedy has been to use tax records in order to account for income earners at the top of the income distribution. While Alvaredo and Atkinson (2010) have provided an extensive analysis of top incomes in South Africa between 1903 and 2007, there has not been an attempt to link tax administration data to a household survey for the analysis of income inequality in South Africa.

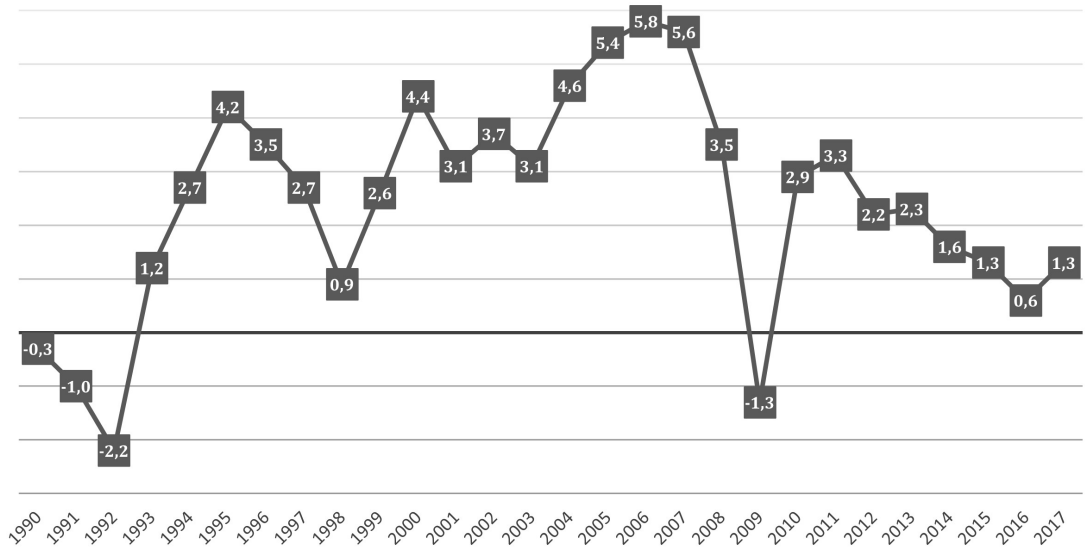
Therefore, this thesis strives to remedy the shortcomings of existing research and fill the extant gap in the South African literature by applying novel methods that allow for dynamic decompositions, integrate top earners in the measurement of income inequality and account for structural unemployment as well as other factors of the labour market while analysing changes in earnings inequality.

1.1.1 Macro-Economic Context

The issue of inequality is particularly ripe in the South African context, yet, the results of the government's commitment towards greater economic equality and redistribution in the post-apartheid era are coming up short for the majority of the population. When analysing the effectiveness of fiscal policy in redistributing income, Inchauste et al. (2015) show that in South Africa the levels of poverty and inequality remain among the highest in middle-income countries. This is despite highly progressive social spending programs and the (slightly) progressive design of the overall tax system and the personal income tax in particular. What is most troubling is the fact that even though fiscal policy significantly reduces the Gini coefficient of income, it is still higher after this reduction than the Gini coefficient of other middle-income countries before policy intervention (Inchauste et al., 2015).

Overall, South Africa does not fare well in comparison to other middle-income countries

when it comes to economic inequality. Income per capita increased in South Africa between 1992 and 2012 as it did in other middle-income countries like Brazil, Mexico or Thailand (Levy et al., 2015).[1] However, several studies find that the level of inequality in South Africa continues to exceed those of comparable middle-income countries and in some instances increased in South Africa where income inequality decreased in other countries. Leibbrandt and Finn (2012) find a Gini coefficient of household income per capita of 0.7 in South Africa in 2008 whereas Brazil reported a Gini coefficient of 0.55 in the same year. Income inequality in Brazil fell from previously 0.6 whereas South Africa's Gini increased from previously 0.64. The high levels of inequality translate into higher levels of poverty in South Africa compared to its peers. Although progressive fiscal policy achieves a significant reduction in poverty, the headcount ratio in South Africa measured at the US$ 2.50 per day poverty line remains high at 36% compared to Brazil, for example, at 11% (Inchauste et al., 2015).



Source: Statistics South Africa

Figure 1.1: Real Annual Growth Rates of total GDP in South Africa since 1990

Since the end of apartheid, the South African government has efficiently expanded fiscal programs and broadened the tax base in order to reduce poverty and inequality. However, these extensive efforts have not translated into the equivalent results. Bourguignon (2004) finds that in highly unequal countries, growth rates for mean income earners are higher than for individuals at the bottom of the income distribution. There is empirical evidence for this in South Africa. Levy et al. (2015) point out that "relative to other middle-income countries, South Africa has an unusually small fraction of the population that gains directly from sustained economic growth." Those who benefit from economic growth are in the top

[1]Brazil, Mexico and Thailand are chosen due to the fact that they share similar characteristics with South Africa. As such, the average income, population size and therefore the development challenges faced by these countries can be assumed to be comparable (Levy et al., 2015).

30% of the income distribution whereas in other middle-income countries, average income earners gain from economic growth. Additionally, Figure 1.1 shows that even though the annual growth rates increased up to 5.8% since the end of apartheid, growth rates took a fall and subsequently slowed down significantly since the global financial crisis in 2007/2008. The decline of the growth rate limits the degree for possible expansion of progressive social spending. This is particularly troubling seeing the strong positive effects of the government's social spending on inequality which will be shown in Chapter 2.

These macro indicators, coupled with previously high levels of fiscal debt and a large fiscal deficit, allow little room for further fiscal policies to bring about greater redistribution. The problem may be aggravated since the downgrading of South Africa's debt to "junk status" by international rating agencies.[2] In order to significantly reduce poverty, unemployment and inequality, the National Development Plan (NDP) foresees an average annual growth rate of 5.4% until 2030. Figure 1.1 highlighted the degree to which the growth rates fall short on this ambitious plan and inequality remains high. This is particularly concerning since redistribution of income can contribute to increased growth (Bourguignon, 2004). The continuously high levels of inequality in South Africa motivate the in-depth analysis of the income distribution both, at the top end and the bottom end, provided in Chapter 3.

1.2 Structure of the Thesis

This thesis investigates changes in income and earnings inequality by taking three distinct but related issues of income inequality and examining questions connected to these. The first issue, which focuses on explaining the changes in household income inequality, is substantively discussed in Chapter 2. The second issue is addressed in Chapter 3 and investigates the distribution of individual incomes when accounting for high-income earners. The final and third issue in Chapter 4 studies how changes in the underlying determinants of wages affect changes in the earnings distribution of working women and men.

Chapter 2 of this dissertation discusses the following questions:

- What are the drivers of household income inequality in South Africa?
- How do the dynamics of micro-determinants of inequality explain changes in aggregate income inequality?
- What effects have household characteristics had on changes in overall income inequality?

[2]Standard & Poor's rating agency rated South Africa's debt as speculative grade or "junk"in April 2017, other international rating agencies are currently reviewing the South African status.

Chapter 3 investigates the following:

- What are the trends in the income distribution in household survey data compared to tax administration data?
- How can data sets be combined to critically assess trends in individual incomes across the entire distribution?
- How does the level of assessed income inequality change when tax administration data is included in the analysis?

Finally, Chapter 4 asks:

- How can changes in the determinants of wages explain changes in earnings inequality over time? How do these effects differ between men and women?
- How do changes in endowments, returns to endowments, occupational choice and demographic variables affect changes in the female/male distribution of wages?
- How do changes in unemployment and participation affect changes in inequality of earnings for women and men?

Each chapter contains separate discussions of the relevant literature because of the focusses on different aspects of income inequality in South Africa. Furthermore, all three substantial chapters use the Gini coefficient as a measurement of inequality. The Gini coefficient is a widely used measurement of inequality. This measurement puts a higher weight on the middle of the distribution and is therefore less sensitive to changes at the top or the bottom of the distribution (see for example Cowell, 2000). This can also lead to certain shifts in the income distribution that would not change the Gini coefficient. For example, imagine one economy where 50% of the population receive no income at all while the other 50% share all of the remaining income equally; this economy would have the same Gini coefficient as an economy in which 75% of the population equally share 25% of total income, while the remaining 25% of the population split the remaining 75% of income equally. Intuitively, the second economy may be considered more equal but such a difference in incomes would not be reflected in the Gini. Despite such short-comings, the Gini coefficient is the sole measurement that enables the various modifications and decompositions performed in the different chapters. Particularly the method used in Chapter 3 is designed for the Gini coefficient rather than other measurements of inequality. Therefore, the Gini is chosen as measurement of inequality in this thesis.

The data sources used in each chapter reflect the requirements that the different research methods place on the data. Therefore, Chapter 2 uses data from the Project for Statistics on Living Standards and Development (PSLSD) together with more recent data from

two waves of the National Income Dynamics Study (NIDS) to assess the effect of different income sources and household composition on the changes in inequality between 1993 and 2014. Traditional static decompositions by income source (see for example Leibbrandt et al. 2010 and Leibbrandt et al. 2012) are extended in this chapter by using micro-simulations to assess how much the different income sources have contributed to changes in income inequality between two years of observation. Through the micro-simulations approach by Azevedo et al. (2013), it is possible to decompose *dynamics* in income inequality rather than decomposing *levels* of inequality. Additionally to the decomposition by income source, micro-simulations make it possible to draw out the contribution of changes in underlying household demographics. Changes in the household composition can lead to changes in income inequality which is assessed as household income per capita. These critical effects are overlooked when only decomposing inequality by income source.

Chapter 3 combines data from the household survey NIDS with data on individual tax records made available for this study by the South African Revenue Service (SARS). SARS provided 20% samples for the 2011 and the 2014 tax year respectively. The income distributions drawn by SARS will be compared with the income distributions presented in the two corresponding waves of the household survey. Findings that incomes at the top tail of the income distribution are often under-represented or understated in household surveys (Atkinson, 2007; Burkhauser et al. 2012) can be tested due to exclusive access to individual tax records. This enables an analysis of the state of the South African household survey with regards to capturing the top tail of the income distribution. Additionally, the two types of surveys are used to create income inequality measurements that actively account for top-income earners. The chapter compares different methods of combining measurements of income inequality, such as the ones by Alvaredo (2011), Diaz-Bazan (2015) or Jenkins (2017), and highlights the benefit of following the more pragmatic approach.

Finally, Chapter 4 applies a dynamic decomposition approach to several waves of the Post-Apartheid Labour Market Series (PALMS). Again, micro-simulations are utilised for an assessment of the development of inequality between 1993 and 2012. In this chapter, the focus is on earnings inequality and how the effects of changes in the labour market may differ between men and women. Of particular interest in the South African context is the effect of structural unemployment on the evolution of earnings inequality - a feature that has been left out in many studies addressing earnings inequality in South Africa (for example Bhorat and Goga, 2013; Wittenberg 2016a,b). However, structural unemployment leads to distortions in the labour market that prevent individuals with otherwise high probabilities of employment from finding a job. Therefore, following González-Rozada and Menendez (2006), the micro-simulation approach by Bourguignon et al. (2008) is extended by a component of randomness in order to account for the effect of distortions through high unemployment.

Additionally, micro-simulations are used to assess changes in participation and employment, as well as changes in the occupational structure of the labour market, individuals' characteristics and returns to those characteristics. Through expansive counterfactual analysis, the effect of changes in each of these components on changes in earnings inequality can be broken down.

Chapter 5 presents the key findings of this study and concludes with a summary of the contributions made by this thesis.

Chapter 2

Drivers of Inequality in South Africa

2.1 Introduction

Levels of inequality have remained high in the recent history of South Africa and analysing the drivers of these inequalities is of importance to both, researchers and policy makers. There exists a well-established literature that investigates the drivers of inequality in South Africa by decomposing income inequality by groups (race and location in particular) or by income sources using South Africa's available household surveys (see Leibbrandt et al., 2012; Leibbrandt et al., 2010). Decompositions of inequality by income source present the opportunity to determine which source(s) led to the overall high levels of inequality while providing a foundation on how to address inequalities in those income sources specifically.

This chapter starts by updating existing work on inequality decomposition using the latest round of National Income Dynamics Study (NIDS) data from 2014/2015. It compares the 1993 Project for Statistics on Living Standards and Development (PSLSD) data, the 2008 NIDS data and the most contemporary picture in detail. For this chapter, interim waves of NIDS have been excluded as static decompositions have been done for these years (see for example Leibbrandt et al., 2010). By offering a long-term analysis from 1993 to 2014, this chapter makes a valuable contribution to existing research on this topic. The decomposition methods utilized in the existing literature are useful but only loosely indicative of drivers of inequality. This chapter complements current research by following a methodology developed by Lerman and Yitzhaki (1985) and models the changes in the distribution of per capita income with a particular focus on the role of different income sources before applying micro-simulations as outlined by Azevedo et al. (2013). Rather than assessing the state of inequality at a given point in time, the use of micro-simulations allows for an assessment of how changes in income sources and relevant demographic factors have influenced changes in inequality over time.

The results show a drop in overall inequality between 1993 and 2014 despite a temporary increase between 1993 and 2008. The Gini coefficient increased between 1993 and 2008 from 0.681 to 0.690 and dropped to 0.655 in 2014. The decrease between 2008 and 2014 seems to be primarily driven by a decrease in inequality in income from labour market sources. The static decomposition by Lerman and Yitzhaki (1985) shows a strong correlation between labour market incomes and total household income and reports a decrease in inequality within this income source. Furthermore, the static approach indicates that government grants had a decreasing effect on inequality, and that remittances have at least the potential to lower inequality.

Applying the dynamic decomposition method made it possible not only to evaluate the effect of changes in different income sources on the overall Gini but also to account for

changes in household composition. While the effects of the increasing share of adults in the household are small, they seem to be driving inequality slightly upwards. The share of employed adults, however, had a decreasing effect between 1993 and 2008. Once these household composition variables are accounted for, the change in the Gini coefficient is driven mainly by changes in the different income sources. The strongest driver was found to be labour income which increased the Gini coefficient by 0.068 points or 10% between 1993 and 2008 and led to a 0.039 or 5.9% decrease in the Gini coefficient between 2008 and 2014. However, the inequality increasing forces of labour income between 1993 and 2008 were largely offset by redistributive efforts by the government through government grants. Other notable changes were driven by investment income, while remittances seem to have had only small effects.

The analysis of this chapter starts by reviewing NIDS data in detail in Section 2.2. Furthermore, the characteristics of NIDS will be compared with the PSLSD data. Section 2.3.1 continues by outlining and applying the so-called static decomposition method by Lerman and Yitzhaki (1985) and extended by Stark et al. (1986) to the PSLSD data set from 1993, NIDS data of 2008 and 2014/2015 before applying a dynamic approach using micro-simulations to the time frame from 1993 to 2008 and 2008 to 2014/2015 respectively in Section 2.3.2. Finally, Section 2.4 concludes by summarizing the main results and offering some policy advice concerning the high levels of inequality South Africa continues to face.

2.2 Descriptive Statistics

The data sets used in this chapter stem from the 1993 Project for Statistics on Living Standards and Development (PSLSD) and the National Income Dynamics Study (NIDS). Using these different data sets enables a comparison of levels of inequality prevalent at the end of the apartheid period and over the last twenty years by applying both, static and dynamic decomposition methods.

Both data sets offer detailed information collected on income from different sources. The PSLSD was a household level survey in which one individual answered all questions regarding the household (Leibbrandt et al., 2012). The household questionnaire comprised areas such as demography, household services and expenditure, land access and use, employment and other income, health and educational status and anthropometry (SALDRU, 2012). Similarly, the NIDS data set offers great detail in the information collected on income from different sources as well as information on a household and individual level. To some degree, NIDS individual data may be considered particularly reliable, based on the fact that in addition to the household questionnaire, NIDS contains individual-specific questionnaires administered to all resident members of the household. As opposed to the one-time study of

the PSLSD, NIDS is a nationally representative panel survey of South African individuals (NIDS, 2013). Every two years, the study collects information on households and individuals with regards to a wide range of topics including labour market participation, individual and household incomes from employment and non-employment sources as well as data on wealth, individual health and well-being, and education.

Leibbrandt et al. (2012) compare and contrast the PSLSD and NIDS data sets in much detail and conclude that the two data sets are largely comparable except for data collected on agricultural income and imputed rent. Particularly agricultural income plays only a small role in overall household income, therefore, these income categories will not be used in the analysis performed in this chapter. In order to compare changes over time, this chapter will use the 1993 PSLSD data and compare it to NIDS data collected in 2008 as well as the data collected in 2014/2015. For simplicity, incomes will be referred to as 2014 incomes from here on as all numbers will be represented in 2014 prices.[1] NIDS provides a set of derived data sets that contain the relevant income variables for the analysis performed in this chapter. These derived variables include some imputations which can be reproduced following the documentation provided by NIDS.[2]

With the information provided by PSLSD and NIDS, household income can be disaggregated into income from labour market sources, such as a salary paid by an employer or income from self-employment, and income from several non-employment sources. Labour incomes in both data sets are net of taxes. The non-employment income sources include income from government grants (such as child support grants, pensions and other government support), income from remittances (including contributions in kind as well as monetary transfers made to an individual in the household) and income from investment sources such as loans, stocks, annuities and rental income.

Table 2.1 presents descriptive statistics of household income and its different income components. All components are represented in 2014 prices in per-capita terms. There is evidence of a significant increase in total household income per capita and as such, income from all income components has increased from 1993 to 2014 in real terms. There has been a slight decrease in overall income inequality with the Gini coefficient falling from 0.68 to 0.66 between 1993 and 2014, despite an increase of the Gini to 0.69 in 2008.

Labour income holds the largest share of total household income among the different income sources. A majority of households receive income from labour market sources, and the pro-

[1] The author uses the monthly updated statistical release of the consumer price index by Statistics South Africa for this conversion. A single national price deflator was used as the data on the consumer price index is not reliable enough for differentiation below the national level.

[2] For more information see the NIDS website: *nids.uct.ac.za*.

Table 2.1: Household Income Components in Per Capita Terms

Variable	**1993**	**2008**	**2014**
Total Household Income			
Mean of HH Income	1328.17	2062.68	2398.57
Gini of HH Income	0.68	0.69	0.66
Labour Income			
Mean of Labour Income	1078.18	1659.86	1971.98
Share in total HH Income	81.18%	80.47%	82.21%
Proportion of HHs receiving Labour Income	60.5%	64.4%	72.6%
Gini of Labour Income	0.73	0.76	0.73
Income from Government Grants			
Mean of Govt Grants	86.17	161.31	187.34
Share in total HH Income	6.49%	7.82%	7.81%
Proportion of HHs receiving Govt Grants	23.5%	56.3%	68.0%
Gini of Government Grants	0.92	0.78	0.76
Income from Remittances			
Mean of Remittance Income	50.56	86.69	93.94
Share in total HH Income	3.81%	4.2%	3.92%
Proportion of HHs receiving Remittances	22.2%	13.9%	38.3%
Gini of Remittances	0.91	0.97	0.91
Investment Income			
Mean of Investment Income	113.28	154.81	145.31
Share in total HH Income	8.53%	7.51%	6.06%
Proport. of HHs receiving Investment In.	3.5%	5.6%	23.3%
Gini of Investment Income	0.99	0.97	0.98
N_unweighted	39 180	28 225	37 965
N_weighted	39 020 805	49 295 750	54 941 051

Source: Author's calculations using PSLSD and NIDS weighted.

Note: HH stands for household.

portion of households receiving labour income has steadily increased from 60.5% in 1993 to 64.4% of all households in 2008 and 72.6% in 2014. The Gini coefficient for this component is relatively large and has fallen slightly by 0.03 points from 0.76 to 0.73 between 2008 and 2014. Considering the strong dependency on this type on income, this slight decrease in the Gini of income from labour market sources as well as the fact that, compared to 1993, many more households receive this type of income may be the driver of the overall decrease in income inequality within the observed period. This will be investigated in more detail later on in the chapter.

Within the same period, income from government grants has increased as has the proportion of households receiving this type of income. The proportion of households receiving some form of government support rose significantly from 23.5% in 1993 to 56.3% in 2008 and as far as 68% in 2014. Income from government grants reports a relatively large Gini coefficient which most likely stems from the fact that many households are reporting zero income from this source. Overall, the Gini of income from government grants decreased slightly from 0.78 in 2008 to 0.76 in 2014. However, the Gini has fallen by 0.16 points since 1993 which is indicative of the fact that so many more households receive grant income by 2014.

In addition, the proportion of households receiving remittances has increased notably between 1993 and 2014, despite a significant drop in 2008. Previously, 22% of households received income in the form of inter-household transfers in cash and in kind. This proportion dropped to a low of 14% in 2008. In 2014, 38% of households received this form of income. Since fewer households are reporting zero income from remittances, the distribution of this type of income has become less unequal. This would also explain the decrease from a Gini of 0.97 to 0.91 between 2008 and 2014. The overall share of remittances in total household income is still relatively small and has increased between 1993 and 2008 from 3.8% to 4.2% only to decrease to 3.9% in total household income in 2014.

Lastly, a strong increase can be observed in the proportion of households receiving income from investment sources between 1993 and 2014. When previously only 3.5% of households reported income from investment, 23.3% report this type of income in 2014. Investment income remains highly unequal over the observed period with a Gini coefficient close to one in all three years. The fact that the share of investment in total household income has decreased would indicate that other income components, namely labour income and government grants, have grown more strongly than investment income and play a more important role in the income composition of the households.

Table 2.1 discussed the development of income per capita in real terms in 1993, 2008 and

Table 2.2: Household Composition from 1993 to 2014

Variable	**1993**	**2008**	**2014**
Household Size	4.38	3.53	3.21
Number of Adults in HH	2.81	2.41	1.94
Number of Employed in HH	1.08	0.96	0.88
Share of Adults in HH	73.37%	77.71%	73.84%
Share of Employed in HH	36.71%	37.57%	41.68%

Source: Author's calculations using PSLSD and NIDS weighted.

2014. In order to interpret these changes, however, it is imperative to study changes in the underlying demographic variables as well. Thus, Table 2.2 presents descriptive statistics for household composition variables. Household size has decreased from 4.38 in 1993 to 3.5 people in an average household in 2008 and then to 3.2 persons on average in 2014. Households are constituted mostly of adults, rather idiosyncratically defined as persons aged 15 years and older. This threshold has been chosen since it presents the start of the working age in South Africa. Although both numbers fall in absolute terms, the share of adults in the household fluctuates from 73% in 1993 to 78% in 2008 and then to just under 74% in 2014. The number of employed adults decreased slightly from an average of just more than one person per household in 1993 to an average of 0.88 employed adults in a household in 2014. Due to the decrease in overall household size, however, this implies that the share of employed adults actually increased from 37% in 1993 compared to a share of almost 42% in 2014. It is important to note that any form of employment activity was accounted for, not only a formally paid job for an employer but also any form of self-employment and other labour market activities were counted into this variable.

The discussion of the variables in this section has shown that the PSLSD and the NIDS data sets offer detailed information on income components and household demographics as a platform to investigate their role in the development of income inequality over time. The remainder of this chapter proceeds to assess the drivers of inequality first by applying a static approach to analysing the effects of different income sources on overall income inequality. These results will be extended by using micro-simulations that allow a dynamic analysis of these changes over time.

2.3 Gini Decompositions

2.3.1 Static Approach

This subsection utilizes an approach conceptualized by Shorrocks (1982) and extended by Lerman and Yitzhaki (1985) and Stark et al. (1986)[3] for a static decomposition of the Gini coefficient. The index presented here is a Gini coefficient decomposed into the different sources of income. This chapter will call this a static approach as income inequality is decomposed by different income sources as it is observed at a particular moment in time. However, by taking the derivative with respect to a small percentage change in income from a particular source, Stark et al. (1986) analysed the effect of a marginal change in an income source on the overall Gini coefficient at that point in time, holding all other income sources constant. This section will briefly analyse the methodology of this static approach before comparing the decomposition of 1993 income source data to data from 2008 and 2014.

Following Stark et al. (1986), the overall Gini coefficient G_0 can be presented as follows.

$$G_0 = \sum_{k=1}^{K} R_k \cdot G_k \cdot S_k \tag{2.1}$$

where S_k and G_k are the share and the Gini coefficient of income component k respectively.[4] R_k represents the so-called Gini correlation of component k with total household income. It shows similar characteristics to the Pearson's and Spearman's correlation coefficients.

As such, equation (2.1) makes it possible to examine three important concepts:

1. the share of the respective income source in overall household income, S_k
2. the inequality within the different income sources, G_k, and
3. the (Gini) correlation R_k between income component k and total household income.

By definition, the share of an income source in overall household income S_k and the Gini coefficient of any income source G_k are always positive and bounded between 0 and 1. The Gini correlation R_k, however, will be positive when an income component contributes positively to the overall Gini, that is when y_k is an increasing function of total income y_0. Correspondingly, R_k will be negative when income component y_k is a decreasing function of total income y_0. R_k is bounded by $-1 \leq R_k \leq 1$ and will be equal to zero when y_k and y_0 are uncorrelated.

[3] This chapter will follow the notation of Stark et al. (1986).

[4] The decomposition and the methodology are discussed in more detail in Appendix 2.A.

In addition to the three concepts outlined above, the assessment of the effect of a small change in any one of the income components k on the overall Gini will be of interest. For this purpose, assume that an exogenous change in any income component k by a factor e occurs. Then, income from k is assumed to change according to $y_k(e) = (1 + e)y_k$ and

$$\frac{\partial G_0}{\partial e} = S_k(R_k \cdot G_k - G_0). \tag{2.2}$$

Equation (2.2) is a partial derivative which simulates a marginal change in a particular income source while holding income from other sources constant. Further, dividing (2.2) by G_0 yields

$$\frac{\partial G_0/\partial e}{G_0} = \frac{S_k \cdot R_k \cdot G_k}{G_0} - S_k. \tag{2.3}$$

Thus, the change in overall inequality due to a small change in income component k is equal to the initial share of k in total inequality less the share of component k in total household income (Stark et al., 1986). Given the characteristics of R_k, this yields two possible outcomes for the overall Gini coefficient. If income component k has a negative or zero correlation between k and total household income y_0, an increase in income from component k will have an equalizing effect, thereby lowering inequality. This is due to the fact that S_k, the share of income from component k, as well as the Gini indices for k and total income, G_k and G_0, are always positive. The other possible outcome is when R_k represents a positive Gini correlation. Assuming that $G_k > G_0$, then the sign of $\frac{R_k \cdot G_k}{G_0}$ determines the impact on inequality and whether there is an increase in inequality associated with component k. $G_k > G_0$ is a necessary condition for an inequality-increasing effect of income component k given that R_k is always smaller or equal to 1.

The results of the decomposition method by Stark et al. (1986) are presented in Table 2.3 for 1993, 2008 and 2014. Equation (2.1) has shown that the Gini coefficient is the sum of the products of the first three columns of Table 2.3 for each component k. These products of $S_k \cdot R_k \cdot G_k$ are reported as the (absolute) contribution in Column (4). The results in Table 2.3 show that income from labour market sources is the most significant contributor to household income inequality. In 1993, labour income contributed 84.4% to overall inequality. This inequality share increased to 87.2% in 2008 and 90.2% in 2014 as can be seen in Column (5) of Table 2.3. The Gini coefficient of labour income was 0.73 in 1993, 0.76 in 2008 and 0.73 in 2014. Labour income is also the most strongly correlated of all income sources with the Gini coefficient. The Gini correlation (or R_k) is close to one for all three years under observation. In this context, it is important to point out that the drop in the Gini of labour income between 2008 and 2014 seems to have contributed to the substantial decrease in overall income inequality as measured by the Gini coefficient. While the Gini of labour income decreased by 0.03 between 2008 and 2014, the relative contribution increased

Table 2.3: Static Decomposition of the Gini Index by Income Sources

	Income share S_k	**Gini correlation** R_k	**Gini index** G_k	**Contribution** $S_k \cdot R_k \cdot G_k$	**%-Contribution** $\frac{S_k \cdot R_k \cdot G_k}{G}$	**Elasticity** $\frac{\partial G/\partial e}{G_0}$
			1993 - PSLSD			
Labour Income	0.820	0.959	0.730	0.575	0.844	0.024
	0.011	*0.003*	*0.004*	*0.009*	*0.015*	-
Government Grants	0.061	0.418	0.924	0.024	0.035	-0.026
	0.004	*0.038*	*0.004*	*0.004*	*0.005*	-
Remittances	0.031	0.115	0.913	0.003	0.005	-0.026
	0.002	*0.050*	*0.005*	*0.002*	*0.002*	-
Investment	0.088	0.915	0.988	0.079	0.117	0.029
	0.010	*0.012*	*0.001*	*0.010*	*0.014*	-
Total	1.000	-	0.681	0.681	1.000	-
			2008 - NIDS			
Labour Income	0.825	0.962	0.759	0.602	0.872	0.047
	0.017	*0.005*	*0.009*	*0.017*	*0.024*	-
Government Grants	0.054	-0.047	0.776	-0.002	-0.003	-0.057
	0.004	*0.034*	*0.007*	*0.001*	*0.002*	-
Remittances	0.036	0.590	0.970	0.021	0.030	-0.006
	0.014	*0.168*	*0.011*	*0.014*	*0.020*	-
Investment	0.085	0.846	0.970	0.070	0.101	0.016
	0.012	*0.022*	*0.003*	*0.011*	*0.016*	-
Total	1.000	-	0.690	0.690	1.000	-
			2014 - NIDS			
Labour Income	0.838	0.964	0.731	0.591	0.902	0.064
	0.017	*0.004*	*0.011*	*0.020*	*0.025*	-
Government Grants	0.047	-0.187	0.758	-0.007	-0.010	-0.057
	0.003	*0.021*	*0.006*	*0.001*	*0.001*	-
Remittances	0.038	0.307	0.914	0.011	0.016	-0.022
	0.003	*0.040*	*0.004*	*0.002*	*0.003*	-
Investment	0.077	0.796	0.978	0.060	0.092	0.015
	0.017	*0.047*	*0.004*	*0.016*	*0.025*	-
Total	1.000	-	0.655	0.655	1.000	-

Note: Author's calculations using NIDS and PSLSD weighted, Standard errors are in italics.
Decomposition following Stark et al.'s approach (1986).

by 3%. As labour income is the largest contributor to overall income inequality, the drop in the Gini of this income source explains partially why overall income inequality in per capita terms has decreased.

The second largest contributor to inequality is investment income. This is despite the fact that between 1993 and 2014, the relative contribution of investment income to overall income inequality decreased from 11.7% in 1993 to 10.1% in 2008 and 9.2% in 2014. The key to this influential role of investment sources is the fact that the Gini coefficient of this source is among the highest of all the income sources with values reported close to one. The discussion of the descriptive statics above has shown that between 2008 and 2014 a small but increasing proportion of households report income from investment. The high Gini coefficient reflects the many households reporting zero income from this source. In 2014, that share was at 76.7% of households. In the previous year, it was significantly higher at 95% in 2008 and 96% of households reporting no income from investment in 1993. Both, investment income and income from labour market sources, have strongly dis-equalizing effects in all periods. This is shown by the elasticities reported in the last column of Table 2.3. Following from equation (2.3), a 1% change in income from labour markets leads to an absolute increase in the Gini of 0.024 in 1993, 0.047 in 2008 and 0.064 in 2014. While the dis-equalizing effect of investment income is stronger in 1993, at 0.029 points in the Gini, the decreasing share S_k of investment income leads to a drop in the disequalizing effect in response to a 1% change in investment income. In 2008, a marginal increase in investment income would lead to an increase of the Gini of 0.016 and in 2014, the Gini would increase by 0.015.

The equalizing forces of government grants and remittances offset to some degree the effects discussed so far. The results of the static decomposition suggest that the absolute and relative contributions of income from remittances and government grants are rather low yet (potentially) lowering the overall Gini coefficient. Government grants report relatively high Gini coefficients of 0.92 in 1993, 0.77 in 2008 and 0.76 in 2014. The large drop in the Gini of government grants is most likely due to increased efforts of the democratic governments since 1994 to address poverty and inequality inherited from the apartheid era through an extensive roll-out of government grants. The persistently high Gini may indicate that many households do not qualify for support from social grants or are ineligible for grant support due to a lack of documentation to support their claim (Leibbrandt et al., 2010). As such, there are many households that report zero income in this category. However, suggesting that income from government grants is relatively well targeted, the results show that for the post-apartheid years, the negative correlation of income from government grants leads to negative absolute and relative contributions of government grants in 2008 and 2014. This highlights the equalizing effect of grants on total income inequality even if these effects are

small. In 2008, government grants lowered the overall Gini by 0.002 points (0.3%) and in 2014 by 0.007 point (1%). The elasticity reported in the last column of Table 2.3 shows that a 1% increase in social grants had a potentially equalizing effect in 1993 already. A marginal increase in government grants would lower inequality measured by the Gini by 0.026 in 1993 compared to 0.057 in 2008 and 2014, holding all other incomes constant.

Income from remittances shows a negative relative effect of a marginal percentage change on inequality as well but is contributing positively to the overall Gini in 2014. The relative contribution of remittances to overall income inequality in 1993 was marginal at 0.5% and increased to 3% in 2008, only to decrease again to 1.6% in 2014. Over the same period, the share of remittances in total household income has remained fairly stable at between 3.1% and 3.8% of total household income. The correlation with the Gini, R_k, increased between 1993 and 2008 and decreased between 2008 and 2014. Overall, the Gini correlation is relatively low but remains positive between 0.11 and 0.59. However, the Gini coefficient within this income source is very high, fluctuating between 0.91 in 1993 and 2014 and 0.97 in 2008. This is due to the fact that many households report zero income in this income category, which drives up the Gini coefficient within the income source. The marginal change analysis shows that remittances have the potential to lower the Gini coefficient. A 1% increase in remittances would lead to a 0.026 decrease in inequality as measured by the Gini in 1993, a decrease of 0.006 in 2008 and a decrease of the Gini by 0.022 in 2014. Thus, the elasticities reveal a stronger redistributive effect of this income source than the static decomposition alone. This highlights the caveat discussed above. The elasticities measured here are given that all other household decisions regarding income remain constant which remains a questionable assumption.

The main shortcoming of the approach proposed by Stark et al. (1986) is in its static analysis. While this decomposition allows a detailed examination of the contribution of different income sources in a given year, it is limited in its evaluation of the effect of changes in one income source on changes in total inequality. For example, if the arrival of a state old age pension pushes a household up along the income distribution, the decomposition is only able to reflect the situation after the arrival of the pension. The elasticities are an attempt to correct for this weakness. However, elasticities can only simulate a small or marginal change in income from one source, holding all other sources constant. This becomes problematic as an increase in labour market income, for example, might lead to disqualification from receiving government grants. The elasticities reported on labour market income cannot account for these consequences. Therefore, even though the static decomposition clearly identifies the considerable contribution of labour market income to overall income inequality, it fails in assessing the role of changes particularly in income sources that are relatively small compared to the large share of income from the labour market. The estimated elasticities are

limited in simulating the influence of income sources that changed quite markedly over the post-apartheid period on the overall income distribution. Thus, this chapter proceeds to a more contemporary approach that uses micro-simulations to assess the effect of changes in one income source on overall inequality.

2.3.2 Micro-Simulations

In this sub-section an approach by Barros et al. (2006) and Azevedo et al. (2013) is introduced that tries to model changes in different income sources using micro-simulations. Following Azevedo et al. (2013), these micro-simulations model counterfactuals by changing one factor at a time in order to decompose the contribution of the effect of measured changes in the different income sources on the aggregate change in inequality. An additional strength of these dynamic decompositions is that a focus on changes in the denominator makes it possible to examine and separate out the impact of changes in household demographics.

Following the notation of Azevedo et al. (2013), household income per capita can be represented as the sum of incomes of all household members over the number of household members n.

$$Y_{pc} = \frac{Y_h}{n} = \frac{1}{n}\sum_{i=1}^{n} y_i, \tag{2.4}$$

where y_i is the income of individual i and Y_h is the total household income. Equation (2.4) can be rewritten assuming that only persons aged 15 years and above are able to contribute to household income. Then, in fact, household income per capita will depend on the number of adults in the household or n_A such that

$$Y_{pc} = \frac{n_A}{n}\left(\frac{1}{n_A}\sum_{i\in A}^{n} y_i\right), \tag{2.5}$$

where $\frac{n_A}{n}$ represents the share of adults in the household. The expression in parentheses is the income per adult which can be written as the sum of income from different income sources. Assume for simplicity that income per adult can be divided into two sub categories, labour income and income from non-labour sources or y_i^L and y_i^{NL} respectively (Azevedo et al., 2013). In the context of this chapter, income from non-labour sources includes income from social grants, pensions and other government sources, remittances or investment income. Regarding labour income, it is important to note that not all adults in the household will be employed, instead only the share $\frac{n_0}{n_A}$ will earn income from labour markets with n_0

being the number of employed adults. Then, equation (2.5) transforms into

$$Y_{pc} = \frac{n_A}{n}\left(\frac{1}{n_A}\sum_{i\in A}^{n} y_i^L + \frac{1}{n_A}\sum_{i\in A}^{n} y_i^{NL}\right) \tag{2.6}$$

$$= \frac{n_A}{n}\left[\frac{n_0}{n_A}\left(\frac{1}{n_0}\sum_{i\in A}^{n} y_i^L\right) + \frac{1}{n_A}\sum_{i\in A}^{n} y_i^{NL}\right]. \tag{2.7}$$

The distribution of household per capita income $F(\cdot)$ depends on the different components outlined in equation (2.7) and in turn, inequality measures depend on the cumulative density function $F(\cdot)$ and as such, can be written as a function of the components discussed above.

Azevedo et al. (2013) show that micro-simulations can be used to estimate the contribution of each component to the observed changes in the inequality measures by changing each component one at a time. With this in mind, assume that ϑ is a measure of inequality and as such a function of the cumulative density function $F(\cdot)$ and the components outlined above. Then

$$\vartheta = \Phi(F(Y_{pc}(n, \frac{n_A}{n}, \frac{n_0}{n_A}, \underbrace{\frac{1}{n_0}\sum_{i\in A}^{n} y_i^L}_{y_{P0}^L}, \underbrace{\frac{1}{n_A}\sum_{i\in A}^{n} y_i^{NL}}_{y_{PA}^{NL}}))), \tag{2.8}$$

with y_{P0}^L representing labour market incomes and y_{PA}^{NL} being the non-labour incomes including government grants, remittances and investment income. In order to estimate the contribution of each component to the observed changes in the inequality between period 1 and period 2, Azevedo et al. (2013) substitute the period 1 level of the respective income source into the counterfactual distribution of period 2. To give an illustration for the case of the study at hand, assume there is a change in the share of adults in a household between 1993 and 2008. Following Azevedo et al. (2013), the counterfactuals will be created by ranking households according to their household income per capita and then assigning the averaged values of each component in equation (2.8) to the corresponding rank in the rescaled 2008 distribution. The rescaling exercise simply rescales the number of observation in one period in such a way that they have the same range as the other period (Azevedo et al., 2012). In order to compute $\hat{\vartheta}$, the 1993 value of $\frac{n_0}{n_A}$ will be substituted into the rescaled distribution of income $f(\cdot)$ observed in 2008:

$$\hat{\vartheta} = \Phi(F(Y_{pc}(n, \frac{n_A}{n}, \frac{\hat{n_0}}{n_A}, y_{P0}^L, y_{PA}^{NL}))). \tag{2.9}$$

Then, $\vartheta - \hat{\vartheta}$ is the estimated contribution of the change in the share of adults that earn labour market income between 1993 and 2008. In the same manner in which $\frac{\hat{n_0}}{n_A}$ was substituted, each of the other components of interest can be substituted into the distribution of

income per capita in 2008 following the rescaling exercise and their contribution to changes in inequality can be estimated. For example, when decomposing the impact of changes in labour market incomes, the 2008 labour market income will be substituted with the average labour market income in 1993 according to the rescaled ranking within the distribution. The effect of labour market income on inequality is then measured by comparing the level of inequality of the simulated data with actual levels of inequality observed in 2008.

The method introduced by Barros et al. (2006) is refined by Azevedo et al. (2013) by computing what they call a "cumulative counterfactual distribution". By adding one variable at a time, the impacts of a change in each of the variables of interest and their interactions can be estimated as the difference between those cumulative counterfactuals (Azevedo et al., 2013).

One technical issue that has to be addressed in the simulation is the fact that the cumulative counterfactuals estimated differ depending on the order in which the different variables are added. In other words, the path that is used for the estimation of the cumulative effects matters. This problem is called path-dependence (Azevedo et al., 2013). In order to overcome this caveat, Azevedo et al. (2013) suggest calculating the Shapley-Shorrocks estimate of each component. The Shapley-Shorrocks estimate calculates the decomposition across all possible paths before averaging the results.[5] Since there are six variables of interest, this aggregates to 720 possible paths, i.e. the result of 6 *factorial*. The estimates of the Shapley-Shorrocks values are reported in the tables below.[6]

While this prevents the analysis from suffering from this major shortcoming, one caveat remains. The counterfactuals calculated in this manner are the result of a statistical exercise rather than actual economic equilibria, in which it is assumed that one component can be changed at a time, keeping all other factors constant. However, an increase in any income source would generally lead to an adjustment in economic behaviour. Households tend to substitute a loss in one income source with an increase in another and, vice versa, an increase in one income source may cause a decrease in efforts to obtain income from another source. The fact that changes in income sources may lead to changes in the household composition is acknowledged by this approach through accounting for these effects separately. The estimated Shapley-Shorrocks values represent the closest possible approximation of these substitution exercises. Therefore, this chapter proceeds to calculate these simulated counterfactuals using the 1993 PSLSD and NIDS 2008 and 2014 data sets.

[5]For details of this cumulative path method, see Box 2 from Azevedo et al. (2013) in Appendix 2.B.

[6]These estimations are done using the *adecomp* program in Stata: Azevedo, J.P., Sanfelice, V. and Nguyen, M.C. "Shapley Decomposition by Components of a Welfare Measure."World Bank. (2012).

The dynamic decomposition will be broken into two steps since the results of the static decomposition exercises showed that there was an increase in inequality between 1993 and 2008, whereas the static decomposition reported a decrease between 2008 and 2014.[7] The dynamic decomposition can be thought of as an accounting tool which helps identify the underlying principles of what drove the increase and later decrease of inequality over these periods.

Before this chapter can proceed and the PSLSD and NIDS data sets can be compared in this dynamic decomposition, it is necessary to undertake a ranking exercise as discussed above. Following Azevedo et al. (2013), households will be ranked by income per capita in the first period, say 1993, and then matched to the corresponding rank in the second period, say 2008. The average of each component in equation (2.8) from the first period will be assigned to the matched household in the second period. This will be done for households ranked according to their household income per capita; however, it is possible to rank order households by the different income components instead. This may offer some insights into the effects of changes in that particular component on overall inequality. For components that are highly correlated with overall inequality, this change in the ranking order will result in small differences. However, when this is not the case, it is possible to differentiate shifts in the overall income distributions from shifts within the distribution of the particular income source by re-ranking according to different income sources. The discussion of Table 2.5 will show that for example with regards to government grants, the re-ranking of the distribution can give valuable insights to the targeting of such grants.

The decomposition in Table 2.4 ranks households according to overall per capita income. It reports the estimation without taking account of changes in household composition. It is therefore implicitly attributing any demographic changes to changes in the per capita values of the different income sources. As such, the results in Table 2.4 serve as benchmarks for Table 2.5 which will provide the results of the simulations following different rankings while also accounting for changes in the household composition.

The estimation in the first part of Table 2.4 shows that between 1993 and 2008, changes in labour income contributed to the observed increase in the Gini over this period, which is similar to the results found in the static decomposition. In the observed period, changes in labour income increased the Gini by 0.043 points or 6.3% of the original Gini. Several studies have found a sharp increase in unemployment after the end of apartheid (Kingdon and Knight, 2007; Burger and Woolard, 2005; Bhorat and Oosthuizen, 2006). This was the result of dismantling previous labour market restrictions that prevented a majority of the population to participate fully in the labour market which lead to an abrupt increase in the

[7]The results of the dynamic decomposition from 1993 to 2014 are reported in Appendix 2.B.

Table 2.4: Dynamic Decompositions of Per Capita Income Sources
1993 to 2008 and 2008 to 2014

	1993 - 2008		2008 - 2014	
Effect	**Gini**	**%-Change**	**Gini**	**%-Change**
Labour Income	0.043	6.3%	-0.021	-3.0%
Government Grants	-0.034	-5.0%	-0.006	-0.9%
Remittances	0.011	1.6%	-0.011	-1.6%
Investment	-0.01	-1.5%	0.003	0.4%

Note: Author's calculations using PSLSD and NIDS weighted.
Number of paths = 24
Number of factors = 4

supply of labour. However, this was not met with an increase in demand for labour. Thus, the changes resulted in increased unemployment. These changes in the labour market are reflected in the results presented here.

Comparing the changes uncovered by the dynamic decompositions from 1993 to 2008 and 2008 to 2014 respectively, the effects are relatively larger in the first 15 years than in the more recent period under observation. Changes in labour incomes have become more equalizing. The results of the static decomposition showed that between 2008 and 2014 the overall Gini coefficient decreased from 0.69 to 0.66. The dynamic decompositions reported in Table 2.4 also show that changes in income from labour market sources were the strongest contributor to the decline in overall inequality. Changes in income from labour markets resulted in a decrease in the Gini of 0.021 or 3%. This supports the results of the static decomposition, which showed that the Gini within labour market income had fallen but the contribution to the overall Gini had increased. As such, a decrease in inequality from labour market income may have driven the overall decrease in inequality in the period between 2008 and 2014.

Between 1993 and 2008, changes in incomes from non-labour sources, on the other hand, mostly had an equalizing effect, lowering the overall Gini coefficient. The strongest equalizing effect is from government grants. Changes in government grants lowered the Gini by 0.034 points (5%). This reflects the substantial roll-out of social grants and, in particular, the Child Support Grant that is documented in Leibbrandt et al. (2010). Government grants helped decrease inequality in both periods even though its decreasing effects weakened relatively between 2008 and 2014. Changes in incomes from government grants resulted in a decrease in the Gini of 0.9% over this period.

Changes in income from investment also had an overall negative effect on the Gini of 0.01 points or 1.5%. This is most likely due to the fact that overall, more households received

this type of income while the relative share of income from investment sources fell sharply in the period observed. This reflects institutional reforms since the end of apartheid that extended access to financial institutions to previously discriminated households. However, Ardington et al. (2004) find that there are still constraints that households particularly at the bottom of the income distribution face in the South African financial system. This may explain, why between 2008 and 2014, investment income leads to a small increase in household income inequality of 0.4%.

Between 1993 and 2008, changes in income from remittances increased the Gini by 0.011 points which translates into a 1.6% change from the 1993 Gini coefficient. During the apartheid era, African workers were not allowed to permanently settle at their places of employment. This led to remittances becoming an important form of income (Posel, 2016; Posel and Casale, 2006; Maitra and Ray, 2003). However, studying data from 1989 and 1992, Jensen (2004) finds a "crowding-out" of income from private transfers when income from government grants increased. Leibbrandt et al. (2010) show supporting results particularly for the period after 2000. This may explain why over this period, changes in this form of income led to an increase in income inequality even though remittances can contribute to redistributing incomes. Such a redistributive effect can be found in the results of the dynamic decomposition using NIDS 2008 and 2014 data. Changes in incomes from remittances resulted in a decrease in the Gini of 1.6% over this period. Posel (2016) finds that between 2008 and 2014, individuals that receive social grants are more likely to receive remittances as well. This indicates the redistributive effects of remittances and explain the negative effect of changes in income from remittances over this period.

The approach developed by Azevedo et al. (2013) allows to differentiate the changes in inequality not only according to changes in different income sources but also according to changes in household demographics. Table 2.2 presented a basic description of household demographics at each of the three points in time. To recapitulate, household size decreased between 1993 and 2014 from an average of 4.4 persons in a household to only 3.2 on average. At the same time, the share of adults as well as the share of employed household members increased between 1993 and 2008, from 73.4% of the household to 77.7% in 2008. However, the share dropped to about 74% in 2014. The share of employed increased between 1993 and 2014 from 37% to 41.7%. Table 2.5 goes on to factor in these changes in household decomposition into the dynamic decompositions. The table also allows for a re-ranking of income sources.

It can be seen from the first part of Table 2.5 that between 1993 and 2008, the change in the share of adults in the household increased the Gini coefficient by 0.002 units (0.3%). The share of employed adults on the other hand decreased the Gini by rather much, 2.9%

Table 2.5: Dynamic Decompositions - 1993 to 2008 and 2008 to 2014
Including Household Composition and Different Ranking Variables

	1993 - 2008		2008 - 2014	
Variable	**Gini**	**%-Change**	**Gini**	**%-Change**
Share of Adults in HH	0.002	0.3%	0.001	0.1%
Share of Employed in HH	-0.02	-2.9%	0.012	1.7%
Labour Income				
Ranked by Total HH Income	0.068	10.0%	-0.039	-5.6%
Ranked by Labour Income	0.068	10.0%	-0.05	-7.2%
Government Grants				
Ranked by Total HH Income	-0.024	-3.5%	-0.002	-0.3%
Ranked by Government Grants	-0.027	-4.0%	-0.004	-0.6%
Remittances				
Ranked by Total HH Income	0.000	0.0%	-0.002	-0.3%
Ranked by Remittances	0.003	0.4%	-0.003	-0.4%
Investment				
Ranked by Total HH Income	-0.016	-2.3%	-0.005	-0.7%
Ranked by Investment	-0.013	-1.9%	0.002	0.3%

Note: Author's Calculations using NIDS 2008 and 2014 weighted.
Number of paths = 720
Number of factors = 6

or 0.02 points. Table 2.2 showed that both these variables had increased slightly between 1993 and 2008. Apparently, the increase in the share of employed in a household contributes to a decrease in the overall levels of inequality. Between 2008 and 2014, both, the share of adults as well as the share of employed adults in a household, had an increasing effect on inequality. The share of employed increased inequality by 1.7% of the 2008 Gini whereas changes in the share of adults in a household only increased the Gini by 0.1%. This could be since the share of adults decreased whereas the share of employed increased between 2008 and 2014 (see Table 2.1). This section moves on to analyse the effects of changes in the different income sources in order to assess the effect of separating the changes in demographic variables compared to the benchmark case of not including household demographics. Since the socio-economic circumstances explaining the different changes in income inequality were discussed in the analysis of Table 2.4, the discussion below will focus on how the effects differ when the methodological advantages of the dynamic decompositions are employed.

Accounting directly for the changes in demographic variables affects the estimated contributions of the changes in income sources markedly. Now, the changes in labour income worsen inequality even more and are driving the increase in the Gini coefficient by 10% or 0.068 points between 1993 and 2008. The result of the simulations after ranking the distribution by labour market incomes leaves this effect unchanged. This supports the finding of the static decomposition that labour income is indeed highly correlated with overall house-

hold income. It indicates that the distribution within this income source remained highly unequal between 1993 and 2008 and that labour income continues to drive inequality, especially after changes in household demographics are accounted for. Once demographic variables are accounted for, income from all sources contributed to the decline in overall inequality between 2008 and 2014. The most substantial driver in the fall of the Gini coefficient over this period is income from labour market sources. The effect is at -0.039 points of the Gini or -5.6%. This large change is even higher when ranked by labour income. This is somewhat surprising, however, Table 2.1 reported that the share of labour income dropped between 2008 and 2014 while the proportion of households receiving this type of income increased from 64.4% to 72.6%. This sharp increase in the reach of this type of income may help explain the decrease in inequality. The increased diversity in income sources and stronger impact of equalizing income sources may have further strengthened this effect. At the same time, the Gini coefficient within labour incomes decreased from 0.76 to 0.73 in 2014, implying that the increased proportion of households receiving this income also report a slightly more equal distribution of this income than in 2008. The interaction of these labour market income effects contributed to the fall in inequality between 2008 and 2014.

Changes in government grants have a strongly equalizing effect when ranked by household income and more so when ranked by grant income. The impact of changes in income from social grants increases from 3.5% to 4% depending on the ranking variable. The fact that the equalizing effect of government grants are stronger when ranked by grants implies that changes in the targeting of government grants was effective in addressing households at the bottom of the income distribution. Compared to the baseline results of Table 2.4, the effect of government grants is lower once demographic changes are separated. This means that the simple dynamic decomposition that does not account for demographic changes overestimated the effect of targeted government grants. Between 2008 and 2014, government grants have a small decreasing effect on income inequality (0.3%) and the effects on changes in the Gini are slightly larger when ranked by grant income (0.6%). It is important to note that between 1993 and 2008 government grants played an active role in decreasing inequality but inequality within labour market income was a much stronger force at increasing levels of inequality. Between 2008 and 2014, these roles have changed in the way that while government grants are still reducing inequality, it is income from labour market sources that dominates the decrease in the overall Gini coefficient. For now, it seems that government grants are still playing an equalizing role, however, this role is relatively small, particularly once the effects of demographic changes are accounted for. Given the changes in income sources and household demographics as they are reported in the second part of Table 2.5, this may indicate that government grants are approaching their marginal limits of redistribution.

Once the effect of changes in household demographics are accounted for, changes in re-

mittances had no effect when ranked by total household income between 1993 and 2008. However, changes in remittances increased inequality slightly when ranked by remittance income. The effect is small at 0.4%. Remittances probably have such a small effect on changes of the Gini coefficient because they play only a minor role in overall household income. The results of the static decomposition in Table 2.3 showed that remittances only contributed between 0.5% and 3% to total inequality and their shares in income were at about 3%. The ranking exercise shows that the inequality increasing effect is larger than when ranked by total household income. This indicates that there is some sort of reverse targeting of remittances such that households on the bottom of the income distribution receive less income from remittances and households on the top are more likely to receive transfers. This supports the "crowding-out" effect discussed above. Low-income households receive more income from government grants which in turn leads to less income from remittances for these households. If low-income households would receive more income from remittances, their effect could be equalizing. This was the finding of the static decomposition which highlighted the equalizing potential of remittances even though they were contributing positively to overall inequality. In the benchmark, remittances increased inequality by 0.011 points which implies that in the benchmark of the dynamic decomposition, inequality changes actually attributed to the change in demographic household variables were implicitly accounted for by this income source. Between 2008 and 2014, remittances have a small effect such that changes in income from this source lead to a 0.3% decrease in the overall Gini when ranked by total household income. This effect is slightly stronger when ranked by remittance income at 0.4%. This supports the potential to decrease inequality that was discussed in the static decomposition. However, the static decomposition reported that remittances are currently contributing to inequality. The dynamic decompositions show that when the effects of household composition variables are netted out, remittances decrease inequality between 2008 and 2014 at least to some degree and have the potential to do more. These findings are in line with the descriptive analysis of Posel (2016) who no longer found a "crowding-out" effect of government grants between 2008 and 2014.

Finally, changes in income from investment sources had a small but equalizing effect on changes in the Gini for the different rankings for the period between 1993 ad 2008. When ranked by total income, this effect is at 2.3%. When assessing the effect of changes in the distribution of investment income, the effect is at 1.9%. These effects are larger than the 1.5% estimated in the benchmark decomposition. This shows that when changes in demographics are not accounted for, the actual impact of changes in investment income is underestimated. The second part of Table 2.5 shows that income from investment sources has a decreasing effect on the Gini between 2008 and 2014. Ranked by total household income, changes in investment result in a -0.7% change in the Gini. However, when ranked by investment income the decomposition shows that investment income increases inequality by 0.3%. The

fact that the changes switch signs when ranked by investment income indicates that the distribution of income from this source is actually quite unequal. Furthermore, it highlights the necessity not only to account for the effects of changes in household demographics but also to assess income sources according to different ranking methods.

Overall, it would seem that increased efforts by the post-Apartheid state in addressing poverty and inequality through government grants have largely offset inequality increasing effects of labour market income after the end of apartheid. However, the different ranking methods and accounting for the effect of changes in household demographics separately have shown that government grants as a redistribution tool may have reached their limit. The different ranking methods as well as accounting for changes in household decomposition presented in Table 2.5 offer important insights into the distribution of the different income sources and their effects on income inequality. The results show that it is necessary to take advantage of the different tools of analysis to truely understand the effects of different income sources.

In Appendix 2.B, Tables 2.6 and 2.7 report the aggregate trends from a broad-period decomposition that applies the methods of Azevedo et al. (2013) to compare changes between 1993 and 2014. In this long-term comparison, very strong effects of government grants on changes in the Gini can be found (-5.6%) in the baseline decomposition. Table 2.6 which does not account for household demographics finds that changes in labour income contributed 3.1% to the overall changes in inequality measured by the Gini and that lowered income inequality by 0.3%. Investment income decreased the Gini coefficient by 0.008 points or 1.2% between 1993 and 2014. Once household composition variables are accounted for, these effects change slightly.

The results reported in Table 2.7 show that the share of adults in a household contributes to a 0.3% change in the Gini coefficient between 1993 and 2014, this is equivalent to 0.002 points. The share of employed adults in a household leads to a 1.3% decrease over the same period. However, it is important to note that the decomposition of Table 2.5 in which the effects of changes in household demographics were accounted for separately reported that the share of employed had a decreasing effect of close to 3% only between 1993 and 2008, whereas between 2008 and 2014 changes in the share of employed contributed 1.7% to an increase in the Gini.

Table 2.7 also reports that throughout the period of 1993 to 2014, changes in labour income contributed to a 3.7% decrease in the Gini when ranked by total household income. When this is ranked by labour income, however, the effect reduces to an increase in the Gini by 2.8%. The fact that the change in the Gini is larger when ranked by total household income

and smaller when ranked by labour income would indicate a slightly less unequal distribution within labour market income. Because this overall decomposition reports the net changes, it falls short in showing the inequality-increasing effect of labour income between 1993 and 2008 or the inequality-reducing effect between 2008 and 2014. These movements in inequality have important policy implications. Therefore, it is highly beneficial to carry out a step-by-step analysis rather than looking at the longer period.

Furthermore, between 1993 and 2014, changes in government grants reduced the Gini by 3.4%, 3.8% when ranked by grant income. This is in line with the inequality-reducing effects found previously and seems to be the aggregate of the effects found between 1993 and 2008 and between 2008 and 2014 respectively. The effect of remittances over the entire period netted out to -0.1% when ranked by total household income and -0.6% when ranked by remittance income. While this appears to be the aggregate of the separate effects found in Table 2.5, this overall analysis ignores the potentially inequality increasing effects of remittances between 1993 and 2008 that were offset by the inequality decreasing effect of changes in this income source between 2008 and 2014.

Finally, investment was found to have a decreasing effect of 2.9% when ranked by total household income and reports a slightly smaller effect of 2.6% when ranked by investment income. It should have become clear that while the decomposition between 1993 and 2014 generally reports the aggregated trends in inequality over the observed period, it cannot account for changes in the effects of the different income source on the increase and then decrease of overall inequality measured by the Gini coefficient uncovered by the separate decompositions.

The dynamic decomposition method is a novel approach in the South African context. Using micro-simulations, the above decompositions of changes in inequality between 1993 and 2008 as well as 2008 and 2014 respectively, show that government grants are strong drivers in the reduction of overall income inequality. The equalizing role of government grants has long been established in the literature (see Leibbrandt et al., 2012, and Leibbrandt et al., 2010), however, the results of the dynamic decomposition prove the powerful effect of correctly targeting these grants. Furthermore, the above analysis revealed that changes in labour market income have substantial effects on overall inequality measured by the Gini coefficient. Between 1993 and 2008, it was income from labour markets that drove the increase in inequality and between 2008 and 2014, income from this source contributed strongly to a decrease in the Gini. These ambivalent effects of changes in income from the labour market have previously been overlooked when using static decomposition methods.

2.4 Conclusion

This chapter has applied different income source decomposition methods in order to ascertain drivers of inequality in South Africa over the post-apartheid period. A static decomposition by income sources established the dominance of contributions in labour market income. Using data from the Project for Statistics on Living Standards and Development from 1993 as well as from the National Income Dynamics Study from 2008 and 2014, labour market income has been shown to be the largest contributor to the high levels of inequality in South Africa. Furthermore, this chapter has shown that while inequality was rising from 0.68 in 1993 to about 0.69 in 2008, it has been declining in recent years to a Gini coefficient of 0.66 by 2014. According to the decompositions performed in this chapter, households benefit significantly from income from government grants whereas labour income as well as investment income contribute actively to overall inequality.

Previous analysis of inequality in South Africa has relied heavily on the type of static decompositions by income source (see Leibbrandt et al., 2009 or Leibbrandt et al., 2012) discussed in this chapter. The analysis in this chapter has shown that the static approach is successful in identifying labour income as the main driver behind overall income inequality in South Africa. The static decomposition method following Stark et al. (1986) suggested that labour market income contributes between 84% and 90% to the overall Gini coefficients between 1993 and 2014, proceeded by large contributions of investment income to inequality. This chapter also found that between 2008 and 2014, labour market incomes remain highly correlated with overall inequality. However, the static decomposition falls short in explaining how changes in the different income sources affect changes in observed levels of inequality. This makes a significant difference especially when assessing the effect of incomes that have a smaller share in overall household income, such as government grants.

The dynamic approach following Azevedo et al. (2013) implements a series of counterfactual simulations to identify the direct effect of a change in a particular income source on the total income inequality. While applying these methods, this chapter has demonstrated that although government grants only make up between 6% and 8% of overall household income, these incomes have significantly contributed to reducing inequality. The results of different ranking exercises within the dynamic decomposition showed that changes in the targeting and extensions to the system of government grants largely offset the inequality increasing effects of labour income between 1993 and 2008. The static decomposition method was unable to differentiate between the effects of labour income and government grants in that way. The role of government grants changed hugely between 1993 and 2008. Poverty-alleviating policies that resulted in an expansion of government grants limited the increase in inequality driven by changes in labour market incomes over this period. Even between

2008 and 2014, government grants played a significant role in reducing inequality and different ranking methods show that the targeting of social policies continued to successfully address households at the bottom of the income distribution. However, their effect shrank significantly, thus indicating that social spending by the government might be reaching its limits with regards to reducing inequality given the household income distribution as it stood in 2014. Additionally to the successful implementation of government grants, the dynamic decomposition has shown that changes in inequality within labour market income contributed strongly to the decrease in inequality between 2008 and 2014 once changes in household composition were separated. The nuances of the effects of the different income sources are largely overlooked when relying on the static decomposition approach.

Furthermore, the dynamic approach has shown that investment income played an inequality reducing role, particularly between 1993 and 2008. The static decomposition failed to detect these lowering effects of investment income on inequality. Additionally, the results of the static decomposition suggested that while remittances have the potential to lower inequality, as shown by negative elasticities, they are contributing positively to inequality. However, the dynamic approach using micro-simulations paints a different picture. Remittances helped decrease inequality between 2008 and 2014.

The more refined analysis using micro-simulations made it possible to account for changes in household demographics as well as changes in income sources. Changes in the share of employed adults had a dis-equalizing effect between 1993 and 2008 but an equalizing effect between 2008 and 2014, whereas changes in the number of adults in a household led to an increase in the Gini in both periods. The effects of both these variables are missing in the decomposition method by Stark et al. (1986). Particularly between 2008 and 2014, it is the household composition variables that drive inequality upwards, whereas all income sources report an equalizing effect. This highlights the importance of separating income changes from demographic changes as both have changed substantially over the post-apartheid period. There is improved information by using the method of Azevedo et al. (2013).

All in all, this chapter analysed the development of inequality over the first twenty years after apartheid using household survey data. Thus, showing that inequality has started to fall in South Africa. The novel decomposition methods using micro-simulations applied in this chapter have allowed for the identification of the role of changes in different income sources on this reduction in ways not observed in any other research to-date. Even more so, this chapter has been able to differentiate between the effects of changes in demographics as well as changes in the income sources. However, the progress in analytical methods highlights the need for further policy improvement in order to deal with continuously high levels of inequality.

The dynamic analysis shows a more significant impact on inequality of social grants than the static decomposition. This makes sense given the massive increase in social grants over the post-apartheid period. Nonetheless, extensive dependency on government grants will not be sustainable in tackling prevailing levels of inequality much further. The role of government grants in counteracting increasing forces on inequality is undeniable. However, there is little scope left for extending their coverage and government funds are strained which makes significant increases unlikely. However, there is potential to improve the redistribution efforts by adjusting the administration and design of government grants to be more efficient.

The analysis of this chapter has identified the changes in inequality within labour market income as a pivotal driver of overall changes in inequality. The high levels of inequality within the distribution of labour market income are caused both by high levels of unemployment, meaning individuals receive no income from earnings, as well as highly unequal earnings of those who are employed (World Bank, 2018). Therefore, the key to lowering income inequality in the long-term remains in labour market policies that are inclusive of those at the bottom of the income distribution through employment and earnings. Potential policies for this might include promoting the development of small enterprises and of the informal sector. Additionally, government could make use of employment or wage subsidies to reduce employment costs. This has the potential to increase labour demand and induce growth which can have beneficial effects on the reduction of inequality through several channels (Bourguignon, 2004; Donaldson, 2014).

Contributing to the high levels of overall income inequality in South Africa are high levels of inequality at the top of the income distribution. Some of the proposed policy solutions are aimed at the bottom of the distribution and would therefore not address this problem. It is widely acknowledged that survey data tend to miss or under-report the top end of the income distribution. Therefore, Chapter 3 utilizes income tax records from the South African Revenue Service. This allows to combine these data with household surveys to get a more comprehensive picture of the income distribution. Chapter 3 will look more closely into the top of the income distribution but also analyse zero earners, those who report no taxable income from earnings or otherwise. Additionally, the dominant role of labour market income during both, the increase of inequality between 1993 and 2008 as well as the decrease in recent years, has been established in this chapter. Therefore, the effect of changes in labour market income will be analysed with more detail in Chapter 4.

2.A Extended Methodology

Assume that households derive income y_K from different income components K. Total household income is then given by the sum of income from all income components K. This can be formalized as $y_0 = \sum_{k=1}^{K} y_k$. Following Stark et al. (1986), the Gini coefficient for total household income y_0 is then given by

$$G_0 = \frac{2Cov\left[y_0, F(y_0)\right]}{\mu_0}, \tag{2.10}$$

where G_0 represents the Gini coefficient of all household incomes and μ_0 denotes the mean of household incomes. $F(y_0)$ in equation (2.10) denotes the cumulative distribution function of overall household income y_0. Given the property that $y_0 = \sum_{k=1}^{K} y_k$, equation (2.10) can be rewritten as

$$G_0 = \frac{2\sum_{k=1}^{K} Cov\left[y_k, F(y_0)\right]}{\mu_0}, \tag{2.11}$$

where $Cov\left[y_k, F(y_0)\right]$ is the covariance between income source k and the cumulative distribution of income, $F(y_0)$. It is necessary to utilize the properties of the covariance in this way in order to arrive at the next step of the decomposition of the overall Gini G_0. As such, let $S_k = \frac{y_k}{y_0}$ denote the share of income from component k in total household income y_0 and let G_k denote the corresponding Gini coefficient measuring the level of inequality within income component k. Using S_k and G_k, equation (2.11) can be rewritten following the steps outlined below.

$$G_0 = \underbrace{\frac{2\sum_{k=1}^{K} Cov\left[y_k, F(y_0)\right]}{\mu_0 \cdot y_0}}_{\text{Relative Gini}} \cdot \frac{Cov(y_k, F_k)}{Cov(y_k, F_k)} \cdot \frac{y_k}{y_k}.$$

Equation (2.11) has been multiplied with $\frac{Cov(y_k,F_k)}{Cov(y_k,F_k)}$ and $\frac{y_k}{y_k}$ to obtain equation (2.1) discussed above. Each of the components of equation (2.1) can be rewritten as follows.

$$\begin{aligned} G_0 &= \sum_{k=1}^{K} R_k \cdot G_k \cdot S_k \\ &= \sum_{k=1}^{K} \frac{Cov\left[y_k, F(y_0)\right]}{Cov\left[y_k, F(y_k)\right]} \cdot \frac{2Cov\left[y_k, F(y_k)\right]}{\mu_k} \cdot \frac{y_k}{y_0}, \end{aligned} \tag{2.12}$$

To repeat, S_k and G_k are the share and the Gini coefficient of income component k respectively. R_k represents the so-called Gini correlation of component k with total household income. This correlation is given by

$$R_k = \frac{Cov\left[y_k, F(y_0)\right]}{Cov\left[y_k, F(y_k)\right]}.$$

Note that in order to dervive equation (2.1) one needs to take into account that $\mu_0 \cdot y_k = \mu_k y_0$ since $\sum \frac{\sum y_k}{N} \cdot y_k = \frac{\sum y_k}{N} \cdot \sum y_k$ with $\mu_k = \frac{\sum y_k}{N}$ and $\sum y_k = y_0$.

2.B Additional Figures and Tables

Figure 2.1: Box 2 in Azevedo et al. (2013)

Box 2. Proposed Methodology along One possible path

1.	$\vartheta_0 = \Phi\left(F\left(Y_{pc}\left(\frac{n_A}{n}, \frac{n_o}{n_A}, y_{PO}^{L}, y_{PA}^{NL}\right)\right)\right)$	Initial inequality rate
2.	$\widehat{\vartheta_1} = \Phi\left(F\left(Y_{pc}\left(\frac{\widehat{n_A}}{n}, \frac{n_o}{n_A}, y_{PO}^{L}, y_{PA}^{NL}\right)\right)\right)$	Contribution of share of household adults is $\widehat{\vartheta_1} - \vartheta_0$
3.	$\hat{\vartheta}_2 = \Phi\left(F\left(Y_{pc}\left(\frac{\widehat{n_A}}{n}, \frac{\widehat{n_o}}{n_A}, y_{PO}^{L}, y_{PA}^{NL}\right)\right)\right)$	Contribution of the share of occupied adults is $\widehat{\vartheta_2} - \widehat{\vartheta_1}$
4.	$\hat{\vartheta}_3 = \Phi\left(F\left(Y_{pc}\left(\frac{\widehat{n_A}}{n}, \frac{\widehat{n_o}}{n_A}, y_{PO}^{L}, \widehat{y_{PA}^{Pens}}, y_{PA}^{Trans}, y_{PA}^{Cap}, y_{PA}^{Oth\ NL}\right)\right)\right)$	Contribution of pensions is $\widehat{\vartheta_3} - \widehat{\vartheta_2}$
5.	$\hat{\vartheta}_4 = \Phi\left(F\left(Y_{pc}\left(\frac{\widehat{n_A}}{n}, \frac{\widehat{n_o}}{n_A}, y_{PO}^{L}, \widehat{y_{PA}^{Pens}}, \widehat{y_{PA}^{Trans}}, y_{PA}^{Cap}, y_{PA}^{Oth\ NL}\right)\right)\right)$	Contribution of transfers is $\widehat{\vartheta_4} - \widehat{\vartheta_3}$
6.	$\hat{\vartheta}_5 = \Phi\left(F\left(Y_{pc}\left(\frac{\widehat{n_A}}{n}, \frac{\widehat{n_o}}{n_A}, y_{PO}^{L}, \widehat{y_{PA}^{Pens}}, \widehat{y_{PA}^{Trans}}, \widehat{y_{PA}^{Cap}}, y_{PA}^{Oth\ NL}\right)\right)\right)$	Contribution of capital income is $\widehat{\vartheta_5} - \widehat{\vartheta_4}$
7.	$\hat{\vartheta}_6 = \Phi\left(F\left(Y_{pc}\left(\frac{\widehat{n_A}}{n}, \frac{\widehat{n_o}}{n_A}, y_{PO}^{L}, \widehat{y_{PA}^{Pens}}, \widehat{y_{PA}^{Trans}}, \widehat{y_{PA}^{Cap}}, \widehat{y_{PA}^{Oth\ NL}}\right)\right)\right)$	Contribution of other non-labor income is $\widehat{\vartheta_6} - \widehat{\vartheta_5}$
8.	$\vartheta_F = \Phi\left(F\left(Y_{pc}\left(\frac{n_A}{n}, \frac{n_o}{n_A}, y_{PO}^{L}, y_{PA}^{NL}\right)\right)\right)$	Final inequality rate. Contribution of labor income is $\widehat{\vartheta_F} - \widehat{\vartheta_3}$

Table 2.6: Dynamic Decompositions of Per Capita Income Sources 1993 to 2014

Effect	**Gini**	**%-Change**
Labour Income	0,021	3,1%
Government Grants	-0,038	-5,6%
Remittances	-0,002	-0,3%
Investment	-0,008	-1,2%

Note: Author's Calculations using PSLSD and NIDS weighted.
Number of paths = 24
Number of factors = 4

Table 2.7: Dynamic Decompositions - 1993 to 2014
Including Household Composition and Different Ranking Variables

Variable	**Gini**	**%-Change**
Share of Adults in HH	0,002	0,3%
Share of Employed in HH	-0,009	-1,3%
Labour Income		
Ranked by Total HH Income	0,025	3,7%
Ranked by Labour Income	0,019	2,8%
Government Grants		
Ranked by Total HH Income	-0,023	-3,4%
Ranked by Government Grants	-0,026	-3,8%
Remittances		
Ranked by Total HH Income	-0,001	-0,1%
Ranked by Remittances	-0,004	-0,6%
Investment		
Ranked by Total HH Income	-0,02	-2,9%
Ranked by Investment	-0,018	-2,6%

Note: Author's Calculations using NIDS and PSLSD weighted.
Number of paths = 720
Number of factors = 6

Chapter 3

The Effect of Top Incomes on Inequality

3.1 Introduction

The previous chapter has established the high levels of household income inequality prevailing in South Africa. Additionally, the dynamic decomposition has shown that previously the effect of investment income, an income source mainly available to the top tail of the distribution, has been understated. Globally, the effect of the top tail on overall income inequality is gaining attention in the recent literature. How this part of the income distribution is accounted for matters. It has widely been acknowledged that household surveys tend to understate top incomes due to a higher non-response rate among high-income households or under-reporting of incomes earned (see Atkinson et al., 2011, van der Weide et al., 2016). In this case, using tax income data has become a commonly used remedy to estimate incomes at the top end. Arguably, tax administration data is more precise on the top tail of the income distribution; however, it offers less detail on the bottom end of the distribution, particularly for individuals earning below the tax filing threshold (Morelli et al., 2014). Therefore, empirical evidence (Atkinson, 2007; Alvaredo and Londoño, 2013; Burkhauser et al., 2012; Diaz-Bazan, 2015) has shown that tax data estimates on the top tail can be combined with estimates on the bottom segment of the population obtained from household survey data in order to estimate the entire income distribution. However, this chapter will show that existing literature that uses both, tax records and household survey data, to estimate the income distribution from top to bottom has not identified a consistent point at which the two types of data sets should be combined.

In the literature on income distributions, the analysis of the development of top incomes using tax data has gained substantial traction across countries. Thus, the collection of international research on top incomes includes numerous countries. Dell (2005), for example, uses data on income tax returns to generate a homogeneous series on top shares for Switzerland and Germany spanning much of the 20th century. Atkinson (2007) uses records from the super-tax/surtax data and the Survey of Personal Incomes of the United Kingdom to analyse the distribution top incomes dating back to the beginning of the 20th century as well. Similarly, Alvaredo and Atkinson (2010) analyse top incomes in South Africa using tabulated data published by the income tax authorities between 1903 and 2007. The advantage of studies that use data on tax returns to examine the development of the income distribution is that tax return data are available for periods that precede any household survey.

In order to understand the development of income inequality, these studies of top incomes add valuable information to research that otherwise relies much on household survey data. However, the analysis of top incomes on their own remains detached to household surveys which are more commonly used. Therefore, recent studies have developed different meth-

ods to combine information from top incomes with household survey data. When external information such as tax records is combined with information from household surveys for the analysis of the entire income distribution, there are different techniques available to the researcher. As such, the analysis may combine household survey data for the bottom of the distribution with (more reliable) tax administration data on the top tail. Garbinti, Goupille-Lebret and Piketty (2016), for example, study the evolution of top incomes from France by statistically matching tax and survey data. Additionally, they use national accounts data to decompose growth by income groups and match macro-economic with micro-economic analysis. Differently from this matching approach, Diaz-Bazan (2015) studies tax income combined with household survey data in Costa Rica by decomposing the measurements of income inequality in such a way that inequality indices from household surveys can be combined with indices estimated from tax records. Finally, Lakner and Milanovic (2016) and Jenkins (2017) use different semi-parametric approaches to combine measurements of inequality.

The different approaches to the top-earner problematic show that there are many ways to account for them when information from tax administration data is available. Lustig (forthcoming) offers a detailed review of alternative methods that correct for high earners in the absence of external data sources. These methods include the replacing of household survey data at the top end through semi-parametric estimates or imputations as well as the re-weighting (post-stratification) of incomes at the top as well as on the bottom of the income distribution. Lustig (forthcoming) highlights the fact that if the true distribution and the sample have the same support, any results obtained by replacing top earners can be replicated using the re-weighting method.

Existing literature on South Africa, in particular, uses both, tax administration data and household survey data, to assess the distribution of income. However, so far, these two sources of information have been studied separately in the South African context but not together. An exception is Orthofer (2016) who combines South African household survey data with tax data to investigate wealth inequality. While income inequality features in Orthofer's analysis, the focus remains on estimating wealth across data sets. Therefore, this chapter will extend the existing analysis by using a novel method to combine information from a unique data set on personal income tax with data from a South African household survey in order to study income inequality across the entire distribution. This chapter will show that the methods chosen are the best way to assess overall income inequality as close to the actual level of inequality as possible in the context of the large share of low-income earners and steep levels of inequality observed in South Africa.

The first set of data for the analysis of this chapter stems from the National Income Dy-

namics Study (NIDS), a nationally representative household survey. The analysis uses two waves of this study, Wave 2 from 2010/2011 and Wave 4 from 2014/2015. The attrition rates reported for the different income deciles in the NIDS panel show that the study is no exception to the typical pattern of households at the top end showing increased non-response (see Finn et al., 2012 for a discussion on Wave 2, Table 3.8 in Appendix 3.A for Wave 4). For this reason, this chapter uses personal income tax (PIT) data provided by the South African Revenue Service (SARS) to better estimate inequality for high-income earners. Individuals who earn above a certain income threshold are required to file taxes. Therefore, the information on incomes above this threshold is likely to be more accurate in the tax income data than the data from the household survey. The data sets provided on personal income tax comprise a 20% sample for the 2011 and the 2014 tax year each. However, having access to two data sets on individual income does not guarantee that the data sets are comparable. Hence, this chapter reviews the information provided by the two data sources in great detail before applying the methodology introduced by Diaz-Bazan (2015) that combines the information from the two data sources to measure inequality across the entire distribution.

The estimated inequality indices show a decline in income inequality between 2011 and 2014. This decrease can be found across data sources and different methods of combining the two types of data. Whenever tax administration data and household survey data are combined, this decrease is even stronger. This shows that relying on household survey data alone, which many studies have done, to assess the income distribution distorts the analysis. Additionally, the discussion shows that for the broader part of the income distribution, household survey data reports relatively accurately on individual taxable income compared to tax administration data. However, some weaknesses are identified at the extreme ends of the distribution which makes the use of a combined analysis including tax administration data the preferred choice.

Overall, this chapter will focus on measuring income inequality. The remainder of this chapter is structured as follows. Section 3.2 provides an overview of the South African tax system before Section 3.3 introduces the two types of data used in this chapter. The distribution of taxable income in the different data sets is discussed in great detail in Section 3.4. The core of this chapter builds the discussion of current levels of income inequality and novel methods on how to improve the assessment of inequality across available data in Section 3.5. Finally, Section 3.6 summarizes key findings and concludes.

3.2 Background

3.2.1 The Tax System in South Africa

The South African tax system is broad-based and generates a relatively high level of fiscal revenue by middle-income country standards (Inchauste et al., 2015). According to SARS (2017), tax collections in 2016/2017 amounted to 25.7% of GDP, with a 60%:40% split between direct and indirect taxes. Of the direct taxes, roughly two-thirds comes from personal income tax. Personal income tax (PIT) is a tax levied on the taxable income (gross income less exemptions and allowable deductions) of a person. Capital gains also form part of taxable income. Individuals generally receive most of their income as salary/wages, pension/annuity payments and investment income (interest and dividends). Some individuals, such as sole proprietors and partners, may also have business income which is taxable as personal income (SARS, 2014).

The South African system of personal income tax is straightforward. Tax filing is done by the individual and the system does not distinguish between married and unmarried persons or provide deductions for children. There is, however, a small additional tax rebate for persons over the age of 65 (SARS, 2014). The employer must register all formal sector employees for PIT and is responsible for calculating and withholding the PIT payable by the employee. A certain level of local interest income is tax-exempt in an effort to promote saving (SARS, 2014). Limited deductions are permitted for travel expenses and contributions to pension funds and medical aid (health insurance) schemes.

Over the past two decades, there have been considerable efforts on the part of the tax authorities to broaden the tax base and to adapt the tax system to conform to international tax laws (Woolard et al., 2005). Fundamental changes included changing from a source-based to residence-based system in 2001 and the introduction of capital gains taxation to extend the tax base and enhance the equity of the tax system. There is general consensus that the reforms to PIT made the system simpler and more equitable. The modernization of the tax administration system and reporting requirements imposed on financial institutions has resulted in extremely high levels of compliance. This makes the PIT data a particularly robust source of information about individual income.

3.3 Data Sources

The analysis of income inequality in this chapter is based on two primary data sources. The first set of data stems from two waves of the National Income Dynamics Study. Secondly, the South African Revenue Service (SARS) provided a 20% sample of anonymised records

on personal income tax records for the 2011 and 2014 tax years.

The analysis in this chapter uses Wave 2 of the National Income Dynamics Study (NIDS) from 2010/2011 as well as Wave 4 from 2014/2015. Even though there are additional waves available for the panel study, the analysis in this chapter is limited to these two because tax administration data was only made available for 2011 and 2014. The cross-sectional data sets used in this chapter contain 34 000 individuals for 2010/2011 and about 42 000 individuals for 2014/2015. The broad scope of this study is one of the main advantages of using the NIDS survey data. Furthermore, NIDS contains detailed information on incomes from primary and secondary employment, self-employment and a list of bonuses that are used to assess labour income. Information on rent and interests earned are used to estimate investment incomes comparable to the information in the tax data. It is important to note that there is an exemption for local interest earned of up to R23 800 for individuals below 65 years and up to R34 500 for individuals 65 years and older. Due to the nature of the questionnaire, it is not possible to distinguish between local and international interest earned in NIDS. This may inflate taxable income in NIDS slightly, however, as only a small fraction of individuals report investment income,[1] the effect should be negligible. Capital incomes in NIDS can be estimated using the available information on retrenchment payments, loans, sales of household assets and other income of a capital nature. With these components, it is possible to construct an income variable in NIDS that is comparable to the information available in the tax data. A detailed comparison of the data provided by SARS and the information available in the NIDS data is included in Table 3.10 in Appendix 3.A. In order to further ensure that income variables in the two data sources are corresponding, all values of the 2014/2015 NIDS survey as well as the 2014 PIT data have been deflated to prices that reflect the 2011 tax year.

The South African tax year spans from March of the previous year to February of the current year. Therefore, PIT data on the 2011 tax year spans from March 2010 to February 2011. The data on personal income tax (PIT) provided by SARS contains almost 1.2 million records for the 2011 tax year and about 1 million records for the 2014 tax year which represent a simple random sample of 20% of all individuals who filed their tax returns in the respective years. The PIT data contains detailed information on investment incomes including capital gains, business profits and losses. Furthermore, it provides data on labour income including income taxed as "pay as you earn"(PAYE) and fringe benefits as well as lump sum incomes linked to earnings. Additionally, the PIT data includes any possible exemptions and deductions as well as some demographic information such as age, gender and office of registration. However, demographic information in the PIT data is scarce and

[1]The data shows that 0.5% and 1.4% of individuals report investment income in 2011 and 2014 respectively.

the information on the office of registration not comparable to provincial information in the survey data. The PIT data only spans eight out of the nine South African provinces and rather than reflecting where an individual currently lives, it reflects where an individual first registered with SARS irrespective of any moves thereafter. Additionally, the gender variable in the PIT data has a third category which will be discussed in more detail later on.

The information on incomes in the PIT data is summed up in a variable on taxable income which forms the base for the taxpayer assessment by SARS. It is important to note that the provided sample includes individuals who filed tax reports despite earning incomes below the mandatory tax filing threshold. Tax filing is mandatory for individuals who earn above a certain threshold. For the 2011 tax year, that threshold was R120 000 annual income. In 2014, however, the threshold had increased to R250 000 annual income. When this threshold is corrected for inflation, R250 000 is equivalent to R199 397 in 2011 prices.[2]

The following section will provide a more detailed analysis of the distribution of the taxable income variable in both, the household survey data and the tax administration data.

3.4 Distributional Analysis

3.4.1 Overview of the Household Survey Data

The previous section discussed the components of the variable of taxable income created in the household survey. In this section, the distribution of taxable income will be reviewed in more detail.

Table 3.1 offers a brief overview of the distribution of taxable income in NIDS for adults 18 years and older. Distributional statistics such as the first, 25th, 50th, 75th, 90th, 95th and 99th percentile are reported for both years. Additionally, the table shows the first non-zero percentiles for both years. In 2011, more than half the population reported zero taxable income. Only at the 53.8th percentile, positive taxable income can be observed. In 2014, less than half the population earned zero taxable income as the first positive incomes are observed at the 46.3rd percentile. Furthermore, Table 3.1 highlights the percentiles that are closest to the different filing thresholds of the two years. In 2011, the tax filing threshold was at R120 000 which is equivalent to the 93.6th percentile in the 2011 household survey data. The filing threshold in 2014 was at R250 000, accounting for inflation, that is equivalent to a value of R199 397 which lies between the 97.4th and the 97.5th percentile of the 2014 NIDS distribution. If the 2014 threshold were to be applied to the 2011 NIDS distribution, the

[2]For simplification, this chapter will continue to refer to it as the R250 000 threshold although all prices have been deflated to 2011 levels for the remainder of this chapter.

inflation-adjusted threshold would lie between the 97.6th and the 97.7th percentile. In the existing literature on combining tax data with household survey data, different percentiles are used as connection points and this chapter will discuss later which one is best in the context of the South African income distribution.

Table 3.1: Taxable Income - Percentiles in the NIDS Data

Percentile	**2011**	**2014**
1%	0	0
25%	0	0
46.3%		453
50%	0	7 654
53.8%	300	
75%	23 700	29 457
90%	82 548	84 449
93.6%	120 000	
95%	148 200	140 030
97.4%		196 464
97.5%		200 383
97.6%	197 200	
97.7%	202 500	
99%	324 000	413 154
Mean	30 220	36 377

Source: Author's calculations using NIDS (weighted)

The distribution of taxable income shown in Table 3.1 highlights one of the major problems when it comes to inequality in South Africa. In 2011, more than 50% of individuals had zero taxable income. There is an apparent upward shift in the distribution in 2014 which can be seen as only about 46% report zero taxable income. However, it is imperative to note that the individuals who report no taxable income may receive income from other sources. Taxable income is comprised of labour income, business profits, capital income and other sources liable for taxation. The large share of individuals that earn no taxable income highlight the strong dependence on other income sources for individuals at the bottom of the income distribution. Non-taxable income sources include government grants which specifically target poor individuals. The targeting as well as the positive effects of government grants on income inequality have been discussed in Chapter 2. Furthermore, Schiel et al. (2016) discuss the strong dependency on government grants in great detail. Other non-taxable income sources include intra-household transfers where household members profit from another household member's income. This type of resource sharing cannot be accounted for when analysing taxable income which is only available at the level of the individual. Finally, inter-household transfers in the form of remittances may be another source of income that cannot be taxed and therefore will be overlooked in this analysis. Al-

though their effect was small, Chapter 2 showed that remittances lowered income inequality between 2008 and 2014. Therefore, not being able to account for these income sources in taxable income will result in higher measurements of inequality.

The aforementioned upward shift in the 2014 income distribution can also be seen by the development of mean incomes reported in the bottom row of Table 3.1. Mean incomes increased from R30 220 in 2011 to R36 377 in 2014 even though all other percentiles in 2014 are less than their 2011 equivalents, except for the 99th percentile. This is another indicator of the high levels of income inequality prevalent in South Africa. In 2011, the mean taxable income for individuals was at R30 220 whereas the median (50th percentile) was at zero. This is a strong indicator that the top of the income distribution is moving away from the median, implying widening levels of inequality. At a median income level of R7 654 and a mean of R36 377 in 2014, these high levels of inequality do not seem to have subsided. The mean and median points of a distribution are reliable indicators of inequality[3] but the substantial differences between the top of the income distribution and the bottom warrant further investigation of these income groups. The following section will assess the group of zero earners in the NIDS data in more detail before Section 3.4.2 will review higher income earners in the tax data more closely.

Zero Earners in the NIDS Data

The discussion of Table 3.1 has shown that in both, 2011 and 2014, there is a large number of individuals with zero incomes. Table 3.2 takes a closer look at the characteristics of these zero earners. In both years, it can be observed that a large fraction of the zero earners are in the age group of 18 to 24 year-olds. This is hardly surprising for two reasons. Firstly, South Africa has a particularly high unemployment rate among youths. Secondly, this age group is most likely to still be studying or training for a job and therefore more likely to have little to no income. In the age group of 25 to 29 year-olds, the fraction of zero earners is only slightly larger than in the overall income distribution with 15.9% relative to 14.6% overall in 2011 and 16.6% relative to 16% overall in 2014. Individuals of this age group may still face a rather large unemployment rate among young people, however, as more have finished their studies, a larger fraction of this age group is starting to earn taxable income. For all age groups above 30, the fraction of zero earners is less relative to the overall distribution. This is also true for the age group of 70 years and older. This age group is less likely to receive income from employment sources so the fact that the fraction of zero earners is much smaller in this age group may indicate a significant amount of wealth income. Wealth may have been accumulated over a lifetime such that income in form of interest can be received at an older age.

[3]See Wittenberg, 2016 for a discussion on mean and median in the context of wage inequality in South Africa.

Table 3.2: Zero Earners in the NIDS Data (Percentages)

	2011		2014	
	Zero Earners	**Overall**	**Zero Earners**	**Overall**
Age Group				
18-24	36.5	24.7	36.6	22.9
25-29	15.9	14.6	16.6	16.0
30-39	21.2	24.6	20.2	24.7
40-49	14.7	18.2	13.5	18.2
50-59	10.3	13.0	12.2	13.0
70 and older	1.5	4.9	0.8	5.3
Population Group				
African	82.8	76.4	82.8	77.6
Coloured	7.9	9.4	8.7	9.3
Asian/Indian	2.5	2.9	2.4	2.9
White	6.9	11.3	6.2	10.2
Gender				
Male	43.0	47.6	43.1	48.1
Female	57.0	52.4	56.9	51.9
N_weighted	17 007 386	31 660 280	16 064 091	35 149 878

Source: Author's calculations using NIDS (weighted)

A large fraction of individuals that report no taxable income are African with 82.8% in 2011 relative to 76.4% in the overall distribution and 82.8% relative to 77.6% in 2014. In both years, white South Africans are significantly less likely to be in the group of zero income earners. In 2011, only half as many whites can be found in the group of zero income earners relative to the overall distribution with about 7% of the individuals in the zero income earners being white relative to 11.3% overall. This difference was less strong in 2014 but with 6.2% of whites in the zero income earners relative to 10.2% overall, this population group is still under-represented in the zero income earners. Similarly, coloureds and Asians or Indians are less likely to be part of the group of zero earners in both years albeit the differences for these population groups being less than for white people.

Finally, women in South Africa are more likely to be part of the zero income group than men. This has changed little over the years under observation. In 2011, 52.4% of the individuals with data on taxable income in NIDS were women. Among the zero earners, 57% were female. In 2014, there were slightly fewer women in the NIDS sample at about 52% but about 57% of women are observed among the zero income earners. In summary, young people, black South Africans and women are more likely to report zero taxable income according to the NIDS data. In later discussions of the levels of inequality in South Africa it will be shown that this large group of zero earners has a significant effect on inequality measured. Another group that leads to severe levels of inequality in South Africa are earners at the top

end of the distribution. In order to assess this part of the population, tax administration data will be used, and has been made available for the purpose of this study by the South African tax authority. This data is discussed in more detail in the following section.

3.4.2 Overview of Data on Taxpayers

Table 3.3: Income Brackets in the PIT Data

	2011		2014	
Income Group	**Mean Taxable Income**	**Percentage of Taxpayers**	**Mean Taxable Income**	**Percentage of Taxpayers**
0	0	4.53%	0	2.34%
1 - 20 000	9 630	4.97%	10 083	3.77%
20 000 - 30 000	25 130	2.58%	25 088	2.03%
30 000 - 40 000	35 146	2.94%	35 136	2.27%
40 000 - 50 000	45 211	3.34%	45 273	2.81%
50 000 - 60 000	55 419	4.59%	55 024	3.81%
60 000 - 70 000	65 109	4.82%	65 048	4.02%
70 000 - 80 000	74 925	5.30%	75 083	4.36%
80 000 - 90 000	85 000	4.63%	84 971	4.39%
90 000 - 100 000	94 967	4.04%	94 982	4.24%
100 000 - 110 000	104 889	3.84%	104 946	4.00%
110 000 - 120 000	115 132	3.67%	114 987	3.93%
120 000 - 130 000	124 965	3.66%	124 929	3.59%
130 000 - 140 000	134 934	3.32%	135 108	3.61%
140 000 - 150 000	145 012	3.09%	144 824	3.68%
150 000 - 200 000	173 340	13.73%	174 203	14.63%
200 000 - 300 000	241 192	13.12%	241 559	16.22%
300 000 - 400 000	344 656	5.63%	344 453	6.70%
400 000 - 500 000	444 726	2.91%	444 974	3.47%
500 000 - 750 000	600 010	3.11%	598 698	3.64%
750 000 - 1 000 000	855 902	1.01%	853 815	1.16%
1 000 000 - 2 000 000	1 322 742	0.90%	1 325 424	1.02%
2 000 000 - 5 000 000	2 860 366	0.24%	2 873 280	0.27%
above 5 000 000	7 704 813	0.03%	7 571 683	0.03%
N_weighted	5 783 360	100%	5 085 060	100%

Source: Author's calculations based on PIT Data (2011 and 2014)

The PIT data contains a taxable income variable, which in the following sections is shown in raw (non-manipulated) form. The data provided is used to create a proxy for mean taxable income by income bracket. This proxy is created based on the continuous distribution of taxable income in the PIT data. Table 3.3 lists 24 income brackets including zero incomes in the PIT data and the percentage of total taxpayers in each bracket. Even though the data on personal income tax provided is continuous, these results are reported in brack-

ets to relate this discussion to the annual publication of the Tax Statistics by SARS (2015).[4]

When looking at the two years for which SARS provided data on taxpayers, there is a definite shift toward the upper end of the distribution between 2011 and 2014. In the PIT data, there are relatively more taxpayers in each income bracket above a yearly income of R130 000 in 2011 prices. At the same time, the number of tax filers reporting zero income reduced to half of its 2011 level, from 4.53% to 2.34% in 2014. It is important to view any statistics below the filing threshold with a certain level of caution as only a few individuals with lower incomes will have filed tax reports. Many income earners will not be captured due to voluntary self-reporting of incomes at this level. Of those that have filed taxes, many will have done so in order to gain refunds. As mentioned above, filing for taxes is compulsory only once an individual earns above the filing threshold in the specific tax year. Recalling that the filing thresholds were at R120 000 for 2011 and at R250 000 in 2014,[5] this chapter will show later on that only above these filing thresholds is the PIT data more reliable.

There is a significant jump in mean taxable incomes that can be observed for the highest income bracket including individuals that earn R5 million a year and more. SARS provided additional statistics on high earners with annual taxable incomes of R10 million. These individuals are included in this bracket and drive up the mean. The following section will take a closer look at these high-income earners after Table 3.4 briefly reviews the demographics provided on all taxpayers by SARS.

Table 3.2 in Section 3.4.1 discussed the demographic characteristics of the individuals at the very bottom of the income distribution in the NIDS data in great detail. Even though household survey data provides a lot more information on individual characteristics than tax administration data, Table 3.4 reports a brief overview of the demographic data available in the PIT data set. The mean age in the overall distribution of registered taxpayers is 47.29 years in 2011 and slightly younger at just under 45 years in 2014. In the NIDS data, there were slightly more women than men. This is slightly reverse in the overall tax data. In 2011, almost 57% of tax filers are male, compared to 55% in 2014. The share of female tax filers increases from about 42% to 43.54%. Most notably, the tax administration data has a "missing" category in the gender variable. This stems from the fact that South African taxpayers register using their national ID. This identification document contains an indicator for the person's sex. Foreign nationals that use a passport to register with SARS will therefore not have any information on their gender. The roughly 1% of individuals with missing gender information are therefore most likely foreign nationals. However, they make

[4]The 2015 Tax Statistics are cited here as it is the last year that reports for both, the 2011 and 2014 tax year.

[5]The inflation-adjusted threshold in 2014 is at R199 397.

up a rather small fraction of the overall registered taxpayers.

Table 3.4: Demographics in the PIT Data

	2011	**2014**
Mean Age	47.29	44.96
Men	56.96%	55.07%
Women	42.04%	43.54%
Missing	0.99%	1.39%
N_weighted	5 770 470	5 069 365

Source: Author's calculations based on PIT Data (2011 and 2014)

High Earners in the PIT Data

The South African tax authority has provided additional statistics[6] for top income earners who report taxable incomes above R10 million a year for each, 2011 and 2014. In order to protect their identity, SARS has only provided summary statistics for earners above this threshold, including the number of earners, maximum and minimum as well as mean and different quartiles of the distribution of top income earners. In 2011, there were a total of 482 individuals who reported a taxable income above R10 million, in 2014, this number had increased to 1048 individuals. SARS took these samples without accounting for inflation. In order to make the data of the two years comparable, prices have been adjusted for inflation in this analysis. Therefore, the 2014 levels appear much lower in the comparison of the different statistics provided by SARS between 2011 and 2014 in Table 3.5. In 2011, the smallest income reported in this segment of the population was at R10 005 439 and the first quartile can be observed at R11.5 million. The smallest value observed in the 2014 data on high earners was at R7 978 814 in 2011 prices. In real terms, the upward shift in the distribution mentioned in the discussion of Table 3.3 cannot be observed here except for the highest value reported. It is important to note that the median or second quartile is lower than the mean or average. The median in 2011 was at R14 million and in 2014 at R11.7 million. The fact that the mean is significantly higher in both years with a value of close to R17 million in 2011 and R15.8 million in 2014 indicates that the maximum values above R100 million in both years are outliers that drive up the average whereas half of these high-income earners actually report R14 million or less annual income. Additional to the income data, SARS provided some demographic statistics on the high earners. However, the information provided is rather scant. Table 3.5 shows that the mean age in this income group is about 56 years in 2011 and close to 53 years in 2014. This is higher than the mean age in the overall PIT data which is about 47 years in 2011 and about 45 years in 2014. Most interesting is the gender composition observed at the highest income levels. In

[6]SARS provided the data on high earners as summary statistics whereas the rest of the data contains a plethora of information for taxpayers.

Table 3.5: High Earners' Income Distribution

Quartiles	**2011 Income**	**2014 Income**
Min	10 005 439	7 978 814
Q1	11 551 323	9 302 885
Q2	14 078 782	11 774 718
Mean	16 987 843	15 783 480
Q3	17 844 600	16 861 538
Max	109 909 354	151 159 382
Demographics		
Mean Age	56.19	52.97
Men	85.89%	82.16%
Women	7.68%	8.78%
Missing	6.43%	9.06%
Number of Obs	482	1048

Source: Author's calculations based on PIT Data (2011 and 2014)

both years, this income group is dominated by men. In 2011, 85.89% of high earners were male compared to 82.16% in 2014. Only a small fraction of women can be observed at this very top end of the distribution. To recall, Table 3.2 found that there were slightly more women among the zero earners in NIDS but the discrepancy was not nearly as substantial as among the top earners. Compared to the overall PIT data provided by SARS, the gender composition reported in Table 3.4 is much more equal with about 56% of men in 2011 and about 55% in 2014. Furthermore, the discussion of Table 3.4 established that the population group of the "missing" gender is most likely to be foreign nationals. This fraction is much larger in the top earners than in the overall PIT data. According to SARS, only about 1% reported missing gender in the overall tax administration data compared to 6.43% among the high earners in 2011 and 9% in 2014.

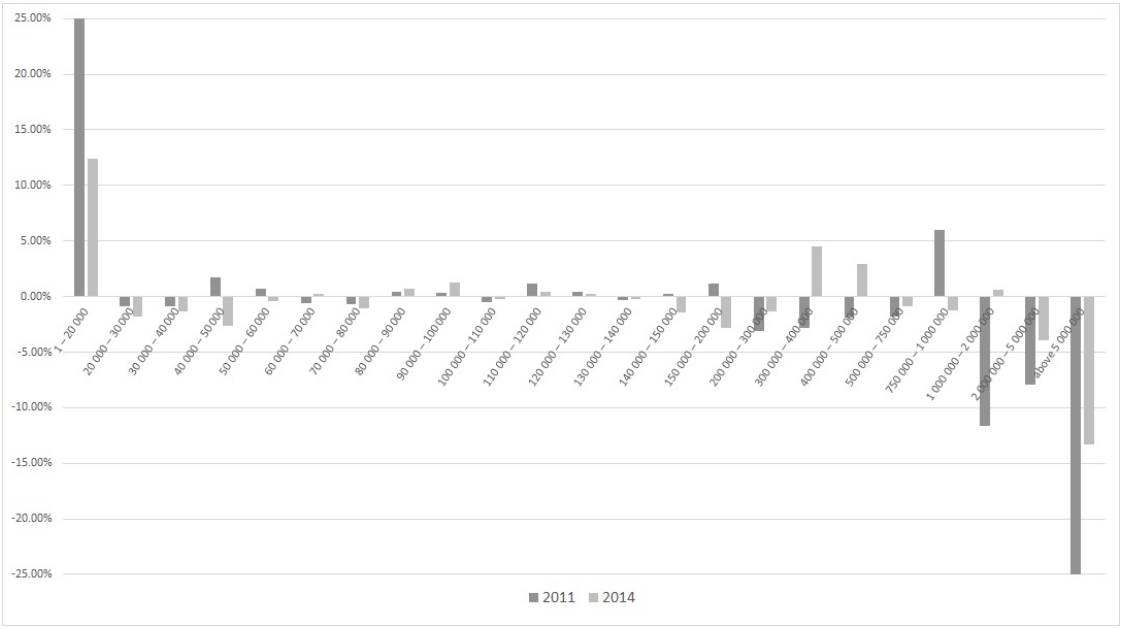
The information on taxable income of high-income earners discussed in this section will be included in the PIT data for the forthcoming analysis of inequality across different data sources.

3.4.3 Comparison of Household Survey Data with Tax Administration Data

As discussed in Section 3.3 above, a taxable income variable is derived in NIDS for individuals 18 years and older. In order to reconcile this analysis with the annual Tax Statistics published by SARS (2015), Section 3.4.2 then discussed the PIT data with regards to 24 income brackets with taxable income limits greater than zero. Figure 3.1 assesses the amount by which the taxable income variable in NIDS is greater or smaller than the PIT taxable

income. For this purpose, the mean of each of the 24 taxable income brackets has been calculated in NIDS and PIT. The results show that, with the exception of the extreme ends of the distribution, the difference of these mean taxable incomes across the different income brackets is negligible between NIDS and PIT.

At the bottom end of the distribution, NIDS estimates low-income earners at much higher percentages when compared to the PIT data. In 2011, the difference was as large as 25% for positive incomes below R20 000. However, the difference of mean taxable income between NIDS and PIT decreased to 12% for the lowest income bracket above zero in 2014. This difference at the bottom of the distribution is hardly surprising as not many individuals will have filed taxes when earning such small incomes and are therefore missing in the tax administration data but will have been covered by the household survey.



Source: Author's calculations based on NIDS and PIT (weighted)

Figure 3.1: Difference in Estimated Mean Taxable Income by Income Bracket between NIDS and PIT

In 2011, any individual earning R120 000 or more had to file tax returns. Around this threshold, the taxable income proxy created in NIDS reports very close estimates to the mean taxable incomes reported in the PIT data for both, 2011 and 2014. The estimated discrepancies between the two data sources start increasing for incomes of R200 000 and higher. Table 3.1 has shown that R200 000 is at the very top of the income distribution in the household survey in 2011 and 2014. At this point of the distribution, NIDS understates mean taxable income by about 3% in 2011 but only by about 1% in 2014. In 2011, the differences between the two data sets remained small until the income brackets above R750 000. For incomes between R750 000 and R1 000 000, NIDS actually overestimates mean

taxable income by about 6%. It is most likely that this is an outlier in the 2011 data as for the remaining higher income brackets, NIDS severely underestimates mean taxable income. In 2011 this underestimation accumulates to a difference of up to 25% compared to PIT for incomes above R5 000 000.

The mandatory filing threshold increased to R250 000 in 2014.[7] For incomes between R200 000 and R300 000, NIDS underestimates mean taxable income slightly. For the income brackets between R300 000 and R500 000, however, NIDS actually overestimates mean taxable income by between 3% and 5%. These are surprising findings as previous literature would indicate that households at the top end of the income distribution tend to under-report their incomes and not over-report as this would suggest. A possible explanation of this conundrum lies in the high attrition rates of top income households which will have affected the fourth wave of NIDS in 2014 to a more considerable degree than the second wave in 2011.[8] Due to increasingly high non-response rates of high-income households, those few that remain in the study will be attributed larger weights to ensure continued national representation of the study. It is possible that this has led to bias through over-weighting this top end of the income distribution. This issue will be investigated further in the following sections.

Furthermore, NIDS underestimates the incomes in the very top income brackets in 2014, however, to a much smaller degree than in 2011. For incomes between R2 000 000 and R5 000 000, NIDS estimates 4% less mean taxable income than PIT. For incomes above R5 000 000, the difference is just above 12%, about half of the difference in 2011. For all other income brackets, the differences between proxy taxable income in NIDS and taxable income provided in the PIT data seesaws around zero. Therefore, the under- and over-estimation of mean taxable incomes appears to behave somewhat like random fluctuation around the true mean in both years. In turn, this implies that NIDS performs reasonably in capturing mean taxable incomes from respondents. This is remarkable given the findings in international literature which compares household survey data and tax administration data of under-reporting of incomes earned in high-income households or higher non-response rate among such households (see for example Atkinson et al., 2011, van der Weide et al., 2016).

However, while taxable income levels and overall magnitudes in NIDS correspond fairly well to the same levels contained in the PIT data, inequality, is determined not by the overall magnitude of the cumulative total but its distribution. PIT and NIDS demonstrate substantial differences in the distribution of total taxable income, especially at the extreme

[7] To recall, the for inflation corrected threshold is at R199 397.

[8] See Finn et al., 2012 for a discussion of attrition rates in Wave 2 and Table 3.8 in Appendix 3.A for Wave 4.

ends of the distributions. As these ends have a significant impact on the overall measurement of inequality, the next sections will pursue these differences further in order to uncover what impact the data source has on the overall level of inequality in both, gross and net incomes.

3.5 Measuring Inequality

Traditionally, inequality in South Africa has been studied using household survey data (see for example Leibbrandt et al., 2012) and labour force survey data (see for example Wittenberg, 2016). Additionally, Alvaredo and Atkinson (2010) provide an exceptional analysis of top income shares over a one hundred year period using South African published data from the tax authorities but are unable to utilize it to assess overall inequality. To the author's knowledge, no study has attempted to use a combination of household survey data and tax administration data to assess income inequality across the entire income distribution in South Africa. As a consequence, it is possible to contribute to the existing literature on income inequality through access to a unique set of tax administration data and high-quality household survey data on South Africa.

The previous sections have discussed the two data sources used for this analysis in great detail. The discussion of the distribution in the household survey data has highlighted a large number of individuals earning zero taxable income. However, the number of these zero earners has decreased from 2011 to 2014 and mean incomes have increased. The vast discrepancy between the mean and median observed in the NIDS data indicated relatively high levels of inequality without using statistically more advanced methods. As in the household survey data, tax administration data indicated an upward shift in the overall distribution of taxable income. The discussion of extremely high earners in the PIT data showed a discrepancy between mean and median incomes, highlighting extreme outliers that drive up the average in this subset. This section uses more advanced methods to look at inequality in the two data sets separately and how best to combine the data sets to improve on existing analysis of income inequality.

At this point, it is essential to recall that this chapter measures inequality of taxable income of individuals rather than households. These measurements are going to be higher than inequality measures at the household level for two reasons. Firstly, the analysis measures taxable income which means other income sources relevant particularly to individuals at the bottom of the income distribution such as remittances and government grants will be left out. Chapter 2 has shown the immense impact that government grants had on lowering income inequality over the past twenty years, as did other studies such as Inchauste et al. (2015) and Schiel et al. (2016), so the large fraction of zero earners will have a significant

effect on the estimated measures of inequality. Secondly, in a household, there is sharing of resources such that even if one household member earns no income, the household as a whole may still report some form of revenue through another person's income. Taxable income of the individual by definition cannot account for such sharing of resources. The measurements of inequality of individual taxable income will be discussed in more detail in the following section.

3.5.1 Inequality within the NIDS and PIT Data

A commonly used measurement for inequality is the Gini coefficient. Table 3.6 presents the Gini coefficient for the household survey data for the different years. As mentioned previously, tax administration data is less reliable below the tax-filing threshold because only a decreased number of individuals will have filed taxes below the mandatory threshold and potential biases within this sample are not known. Therefore, the analysis of the Gini coefficient is broken up into segments above and below the tax-filing threshold and measurements of inequality for the PIT data are only reported above the filing threshold.

Table 3.6: Gini Coefficients at Different Thresholds

Threshold	**2011**		**2011 at 2014**		**2014**	
Data Source	**NIDS**	**PIT**	**NIDS**	**PIT**	**NIDS**	**PIT**
Overall	0.823	-	0.823	-	0.813	-
Below Filing Threshold	0.762	-	0.783	-	0.735	-
Above Filing Threshold	0.359	0.367	0.369	0.326	0.472	0.349

Source: Author's calculations using NIDS and PIT (weighted)

In 2011, the mandatory tax-filing threshold was at R120 000. Whereas in 2014, the mandatory tax-filing threshold was at R250 000. Due to this significant shift, one cannot merely compare the levels of inequality in 2011 with inequality in 2014. The two income distributions can only be sensibly compared when the same filing thresholds are applied to both years. For this reason, Table 3.6 assesses the level of inequality below the 2014 filing threshold in the 2011 data instead of only looking at the 2011 filing threshold. However, it is important to remember that the 2014 threshold is rather high up in the 2011 income distribution (between the 97.6th and the 97.7th percentile, see Table 3.1) which means the Gini coefficient below the filing threshold relies more strongly on incomes reported from high earners which is argued to potentially be biased in the household survey data. Therefore Table 3.7 reports the aggregation of inequality above and below the 2011 filing threshold as well as above and below the 2014 threshold.

Looking at taxable income in the NIDS data, the overall Gini coefficient in 2011 was relatively high at 0.823. When applying the 2014 filing threshold to the 2011 NIDS data, the level of inequality below the threshold increased from 0.762 to 0.783 in the household

survey. At the same time, the 2011 NIDS data also reports a higher Gini coefficient above the 2014 threshold. At this level, the Gini coefficient increased from 0.359 to 0.369. In 2014, a decrease in inequality can be examined. Table 3.1 indicated a significant reduction in the zero earners paired with an overall upward shift of incomes earned which might explain the decrease to a Gini coefficient of 0.813 in 2014. The level of inequality below R250 000 is lower at 0.735 compared to the 2011 values at the same threshold. At the same time, the level of inequality above the filing threshold has increased from 0.369 in 2011 to 0.472 in 2014.

The spike in inequality above the filing threshold as measured by NIDS between 2011 and 2014 in Table 3.6 is somewhat concerning. Therefore, Table 3.6 also compares the Gini coefficients above the respective tax-filing threshold for 2011 and 2014 to the PIT data sets. The estimation of the Gini below the filing threshold would be biased in the PIT data and is therefore not reported. The large increase in inequality above the filing threshold in the household survey data cannot be observed in the tax administration data. According to the PIT data, inequality increased only slightly from 0.326 to 0.349 when the same filing thresholds are applied. Otherwise, when the lower filing threshold is applied to the 2011 PIT data, the Gini coefficient actually decreases from 0.367 in 2011 to a Gini coefficient of 0.349 in 2014. However, it is not possible to draw a reasonable conclusion from this decrease as the distributions above the different filing thresholds cannot be compared.

As mentioned in the discussion of Figure 3.1, it can be assumed that the PIT data is much more precise in reporting higher incomes than the household survey data. Therefore, the discrepancies observed in Table 3.6 indicate a potential weighting issue in the NIDS data set in 2014. Such weighting issues can be caused by higher refusal rates in the upper income deciles compared to lower income deciles (see Table 3.8 in Appendix 3.A) or outliers in the household survey data. It is for these reasons that tax administration data is the preferred source of information to assess income inequality at the top of the income distribution.

3.5.2 Inequality across the NIDS and PIT Data

The comparison of the Gini coefficients in Section 3.5.1 provides some insight into the overall level of inequality as well as the difference in inequality measured in the NIDS data and the PIT data. Both of these types of data sets suffer from certain shortcomings. Previous studies have argued that household survey data suffers from distortions at the top end of the distribution (for example van der Weide et al., 2016 or Atkinson et al., 2011). Tax administration data, on the other hand, is much less reliable below the mandatory filing threshold as the subset of tax filers that report their incomes at this point of the distribution are unknown. Therefore the tax records below the filing threshold will not be representative. In order to understand the distributions of the household survey data and tax administration

data better, Figures 3.2 to 3.5 plot the kernel density functions of the NIDS vis-à-vis the PIT data. The results draw a clear picture of the shortcomings presented above.

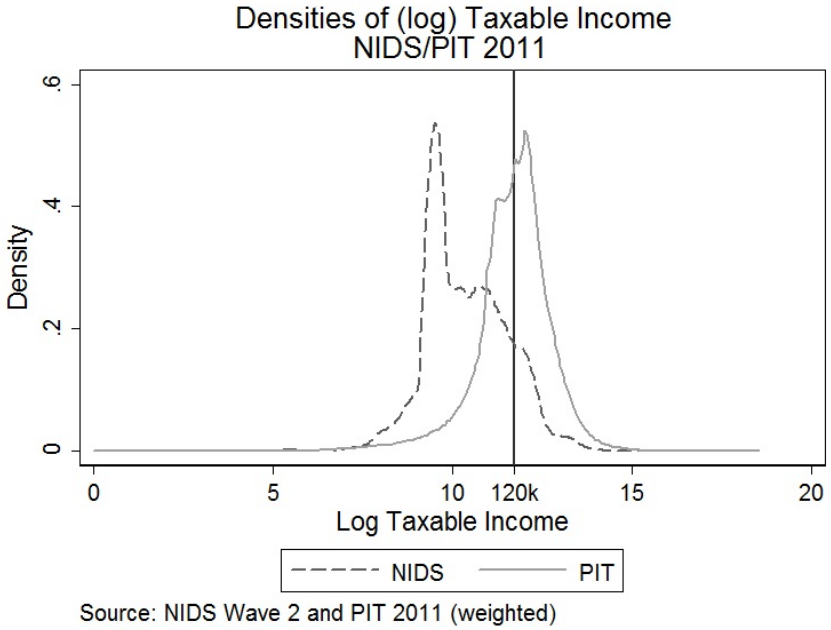


Figure 3.2: 2011 Taxable Income in NIDS vs. PIT

Figure 3.2 reports the densities of the continuous taxable income variable in logs from the household survey data and the tax administration data. In this graph, the household survey reports much more income below the filing threshold of R120 000 but is outperformed by the tax data in capturing incomes above the filing threshold which is marked as *120k*.[9] NIDS thereby shows a spike in incomes below the filing threshold that PIT lacks completely, implying that there many incomes that were successfully captured in the household survey whereas these incomes are missing in the PIT data. Because filing for tax reports is voluntary below the filing threshold, these incomes were not reported to the tax authority and are therefore not reflected in the PIT data. Due to the nature of the logarithmic transformation, Figure 3.2 does not picture the numerous zero earners discussed previously, another relevant part of the distribution missed by PIT. On the other hand, the tax administration data starts picking up incomes from just below the mandatory filing threshold that cannot be found in NIDS as evident by a spike in incomes around the *120k* mark that PIT captures but NIDS does not. In order to assess the accuracy of the household survey at measuring top incomes, Figure 3.3 takes a closer look at the distribution above the mandatory filing threshold.

The high spike in Figure 3.3 around a log of 12, an actual income of about R162 754,

[9]The logarithmic of R120 000 is about 11.7.

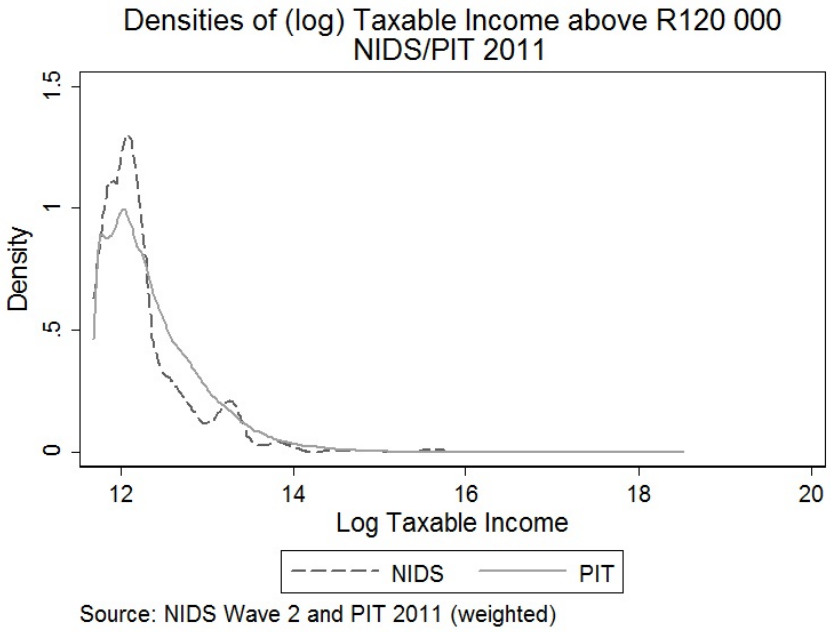


Figure 3.3: 2011 Taxable Income in NIDS vs. PIT above the Filing Threshold

may be the reason NIDS performs reasonably well when comparing mean incomes by income bracket at these levels of the income distribution. However, the household survey data trails off at the very high income levels and the PIT data reports incomes much higher than those reported in the NIDS data. The significance of the income levels above the filing threshold will be discussed in more detail in the following section.

Figure 3.4 shows a similar overall pattern for 2014. NIDS captures a lot more incomes below the filing threshold of R250 000 which is marked as *250k*.[10] Even though zero earners are again not pictured, NIDS captures a number of incomes very close to zero. The overall distribution in 2014 is rather similar to the 2011 distribution and so Figure 3.4 reports a large spike in incomes below the filing threshold as well. Again, the tax administration data fails to report any significant incomes at this segment of the distribution. However, closer to the mandatory filing threshold, PIT starts capturing incomes that NIDS does not report. Above the filing threshold, NIDS fails to capture a significant amount of incomes reported in the PIT data. This is emphasized by the distribution pictured in Figure 3.5.

In the graphs reported in Figure 3.5, the tax administration data shows a spike very close to the mandatory filing threshold that is missing in the NIDS data. Further up the income distribution, NIDS reports higher incomes than the distribution of the tax administration data. This is a surprising observation as existing literature focuses on the issue of under-reporting of top incomes in household surveys. However, it would seem that the issue of missing

[10]Corrected for inflation, the logarithmic of the R250 000 threshold is at 12.2.

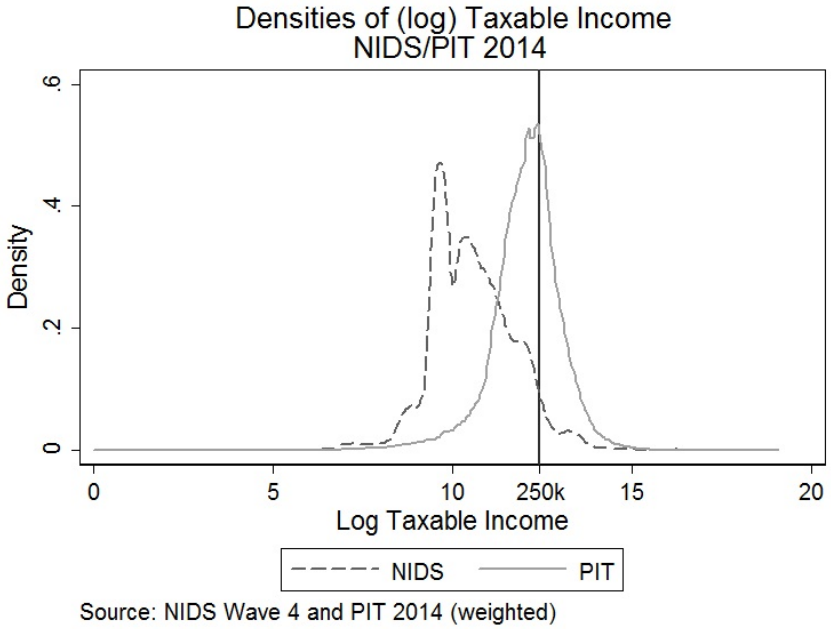


Figure 3.4: 2014 Taxable Income in NIDS vs. PIT

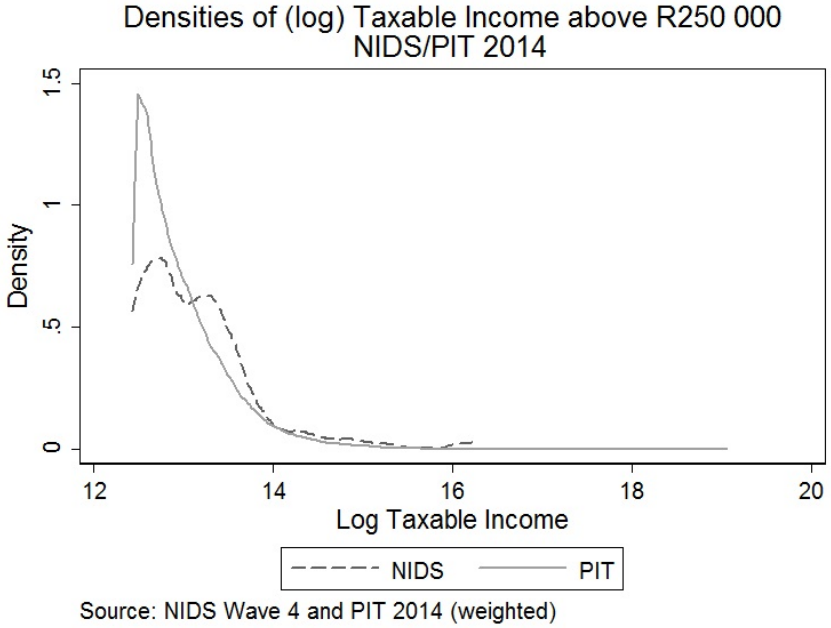


Figure 3.5: 2014 Taxable Income in NIDS vs. PIT above the Filing Threshold

observations in the top tail due to higher non-response of high-income households is more prevalent in the South African household survey. The few high-earning individuals that are observed will receive larger weights in order to represent everyone in the top tail of the income distribution. It appears that after four waves of the household survey, the higher attrition rate among high earners has introduced a bias at this part of the income distribution which leads to this surprising over-reporting of taxable income in NIDS. Nonetheless, after a significant hump at a log of about 16, a real income of about R8.9 million, the NIDS data fails to capture any further incomes. This outlier and the generally higher levels of income reported by NIDS up to that point may explain why inequality in the NIDS data was so much higher than in the PIT data as reported in Table 3.6; the distribution shown in Figure 3.5 is a lot more equal in the PIT data than in the household survey despite the fact that PIT captures high earners (with a log income above 16) that are absent from the NIDS data.

The figures discussed in this section show that in both years, household survey data captures incomes at the bottom of the income distribution that are not captured by the PIT data. At the same time, the tax administration data starts picking up incomes that are not reported in the NIDS data from income levels just below the respective filing thresholds. Additionally, Figure 3.5 highlighted a potential weighting issue in the NIDS data above the filing threshold. Because of these characteristics, the following section will discuss a method of how to assess income inequality across the two types of data in a way that captures all segments of the income distribution and estimate the true level of income inequality more accurately.

3.5.3 Combining Household Survey Data and Tax Administration Data

The discussion of the two types of data sets in this chapter has shown that the NIDS household survey captures the broad spectrum of incomes relatively well when compared to the tax administration data provided by SARS. However, the above figures highlight certain shortcomings in the capturing of the top incomes in the NIDS data. Since the top end of the income distribution is crucial to the analysis of overall inequality, it follows that tax administration data should be utilized to improve the overall analysis of inequality. Previous studies on the topic support this idea.

Atkinson (2007) highlighted that changes in the incomes of the top income earners can significantly affect overall inequality. This holds true even when the group of top income earners is rather small: "If we treat the very top group as infinitesimal in numbers, but with a finite share S* of total income, then the Gini coefficient can be approximated by S*+ (1 - S*) G, where G is the Gini coefficient for the rest of the population"(Pg. 19).

In a numeric example, an increase of 8% in S* would then lead to an increase of the Gini coefficient by 4.8% if the Gini for the rest of the population (i.e. 1 - S*) was at 0.4 prior to the increase. Several studies have therefore combined data from household surveys with tax administration data and Alvaredo (2011) further extends Atkinson's argument, asserting that depending on the degree of under-reporting in household surveys, one can use tax data from the top 1% up to top 5% or 10%. However, neither Atkinson (2007) nor Alvaredo (2011) provide a guideline as to which threshold would be preferable in order to analyse the entire income distribution.

Therefore, there exist many studies that attempt to solve the problem of including high-income earners in the assessment of inequality in different ways. Jenkins (2017), for example, uses a more principled approach and determines the appropriate threshold through statistical testing. Using both, Pareto Type I and Type II distributions, to estimate the top end of the income distribution, he concludes that the Pareto Type II estimates are more robust and that the threshold at which inequality estimates from tax return data and household survey data should be combined vary across years. Jenkins' (2017) detailed statistical approach is possible because of the availability of household survey data and tax administration data for the UK. Lakner and Milanovic (2016) use a different semi-parametric approach to combine inequality indices in order to account for high-income earners. Because of a lack of reliable tax administration data particularly in developing countries, the authors allocate the excess of national accounts private consumption over the mean of the household survey data to the top 10% for each country, thereby "elongating" the distribution using Pareto imputations. While this innovative approach is useful in the absence of reliable tax administration data, not only is the 10% threshold chosen somewhat arbitrarily but as Jorda and Niño-Zarazúa (2019) point out, the choice of Pareto imputations might not be the optimal distribution to use.

Alternatively to the semi-parametric approaches of Jenkins (2017) and Lakner and Milanovic (2016), Jorda and Niño-Zarazúa (2019) use a fully parametric approach to estimate the threshold at which household survey data is truncated. For this, they argue that a general class of distributions, the so-called Generalized Beta function of the second kind (GB2), is the most useful model specification as it has been shown to be an adequate fit to income data. The optimal truncation point is chosen to be that threshold at which the squared differences between the estimated top incomes using the GB2 distribution and the top incomes available from the tax administration data are minimized. This allows the optimal threshold to vary across countries. However, this is only possible for countries with tax administration data available. To account for high-income earners when measuring inequality in countries around the world, Jorda and Niño-Zarazúa (2019) have to resort to varying the threshold more intuitively and use GDP per capita as a proxy when tax administration data

is not available. Therefore, while Jorda and Niño-Zarazúa (2019) use the optimal threshold where possible, the overall choice at which to combine household survey data with data that accounts for high earners remains arbitrary.

In an effort to actively account for distortions due to reporting bias and higher non-response rates of high earners in household survey data, Diaz-Bazan (2015) argues that rather than choosing an arbitrary percentile at which to combine the information from the two sources of data, the inequality indices should be combined at an optimal threshold b. This optimal threshold should fulfil the following conditions: (a) individuals with incomes below threshold b are well represented in the household survey data; (b) the segment of the population reporting incomes above b are captured accurately in the tax records; and (c) the threshold is chosen such that distortions attributed to under-reporting at the top of the income distribution in household survey data are minimized. The optimal b that fulfils these conditions minimizes dependence on household survey data and ensures reliable information in the tax administration data. Therefore, the optimal threshold b is the lowest level of income that requires individuals to file taxes. As previously discussed, the mandatory tax filing in South Africa started at annual incomes of R120 000 in 2011 and R250 000 for 2014. Following Diaz-Bazan's argument, these are the optimal thresholds for combining the two data sources. Table 3.1 above has shown that in South Africa, the mandatory filing thresholds are well below the 1% thresholds but also above the 10% thresholds used in existing literature of combining tax administration data with household survey data. To recall, the filing threshold in 2011 was at the 93.6th percentile and in 2014, the filing threshold was close to the 97.5th percentile. The following sections will discuss the methodology of the approach introduced by Diaz-Bazan (2015) as well as the results of its application to the NIDS and PIT data in 2011 and 2014.

Methodology following Diaz-Bazan (2015)

If the optimal threshold b that satisfies the conditions outlined above is at the filing threshold, the Gini coefficient for the entire (unconditional) income distribution can be estimated by combining the (conditional) Gini coefficient below b, where F_1 is estimated using the household survey data, with the (conditional) Gini coefficient above b where F_2 is estimated from the tax administration data such that

$$F_1 = Pr\{Y \leq y | Y \leq b\}$$
$$F_2 = Pr\{Y \leq y | Y > b\}.$$

Given the definition of the conditional distribution, the distribution for $y \leq b$ is represented by

$$F_1(y) \equiv Pr\{Y \leq y | Y \leq b\} = \frac{Pr\{Y \leq y\}}{Pr\{Y \leq b\}} = \frac{F(y)}{F(b)}$$

so that $F(y) = F_1(y)F(b)$.

Similarly, the distribution for $y > b$ can be written as

$$F_2(y) \equiv Pr\{Y \leq y | Y > b\} = \frac{Pr\{Y \leq y \ and \ Y > b\}}{1 - Pr\{Y \leq b\}} = \frac{F(y) - F(b)}{1 - F(b)}$$

so that $F(y) = F(b) + (1 - F(b))F_2(y)$.

This implies that it is possible to reconstruct the underlying income distribution $F(y)$ if the two conditional distributions above and below the threshold b can be observed and the fraction of the population with income above and below threshold b are known.

In order to obtain a measurement of inequality for the entire population, it is possible to estimate the Gini coefficient using a linear combination of the Gini coefficients computed on the conditional distributions above and below threshold b. In this case, the Gini coefficient of a conditional distribution, $G[F_j]$, $j \in 1; 2$ can be written as

$$G\left[F_j\right] = \frac{1}{\mu_j} \int_0^\infty F_j(y)(1 - F_j(y))dy,$$

where $\mu_1 \equiv E[Y|Y \leq b]$ and $\mu_2 \equiv E[Y|Y > b]$. This G is equivalent to the standard Gini coefficient calculated on the two distributions.

This can then be combined into a full population, unconditional Gini coefficient defined as

$$G\left[F\right] = (1 - F(b))F(b)[\frac{\mu_2 - \mu_1}{\mu}] + \left(F(b)^2 \frac{\mu_1}{\mu}\right) G[F_1] + \left(F(b)^2 \frac{\mu_2}{\mu}\right) G[F_2],$$

where $\mu = F(b)\mu_1 + (1 - F(b))\mu_2$ and $G[F_i]$ $i \in \{1, 2\}$ are the Gini coefficients of F_1 and F_2.

Therefore, the (unconditional) Gini coefficient can be written as a linear combination of the two Gini coefficients of the conditional distributions F_1 and F_2 plus an adjustment term that corrects for the fact that the two conditional distributions are obtained from different parts underlying the unconditional income distribution.

Application to the South African case

In order to obtain this derived Gini coefficient, Diaz-Bazan (2015) highlights the fact that consistent estimates of the distributions above and below threshold b are required. The availability of a high-quality household survey as well as a robust data set on tax records facilitates the application of this novel method to household survey data and tax administration data on South Africa. Table 3.7 reports the levels of inequality observed when the Diaz method is used to combine household survey data with tax administration data at the mandatory tax-filing threshold for 2011 and 2014 respectively. The Gini coefficients are calculated using the individual taxable income of adults 18 years and older. The decrease observed in the combined inequality measure is much larger between 2011 and 2014 than what was reported when using household survey data alone in Section 3.5.1. From a total level of 0.83 in 2011, the Gini coefficient for taxable income decreased to 0.79. This extreme drop in inequality observed is driven by the fact that both conditional distributions report lower levels of inequality in 2014. Table 3.6 reported a slight decrease in inequality between the two years when using NIDS alone. However, in the household survey data, the decrease in the (conditional) Gini coefficient below the filing threshold was offset by a large increase in estimated inequality above the filing threshold. When using tax administration data above this threshold in order to explicitly account for the distortions caused by reporting bias and non-responses in the household survey data, a much lower Gini is estimated for 2014. The decrease in overall inequality can also be observed when the 2011 data is assessed at the 2014 filing threshold albeit to a lesser degree. Applying the 2014 threshold to the 2011 data set increases dependence on household survey data at the top of the income distribution, however, it ensures comparability across the two years despite the change in the underlying tax system. In this case, inequality decreased from 0.826 in 2011 to 0.79 in 2014 when the two types of data sets are combined at the same filing threshold.

Table 3.7: Gini Coefficients Combined at Different Thresholds

Year	**2011**	**2011 at 2014**	**2014**
Combining data sets at different filing thresholds			
Threshold	R120 000	R250 000	R250 000
Overall	0.832	0.826	0.790
Below Filing Threshold (NIDS)	0.762	0.783	0.735
Above Filing Threshold (PIT)	0.367	0.326	0.349
Combining data sets at the 99th percentile			
Overall	0.826	-	0.791
NIDS	0.782	-	0.742
PIT	0.328	-	0.343

Source: Author's calculations using NIDS and PIT (weighted)

Additionally, Table 3.7 reports the Gini coefficients derived when household survey data and

tax administration data are combined at the 99th percentile. For 2011, this reduces the Gini coefficient slightly from 0.832 according to the Diaz-Bazan method to 0.826 when combined at the 99th percentile. For 2014, however, the difference is barely noticeable as the Gini increases from 0.790 according to the Diaz-Bazan method to 0.791 when combined at the (arbitrarily) chosen 99th percentile. This increase is most likely driven by the large Gini coefficient above the filing threshold mentioned earlier that drives up the Gini coefficient measured in the household survey data. The fact that the change in the Gini coefficient resulting from the change in thresholds is less significant in 2014 is most likely owed to the filing threshold of R250 000 being much closer to the 99th percentile in 2014 than it was in 2011. In the 2014 distribution, the filing threshold was at the 97.5th percentile. In 2011, the filing threshold was relatively lower at the 93.6th percentile. The 2014 filing threshold was at the 97.7th percentile of the 2011 distribution. When the 2011 data set is assessed at this threshold, the level of overall inequality observed is virtually the same as at the 99th percentile. Nonetheless, choosing a more pragmatic threshold to combine the two data sets to explicitly account for reporting bias and non-response in the household survey data is necessary for consistent estimation of inequality. Therefore, the filing threshold is the preferred level at which the two types of data sets should be combined.

3.6 Conclusion

This chapter takes advantage of the availability of high-quality household survey data and access to a unique set of tax administration data and uses a novel method in order to assess levels of inequality prevalent across the entire population in South Africa. The analysis of taxable income in the household survey data highlights a large fraction of zero income earners. This income group may earn income from other sources such as remittances or government grants but reports no taxable income. This segment of the population is mostly composed of young adults between 18 and 30 years old, both in 2011 and 2014. Furthermore, Africans and women are more likely to be in this income group than others. A similar analysis of the individuals at the top end of the income distribution of the tax administration data has shown that of the individuals earning R10 million or more annually, more than 80% are male in both years and an increasing fraction of high earners is likely to be foreign.

The examination of the income distribution in the two underlying data sets has shown that there was a significant upward shift in taxable income between both years that can be observed across data sets. Using the Gini coefficient as a measure of inequality, this chapter has shown that inequality decreased when using household survey data to assess the overall population. The NIDS data reports a Gini coefficient of 0.823 for taxable income in 2011 and a Gini coefficient of 0.813 in 2014. Similarly, inequality measured above the filing threshold in the tax administration data has decreased over this period. Naturally,

the measured inequality above the filing threshold is much lower than overall inequality and decreased from 0.367 in 2011 to 0.349 in 2014 according to the PIT data. Tax administration data below the filing threshold is unreliable and therefore cannot be used to assess the income distribution as a whole.

In the discussion of inequality across the different types of data sets, the analysis showed that there is an optimal threshold at which household survey data and tax administration data can be combined. This threshold minimizes dependence on household survey data while ensuring reliable information from tax administration data. The chapter argues that this threshold should not be chosen arbitrarily and should instead be defined as the mandatory filing threshold. Detailed analysis of the different data sets has shown that household survey data in South Africa is unreliable above the filing threshold, especially in 2014. Therefore, household survey data from the NIDS survey is used to estimate the conditional distribution below the tax filing threshold and administration data on personal income tax is used for the estimation above the filing threshold. The results show that the decrease in inequality between 2011 and 2014 is even more significant in the estimation of the Gini coefficient of the thus combined income data. Applying the methodology introduced by Diaz-Bazan (2015), the Gini coefficient of taxable income decreased from 0.832, or 0.826 depending on the threshold applied, in 2011 to 0.79 in 2014. A decrease in overall inequality is also observed when NIDS data and PIT data are combined at the 2014 filing threshold. Assessing the 2011 data sets at the 2014 filing threshold increases dependence on household survey data at the top of the income distribution for that year, however, it ensures comparability across the two years despite the shift in filing thresholds. This comparison shows a slightly smaller decrease in inequality from a Gini coefficient of 0.826 to 0.79. The fact that the results differ highlights why it is important to analyse the shift in thresholds at which the different types of data are combined. When comparing the estimated Gini coefficients of the data sets combined at the 2014 threshold, the overall decrease in inequality is driven by lower inequality below the threshold, most likely due to fewer zero earners. However, increasing inequality at the top of the distribution counteracts these effects.

Despite the decrease in inequality between 2011 and 2014, the results are quite alarming for several reasons. Firstly, the extremely large fraction of zero earners indicates a strong dependency on non-taxable income sources. This holds pitfalls not only for the state and its capacity to generate revenue but also for the individuals at the bottom of the income distribution. Most forms of non-taxable income, such as informal labour, will either be small, less reliable, or both compared to more structured forms of income that can be taxed. Government grants form a reliable and important source of non-taxable income but leave individuals at the bottom of the distribution dependent on external forces, like the government, to receive an income rather than being able to generate an income for them-

selves. The fact that there are relatively more women in this group than men highlights an additional factor of vulnerability of a group that is already more likely to be poor (Posel and Rogan, 2012). Contrary to this, the group of extremely high-income earners is almost exclusively made up of men. The gendered discrepancies at different parts of the distribution are cause for further investigation. Chapter 4 will look into changes of the entire earnings distributions of men and women more closely.

There is a strong need for the state to address the severe deficits of individuals at the bottom of the income distribution. Chapter 2 has shown the strong positive effects of increased government grants in addressing household at the bottom of the income distribution, however, seeing the macro-economic state of South Africa, it becomes clear that continued heavy reliance on government grants may no longer be feasible. Indeed, Van der Berg and Moses (2012) also highlight the effectiveness of targeting South Africa's poor through government grants, however, the authors caution that the South African government is reaching its fiscal limits with regards to social spending as a tool for redistribution. Therefore, more structured ways of generating income need to be facilitated. Such solutions will require the government to introduce policies that induce growth and develop labour market policies to promote (self-)employment as well as skill development.

One way of improving individuals' chances to find employment is through increasing human capital. This is particularly relevant in order to address the high levels of youth found in the group of zero earners. As such, government has tried to improve access to tertiary education for people particularly from the bottom of the income distribution since the announcement of free higher education for new first year students from lower income families in 2017. This is one way to address the high levels of inequality in education overall (Crouch et al., 2009; Spaull, 2013). However, Inchauste et al. (2015) found that spending on higher education was the least progressive of any possible social expenditure. Thus, in order to spend already strained funds more efficiently, the focus should be on improving the access to quality primary and secondary education.

The inclusion of tax administration data enables better estimation of the current level of inequality as well as of the trends in inequality. Otherwise the downward trend in inequality would have been underestimated considerably. A method that solely uses household survey data for the analysis of inequality would have underestimated the actual decrease in inequality due to potential weighting issues at the top of the income distribution. Therefore, the use of tax administration data not only improves the understanding of how the Gini coefficient is constituted but also of what role different parts of the distribution play. Access to this data therefore enabled a new type of analysis in the South African context which means this chapter makes an important contribution to the understanding of income inequality in

South Africa and the role of high income earners. The novelty of the method by Diaz-Bazan (2015) used in this chapter is only exceeded by the fact that this chapter pioneers a type of analysis that combines household survey data with tax administration data in the South African context. However, more research can be done to identify the causes of the decrease in inequality observed in this chapter and to determine their long-term feasibility in order to ensure continued efforts to transform South Africa into a more equal society.

3.A Additional Tables and Figures

Table 3.8: NIDS Attrition Rate by Deciles

Deciles by Wave 1 HH Income (2008)	**Percent refused in Wave 4 (2014/2015)**
1	5.45%
2	5.92%
3	5.63%
4	7.50%
5	5.33%
6	7.33%
7	9.44%
8	9.20%
9	15.30%
10	28.90%
Total	100%

Source: Author's calculations using NIDS.

Note: HH is short for household

Table 3.9: Statutory Tax Rates in South Africa

2011	**2014**	**Marginal PIT rates**
0 - 140 000	0 - 165 600	18%
140 001 - 221 000	165 601 - 258 750	25%
221 001 - 305 000	258 751 - 358 110	30%
305 001 - 431 000	358 111 - 500 940	35%
431 001 - 552 000	500 941 - 638 600	38%
552 001 and over	638 601 and over	40%

Source: SARS (2015)

Table 3.10: Comparison of the Information in the PIT and NIDS Data sets

	PIT Data		NIDS Equivalent	
	Item	Explanation	Variables	Notes
	Local interest	Local interest earned†	indi (dividends and interest)	indi does not differentiate local and foreign dividends, cut-off of R22,300
+	Local capital gains	Excludes the basic exemption for capital gains (exclusion rate in 2010-11: 75%)	rnt	
−	Local capital losses	Local dividends†; rental profits/ losses; income from building societies; income from fixed period shares and deposits; royalties; foreign investment income (interest, dividends, capital gains/losses); gambling gains/losses	*not covered*	
+	Other gains∗		*not covered*	
−	Other losses∗		*not covered*	
=	*Investment income incl. capital gains*			
+	Business profits	Profits/losses from unincorporated businesses	emsprof_a	Profit after tax
−	Business losses	or trades	emsloss_a	
=	*Business income*			
+	Normal income	Local and foreign labour and pension income	em1inc em2inc swag prof ppen spen	does not differentiate between local and foreign benefits
+	Fringe benefits	Local and foreign lump-sum income, including remuneration and pension/ provident fund lump-sums	*not covered*	
+	Lump sum income		extr cheq bonu, em1prflm_a	does not differentiate between local and foreign benefits
=	*Labour income*$^{++}$			
=	*Gross income*	All incomes received by the individual		
−	Deductions	E.g., Taxes paid under pay-as-you-earn; pension, provident or medical fund contributions The exempted portion of interest income, all local dividends	em1dedmed_a em1dedpen_a em1deduif_a	
−	Exemptions		*not covered*	
=	*Taxable income*	Taxable income used to determine the normal tax due (before any rebates and tax credits)		

Source: Orthofer (2016) and NIDS

Note: ∗ Asterisks refer to the subset of items under the respective SARS Code that not mentioned separately in the table.

† Local interest below the threshold of R22 300 and local dividend income in its entirety is exempt from the PIT, and the accuracy of exempt incomes is not verified in the tax inspection process.

$^{++}$ Employment income derived from taxable normal and lump sum income (only taxable portion of normal and lump sum income provided by SARS).

Chapter 4

Evolution of Earnings Inequality in the Post-Apartheid Era

4.1 Introduction

In 1994, the first democratic government of South Africa was elected and implemented numerous policies to redress the systematic discrimination of the apartheid regime. However, income inequality has remained high over the post-apartheid era and existing literature shows that overall income inequality is driven by inequality in the labour market (see for example, Leibbrandt et al., 2010). The results of the decomposition of overall household incomes performed in Chapter 2 support these findings. Therefore, advanced micro-simulations will be used again in this chapter. However, using detailed labour market data in order to identify the driving effects behind changes in earnings inequality between 1993 and 2012, this chapter applies micro-simulations to the earnings distribution.

Many methods can be used to explain differences in distributions across genders, nations or time. This chapter uses a micro-simulation approach initially developed by Bourguignon et al. (2008) that moves beyond the traditional Oaxaca-Blinder (1973) decomposition method. The Oaxaca-Blinder (1973) decomposition method was groundbreaking at the time and enabled a dissection of the effects of differences in characteristics and differences in returns on wages when comparing two distributions. However, these effects could only be ascertained at the mean of the wage distribution. The aim of this chapter is to understand what factors are driving the evolution of earnings at different points of the distribution, not only at the mean. This can be achieved through a set of counterfactual distributions that account for changes in the underlying determinants of the earnings distribution which make it possible not only to identify the 'price effect' and 'endowment effect' of the traditional Oaxaca-Blinder (1973) decomposition but also extract the effects of other important labour market determinants.

Since the end of the apartheid regime, South Africa is characterized by high levels of structural unemployment. The official unemployment rate increased from 22% in 1994 to 27.2% in 2018 (Statistics South Africa, 2018). However, the application of micro-simulations by Bourguignon et al. (2008) assumes clearing conditions in the labour market. This would make simply applying the approach of Bourguignon et al. (2008) to the South African context a poor choice. Therefore, the analysis in this chapter follows a modification of the Bourguignon-style micro-simulation introduced by González-Rozada and Menendez (2006). This approach actively accounts for the effect of high structural unemployment on labour participation and the employment outcome.

The micro-simulations performed in this chapter take advantage of a collection of long-term data consolidated in the Post-Apartheid Labour Market Series (PALMS). The particular data sets used in this chapter are from 1993, 1995, 2001, 2003, 2010 and 2012. These years

are chosen based on comparable data and instruments which are discussed in Section 4.3.

By estimating participation and employment outcomes, this chapter finds that the participation rate of women increased significantly without corresponding increases in the female employment rate. This makes women particularly vulnerable to finding themselves in informal labour (Casale and Posel, 2002). Through carefully applying micro-simulations and estimating counterfactual wage distributions, the analysis performed in this chapter indicates that changes in endowments contributed significantly to the increase in earnings inequality for both, men and women. Additionally, the results show that the employment effect contributed considerably to the rise in the Gini coefficient for women's earnings between 1993 and 2012. The employment effect measures how changes in the make-up of those who find employment as well as in the employment rate affect changes in inequality over time. To a lesser degree, the employment effect also increased earnings inequality for men. For both, men and women, changes in returns to observed characteristics counteracted some of these increases. Overall, female earnings inequality measured by the Gini coefficient increased from 0.583 to 0.6 whereas male earnings inequality increased from 0.591 to 0.617 between 1993 and 2012.

The micro-econometric decompositions performed in this chapter aim at providing empirical evidence of the evolution of earnings inequality by estimating the contribution of the different effects to the overall changes. This evidence can enhance the understanding of labour market trends which is particularly relevant in a country faced with the detrimental combination of high unemployment and high earnings inequality.

The remainder of this chapter is structured as follows. The existing literature on the South African labour market as well as the literature on micro-simulations will be discussed in Section 4.2 followed by a detailed discussion of descriptive statistics of the variables underlying the analysis performed in this chapter in Section 4.3.1. The underlying estimations of this analysis are discussed in Section 4.4 as well as the methodology of the counterfactual micro-simulations. The results of these simulations are presented in Section 4.5. Finally, Section 4.6 summarizes the findings and concludes.

4.2 Literature Review

It has been established in the literature that inequality in the labour market is driving overall household income inequality in South Africa. Chapter 2 analysed household income using both, static and dynamic decomposition methods. The use of dynamic micro-simulations to assess the role of changes in different income sources in changing inequality thereby offered new insights into understanding the drivers of income inequality. Micro-simulations are

rare, particularly in the South African context, but very useful in decomposing changes in inequality. Furthermore, Chapter 2 found that labour market income drives overall income inequality in South Africa. Therefore, this chapter uses another micro-simulation approach to analyse earnings inequality in South Africa following a method first introduced by Bourguignon et al. (2008). This section discusses existing literature on the South African labour market before reviewing the literature on Bourguignon-style micro-simulation.

4.2.1 Literature on the South African Labour Market

South Africa's consistently high levels of income inequality in general, and earnings inequality in particular, have drawn attention from numerous researchers. Most recently, Finn and Leibbrandt (2018) use re-entered influence functions (RIF) to trace changes in earnings inequality in South Africa between 2000 and 2014. The authors identify a significant shift in earnings inequality after 2011 which is possibly caused by a change in the methodology of imputations performed by Statistics South Africa (Stats SA) from that year forward. Up until 2011, the Gini coefficient of earnings had been relatively high but stable. After 2011, there is a sudden increase in this inequality measurement that cannot adequately be explained even when the authors use different methods to analyse the changes. The change in the methodology that potentially caused this shift will be discussed in the data concerns raised in Section 4.3.

Wittenberg (2016a,b) offer a detailed analysis not only of the development of earnings inequality between 1994 and 2011 but also presents a detailed discussion of measurement issues and methodological changes underlying the PALMS data set. The first of these papers discusses the earnings of employees (without the self-employed) between 1994 and 2011. The overall review of the earnings distribution showed that incomes below the median have become more compressed during that period whereas the top of the distribution of earnings moved away from the median. The second paper stresses the importance of measurement issues and changes in the survey instruments. Wittenberg (2016b) finds that the Gini coefficient and other inequality measurements increased between 1994 and 2011 and that trends in wage inequality are driving the trends in overall income inequality. Finally, through a decomposition of the inequality of earnings by race and education, it becomes clear that within-race earnings inequality explains a growing share of overall earnings inequality.

When studying labour markets and the earnings distribution, selection into the labour market plays a critical role. Particularly in the South African context, it is essential to account for potential selection bias (Heckman, 1979). Bhorat and Leibbrandt (1999a) address the decision to participate separately from the selection into employment. The authors achieve this by building a three-part model for participation, employment and earnings. Van der

Westhuizen et al. (2007) relate to the work of Bhorat and Leibbrandt (1999a) as well as Casale (2004) by adopting a three-step model for participation, employment and earnings. Their findings for the period between 1995 and 2005 ring similar to the results of Casale (2004) which manifests the significant increase in female unemployment in the first decade after apartheid, particularly for African women. Additionally, van der Westhuizen et al. (2007) found that the majority of women who found jobs between 1995 and 2005 were employed as unskilled and low-paid elementary workers, supporting the findings of other studies such as Bhorat and Leibbrandt (1999b), Oosthuizen (2012) and Bhorat et al. (2016) regarding the vulnerability of African women in particular.

Focussing further on gender discrimination in the South African labour market, Wittenberg and Ntuli (2013) investigate the changes in the labour force participation of black women using regression analysis. Their results indicate that the increase in labour force participation among black women was mainly associated with increases in returns to individual characteristics rather than changes in the women's characteristics between 1995 and 2004. Other papers that focus on gender discrimination, such as Bhorat and Goga (2013) show that there exists a 'sticky floor' in the earnings distribution in South Africa, where the gender wage gap is larger at the bottom of the distribution than at the top. Kwenda and Ntuli (2018) use re-centered influence functions to show that the gender pay gap in South Africa is larger in the private sector which is driven, at least at the bottom of the earnings distribution, by lower endowments of workers in the private sector.

The papers reviewed above offer significant insights into the structure of the South African earnings distribution. However, they fall short in accounting for the non-market clearing effects of structural unemployment, a problem that characterizes the labour market in South Africa. Therefore, this chapter aims to contribute to existing literature by actively accounting for the high levels of unemployment in South Africa while using micro-simulations to decompose the effect of returns to individual characteristics, the change in observed characteristics as well as the effects of changes in participation, employment and occupational choices. The literature that this methodology is based on is reviewed in the following section.

4.2.2 Literature on Micro-Simulations

The literature has long been assessing the effect of changing characteristics and changing returns on (mean) wages and one of the most well-known decomposition methods is the Oaxaca-Blinder decomposition (Oaxaca, 1973 and Blinder, 1973). Oaxaca and Blinder (OB) both find that the difference in mean wages across two distributions can be decomposed into $\bar{y}^A - \bar{y}^B = \bar{x}^A(b^A - b^B) + (\bar{x}^A - \bar{x}^B)b^B$, where A and B are the two distributions under investigation and b^i are the returns to characteristics x^i in distribution i (Fortin et

al., 2011). In this decomposition, $(\bar{x}^A - \bar{x}^B)b^B$ is the *endowment effect*, the effect of changes in observed characteristics, and $\bar{x}^A(b^A - b^B)$ is the *price effect* which is the effect of changes in prices or returns to observed characteristics at constant mean endowments. The Oaxaca-Blinder method is an invaluable tool to decompose differences in mean wages, however, there now exists a growing literature (including for example Dinardo et al., 1996; Ferreira and de Barros, 1999; Lemieux, 2002; Firpo et al., 2009; Fortin et al., 2011) that is looking for methods to decompose differences across entire distributions rather than across means. As such Bourguignon et al. (2008) move "Beyond Oaxaca-Blinder" by estimating the 'price effect' and 'endowment effect' not only at the mean but over the entire distribution by utilizing micro-simulations. Furthermore, the implementation of micro-simulations allows for decomposing the changes in the distribution into factors beyond price and endowment effect by accounting, for example, for the impact of changes in the participation and employment rate or occupational structure.

González-Rozada and Menendez (2006) add to the method developed by Bourguignon et al. (2008) in that they apply a modification that accounts explicitly for non-clearing labour markets. González-Rozada and Menendez (2006) use the micro-simulation approach to study how changes in the labour market affected income inequality and poverty in Argentina. The Argentinian labour market underperformed during an era of economic growth in the 1990s which led to an increase in income inequality, growing unemployment and increased levels of poverty. Therefore, the authors introduce disequilibrium into the method to assess whether the changes in inequality are correlated with changes in unemployment. González-Rozada and Menendez (2006) use a sequential Poisson-sampling approach for selection into participation and employment to mimic imperfect selection outcomes in the labour market. As such, they define a random term that is used to re-scale the estimated probability of participation or selection respectively. However, Garlick (2016) notes that the definition of the random term chosen by González-Rozada and Menendez (2006) is an unstable one. Therefore, she defines an alternate randomization process which the analysis performed in this chapter uses as well. The overall methodology will be discussed in more detail in Section 4.4. Both, González-Rozada and Menendez (2006) and Garlick (2016) focus on the effect of changes in the educational structure of the population in their assessment of changes in household income and earnings respectively. Garlick (2016) thereby offers the first analysis of this kind in the South African context and finds that changes in educational attainment increased earnings inequality through its effect on participation but decreased inequality through the effect on returns to education from 1993 to 2008.

This chapter complements and integrates these studies through analysing the effect of changes in participation, employment, occupational structure, endowments and returns to these endowments, and changes in unobservables on changes in labour market earnings over

twenty years in South Africa. Due to the high levels of unemployment that characterize the South African labour market, this study applies the modification of the Bourguignon et al. (2008) approach that allows for non-clearing markets. Finally, since the differences in the treatment of men and women in the labour market have been established by the literature reviewed in Section 4.2.1, all estimations will be reported separately for men and women. Section 4.4 explains the details of the methodology applied in this chapter after Section 4.3 discusses the data.

4.3 Data Issues

The data used in this study comes from the Post-Apartheid Labour Market Series (PALMS). PALMS harmonises micro-data from a total of fifty-four household surveys into a stacked cross-sectional data set (Kerr and Wittenberg, 2016). The data set covers household surveys conducted by Statistics South Africa between 1994 and 2017, including the October Household Surveys (OHS) from 1994-1999, the bi-annual Labour Force Surveys (LFS) from 2000-2007, and the Quarterly Labour Force Surveys (QLFS) from 2008-2017. Additionally, PALMS also includes the 1993 Project for Statistics on Living Standards and Development (PSLSD) conducted by the Southern Africa Labour and Development Research Unit (SALDRU).

For the purpose of year on year comparison, this study uses data from the 1993 PSLSD, the 1995 OHS, the 2000 and 2003 LFS as well as the 2010 and 2012 QLFS. These years were chosen due to the availability of comparable information across years. All of the surveys chosen contain data on individuals, such as age, population group, education and marital status as well as detailed information on economic activity including real earnings, sectoral and occupational choices. Additionally, the surveys offer some information on the household level such as household size and geographical area of the household.

As the name indicates, the October Household Surveys (OHS) were sampled in October of each year. Therefore, in order to ensure comparability across all years chosen for the analysis in this chapter, the second instalment of the relevant LFS is used which were conducted in September of 2001 and 2003 respectively. Furthermore, the second quarter of the QLFS for 2010 and 2012 are used.

Wittenberg (2016a) offers a detailed discussion of the measurement issues and changes in instruments that are inherent to a collection of different data sets such as PALMS. Some of the most noteworthy changes within the PALMS data set are the fact that between the OHS and LFS, two earnings questions regarding wage work and self-employment were summarized into one. In the QLFS, which followed the LFS's, the questions were sepa-

rated again. However, comparison across surveys is complicated by the fact that in the QLFS, the question was structured in a way such that individuals could not report both types of income. Furthermore, the question on earnings was removed in the LFS and QLFS between 2007 and 2009 and was only added back in the 2010 QLFS. As the variable on real earnings forms the basis of the wage analysis of this study, the years surrounding the global financial crisis could not be included despite potentially interesting effects at the time.

Additionally, Wittenberg (2016a) reports a significant increase in the coverage of marginal workers between the 1999 OHS and the 2000 LFS. This led to a dramatic shift in the earnings distribution that cannot be sensibly corrected for. Finally, in the PALMS V3.1 user manual, Kerr and Wittenberg (2016) note a significant change in the way Statistics South Africa (Stats SA) published the data from the 2010-2015 QLFS. The review of other studies using PALMS data in Section 4.2.1 alluded to changes in the survey methodology implemented by Stats SA. Between 2010 and 2012, Stats SA imputed values for individuals who refused or gave a bracket response without the researcher being able to identify who reported an income directly and who was assigned a value in this manner. After the third wave of the 2012 QLFS, only partial imputations were implemented (Finn and Leibbrandt, 2018) and Kerr and Wittenberg (2016) observe some indication of top coding. The extent of the effect of this change in the methodology of imputations cannot be quantified. Due to these inconsistencies in how non-responses and bracket responses are dealt with after that year, 2012 was chosen as the final year of observation for this chapter. As discussed above, only the second quarter of the 2012 QLFS data will be used for the analysis performed in this study to ensure comparability across the years chosen. The following section discusses the descriptive statistics of the relevant years in more detail.

4.3.1 Summary Statistics

The PALMS data set contains harmonised variables and the surveys used in this analysis have been chosen based on comparable information available. The OHS, LFS and QLFS have been stratified by province, district, geographical area (urban/rural) and population group. All subsequent descriptive statistics and analyses take into consideration sampling weights, specifically, PALMS offers a number of different weights and the analysis performed below uses cross-entropy weights to ensure consistency of demographic and geographic trends over time.

Table 4.1 below reports the demographics and geographical information on all individuals aged 15 to 64 years old, including sector and occupations chosen by individuals that participate in the labour market. The age range was set at 15 to 64 since South Africans are legally allowed to work from 15 years old and this definition is broadly accepted as the

description for working age in South Africa. However, this broad definition of the working age leads to the following two problems. Firstly, most South Africans stay in school beyond the age of 15. Thus, they do not participate in the labour market yet. Additionally, eligibility for the state old age pension starts at the age of 60 for most of the period under observation. Meeting the age qualification for this source of income also influences the participation choice of individuals. Therefore, it is essential to include age and age squared when formally modeling participation. This should insulate the analysis of earnings to some degree from the problems mentioned. All demographics are reported for male and female separately since previous work has shown the differential treatment of the two groups in the labour market.

The first section of Table 4.1 reports demographics of working age men and women. In 1993, the mean age of men between 15 and 64 years old was 40.21 years. For women, that mean was slightly higher at 43.67 years. There is a small upward trend so that the mean age in 2012 for men of working age was 40.96 years and for women was 43.49 years. Therefore, the mean age of women is consistently higher than the mean age of men. Additionally, there are consistently more African women of working age over the period from 1993 to 2012, and the share of African women increased from 82.2% to 86.6% over this period. The share of African men also increased from 67.3% to 75.1%. At the same time, the share of white men and women decreased between 1993 and 2012, from 20.3% to 13.3% and from 6.7% to 5.9% respectively. Finally, the pattern for the share of Coloured individuals changed in a similar way. In 1993, 9.1% of working men and 9.8% of working women were Couloured. In 2012, these shares decreased to 8.3% and 6.2% respectively. The share of Indian and Asian individuals of working age stayed relatively stable between 1993 and 2012, around 3.3% for men and 1.3% for women.

Across the period under observation, men tend to report slightly more years of education than women and the mean years of education obtained increased for both genders. In 1993, men reported an average of 7.73 years of education whereas women reported an average of 5.51 years. In 2012, mean years of education increased to 9.85 years of mean education for men and 8.58 years of mean education for women. The increase in mean education is at least partly driven by the fact that there are fewer individuals with no schooling in 2012 and the fact that more working age men and women have obtained education beyond the matric, the South African certificate obtained after the final year of high school (Grade 12). Additionally, the difference in average education between men and women decreased for the period under observation. The democratic government of South Africa set out to decrease educational barriers to women and girls (Chisholm and September, 2005). Chisholm and September (2005) find that by 2001, enrollment in education was at least equal for girls and boys with a slight favour for girls. Additionally, insights from the Gender Series on educa-

tion by Statistics South Africa also report relevant changes between 2001 and 2011. The study reveals that women have a stronger tendency in attaining higher levels of education than men for all age groups (Statistics South Africa, 2015). This might explain how women have caught up on average years of education presented in Table 4.1.

Table 4.1: Individual Characteristics

	1993		1995		2001		2003		2010		2012	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
Mean Age	40.21	43.67	41.22	44.22	41.21	42.89	40.82	42.93	41.04	43.16	40.96	43.49
African	67.3%	82.2%	65.0%	82.3%	70.4%	85.5%	71.7%	85.9%	74.1%	86.6%	75.1%	86.6%
Coloured	9.1%	9.8%	9.9%	7.3%	9.6%	6.3%	9.1%	6.5%	8.7%	6.2%	8.3%	6.2%
Indian/Asian	3.3%	1.3%	3.5%	1.5%	3.2%	1.4%	3.2%	1.3%	3.3%	1.2%	3.2%	1.3%
White	20.3%	6.7%	21.6%	8.9%	16.7%	6.8%	16.0%	6.2%	13.9%	5.9%	13.3%	5.9%
Mean Years of Education	7.73	5.51	8.18	6.47	8.42	6.99	8.80	7.42	9.60	8.18	9.85	8.58
Married	71.5%	28.7%	84.5%	35.0%	74.0%	25.5%	71.1%	24.8%	66.7%	21.4%	65.8%	22.2%
Widowed	2.8%	32.7%	1.7%	28.5%	1.9%	24.7%	2.1%	23.7%	2.3%	20.7%	2.3%	19.1%
Divorced	16.7%	27.2%	1.9%	12.9%	3.0%	11.7%	3.1%	11.6%	2.7%	9.7%	2.9%	9.2%
Never Married	9.0%	11.4%	11.8%	23.6%	21.1%	38.0%	23.7%	39.9%	28.3%	48.2%	29.0%	49.5%
Sector												
Agriculture	15.5%	7.5%	19.2%	5.9%	11.4%	6.0%	12.5%	6.8%	6.2%	3.3%	6.1%	4.5%
Mining	17.4%	0.9%	8.9%	0.5%	9.7%	0.4%	8.6%	0.5%	4.5%	0.7%	4.4%	0.9%
Manufacturing	9.6%	9.2%	16.0%	11.6%	16.3%	10.6%	16.2%	10.2%	15.6%	8.9%	14.9%	7.4%
Utilities	5.6%	0.0%	1.5%	0.2%	1.4%	0.4%	1.2%	0.4%	1.1%	0.2%	1.2%	0.2%
Construction	6.6%	1.2%	6.7%	0.7%	8.8%	1.5%	7.9%	1.6%	12.1%	1.9%	11.4%	2.2%
Trade	8.2%	11.5%	14.9%	19.2%	17.7%	23.8%	17.5%	22.8%	18.6%	23.4%	19.4%	22.0%
Transport	12.1%	1.3%	7.2%	1.3%	7.7%	1.9%	7.0%	1.7%	8.5%	2.2%	8.4%	2.5%
Finance	2.4%	2.1%	5.7%	5.1%	9.1%	5.2%	9.7%	5.4%	13.4%	8.1%	13.3%	8.4%
Services	21.1%	22.8%	18.6%	30.6%	15.2%	20.9%	16.4%	22.3%	16.6%	26.5%	17.5%	29.4%
Domestic Services	1.4%	43.5%	1.4%	24.9%	2.7%	29.4%	2.9%	28.3%	3.4%	24.8%	3.4%	22.5%
Occupation												
Manager	4.7%	3.2%	8.5%	3.5%	9.4%	2.2%	10.7%	3.3%	11.8%	4.1%	11.4%	4.8%
Professionals	10.9%	11.3%	3.9%	2.7%	4.3%	3.7%	4.6%	3.7%	6.4%	4.9%	6.6%	4.1%
Clerk	6.6%	12.2%	6.2%	11.7%	5.3%	8.3%	5.3%	8.5%	5.6%	10.6%	5.0%	11.4%
Services	21.1%	22.8%	18.6%	30.6%	15.2%	20.9%	16.4%	22.3%	16.6%	26.5%	17.5%	29.4%
Agriculture	5.0%	1.9%	2.0%	0.5%	5.5%	2.2%	2.4%	2.6%	1.3%	0.7%	0.8%	0.4%
Artisan	9.3%	0.3%	17.6%	5.0%	20.4%	6.0%	18.9%	5.4%	18.7%	3.1%	18.3%	2.8%
Machine Operator	18.7%	6.2%	16.3%	3.4%	16.6%	2.8%	16.3%	3.0%	13.0%	2.3%	13.2%	2.2%
Elementary	23.1%	19.4%	24.8%	14.5%	17.9%	22.8%	21.4%	23.6%	20.6%	24.4%	20.8%	26.0%
Province												
Western Cape	12.4%	9.3%	12.3%	9.1%	12.5%	8.2%	12.3%	8.4%	11.8%	8.7%	12.2%	9.5%
Eastern Cape	1.0%	2.9%	10.1%	19.8%	9.8%	15.7%	10.2%	15.1%	9.8%	14.1%	9.6%	13.8%
Northern Cape	20.6%	10.7%	3.0%	2.0%	2.6%	1.9%	2.5%	1.7%	2.0%	2.2%	2.0%	2.2%
Free State	4.8%	5.9%	8.2%	6.5%	7.8%	5.7%	7.1%	5.3%	5.8%	5.4%	6.0%	5.6%
KZN	22.5%	19.9%	17.5%	20.3%	17.2%	22.5%	17.8%	22.3%	16.0%	20.2%	15.8%	19.9%
North West	8.6%	7.3%	7.0%	6.4%	7.1%	5.9%	6.7%	6.1%	7.3%	7.2%	6.5%	7.1%
Gauteng	14.8%	12.7%	26.6%	17.5%	28.3%	19.9%	29.3%	20.1%	31.9%	21.1%	32.1%	22.4%
Mpumalanga	6.4%	6.4%	7.3%	5.0%	6.9%	7.3%	6.9%	7.8%	7.2%	8.9%	7.4%	8.0%
Limpopo	8.8%	24.8%	8.1%	13.5%	7.7%	12.9%	7.2%	13.2%	8.2%	12.1%	8.5%	11.5%

Source: Author's calculations based on weighted PALMS Data (1993 - 2012)

All statistics presented for working age individuals.

Furthermore, there has been a significant decrease in the proportion of married men and women between 1993 and 2012. In 1993, 71.5% of working age men and 28.7% of working age women were married respectively. In 2012, that share decreased to 65.8% for men and 22.2% for women. At the same time, the share of individuals that reported they have never been married increased from 9% to 29% for men and from 11.4% to 49.5% for women.

The remainder of Table 4.1 shows the allocation of men and women into different sec-

tors, occupations and the distribution across provinces in South Africa. There have been significant shifts in the allocation of individuals across sectors. The three most important sectors for men in 1993 were services, mining and agriculture with 21.1%, 17.4% and 15.5% working in these sectors respectively. Domestic services was by far the most important sector for working women at the time with 43.5%, followed by services with 22.8% and trade with 11.5% of females working in these sectors. In 2012, trade became the most important sector for men with 19.4% working in that sector compared to 22% of women. Services is the second most common for men with 17.5% working there and the most common sector for women with a share of 29.4%. A significant change happened for women working in the domestic service sector. In 1993, it was the most important sector for female workers. In 2012, it is only the second most important sector with 22.5% of women working there. In the past, mining and agriculture were the two most important sectors in the South African economy. Inwardly focused industrialisation and rich resources of diamonds and gold made mining a historically important sector. Therefore, it is not surprising, that it is still highly relevant at the end of the apartheid regime. With the lifting of international trade bans, the South African economy went through an output expansion in the services sector as well as in international trade (Bhorat and Hodge, 1999). Since the end of apartheid, South Africa went through deindustrialisation and the economy moved away from a dependence on mining and other primary sectors towards the service industry (Bhorat et al., 2018). Nonetheless, there is still a majority of individuals working in labour-intensive sectors that require low skills.

There have also been significant changes in the occupational structure between 1993 and 2012. As such, the share of women working in elementary occupations increased from 19.4% to 26%. It is important to note that the category of elementary occupation includes domestic workers. In most surveys, domestic workers were allocated in a separate category. However, the 1993 PSLSD did not differentiate between elementary occupations and domestic workers which is why the categories had to be recoded to be consistent across periods. Other important occupations for women are service workers who reported 29.4% in 2012 compared to 22.8% in 1993 and clerks with 11.4% in 2012 compared to 12.2% in 1993. For men, there have been some shifts in their occupations as well. In 1993 there were 23.1% of men working in elementary occupations, 21.1% as service workers, 18.7% as machine operators and 10.9% as professionals. In 2012, 20.8% of men were employed in elementary occupations, 17.5% as service workers, 13.2% as machine operators, 6.6% as professionals and 18.3% worked in artisan and related crafts. Finally, there was a significant drop in men employed as agricultural workers. In 1993, there were about 5% of men employed in this occupation. In 2012, that percentage was reduced to less than 1%.

Lastly, Table 4.1 reports the geographical distribution of individuals across provinces. KwaZulu-

Natal (KZN) was the most populous province in 1993 with 22.5% of working age men and 19.9% of working age women living there. In 2012, these share changed to 15.8% of working age men and 19.9% of working age women. The most populous province in 2012 was Gauteng, and relatively more men and women lived there in 2012 than in 1993. In 1993, 14.8% of working age men and 12.7% of women lived in Gauteng. In 2012, these fractions increased to 32.1% and 22.4% respectively. These increases in population for Gauteng and KwaZulu-Natal seem to be driven by movers from the Northern Cape, the Free State and for women also Limpopo. The fraction of men of working age living in the Northern Cape decreased from 20.6% in 1993 to 2%, at the same time, the fraction of women decreased from 10.7% to 2.2%. Finally, Limpopo reports 8.8% of men and 24.8% of women living there in 1993 compared to 8.5% of men and 11.5% of women in 2012.

Table 4.2: Household Level Characteristics

Variable	1993	1995	2001	2003	2010	2012
Household Size	4.09	4.57	3.92	3.72	3.59	3.55
Number of Adults in HH	2.60	2.94	2.62	2.53	2.52	2.50
Number of Children in HH	1.49	1.63	1.30	1.19	1.07	1.05
Number of Women in HH	1.37	1.55	1.38	1.33	1.31	1.30
Number of Employed in HH	1.11	1.16	1.03	0.98	1.01	1.01
Number of Employed Women in HH	0.43	0.45	0.45	0.43	0.44	0.44
Urban	60.6%	55.3%	62.3%	61.4%	67.6%	68.7%

Source: Author's calculations based on weighted PALMS Data (1993 - 2012).
Demographics for households with at least one working age individual.

The discussion of Table 4.1 showed the movement of working age individuals across sectors, occupations and provinces as well as changes in their demographics over time. Table 4.2 further highlights the move towards urban areas during the period under observation by looking at household characteristics. In 1993, 60.6% of households lived in urban areas. In 2012, that share had increased to 68.7%. At the same time, the average size of households decreased from 4.09 people in a household in 1993 to 3.55 people in 2012. The number of children in a household decreased more rapidly than the number of adults. In 1993, there were on average 2.6 adults in each household and 1.49 children. In 2012, these numbers decreased to 2.5 adults and 1.05 children. Table 4.2 presents these demographics while reporting on households with at least one working age individual present.[1]

The analysis in this chapter is focused on the trends of wages over time. As such, it is interesting to see how the number of employed in general and employed women in particular changed over time. The number of employed individuals in a household decreased from

[1]The randomization process employed for the counterfactual micro-simulations performed in this chapter can select anyone into employment which is why household demographics need to be estimated for every working age person in the population.

1.11 on average in 1993 to 1.01 in 2012. Furthermore, there were 0.43 employed women per household in 1993 and in 2012 there were 0.44 on average per household.

Tables 4.1 and 4.2 discuss the demographics of individuals of working age and their households. Table 4.3 and Table 4.4 below summarize the rates of participation and employment of those individuals that entered the labour market between 1993 and 2012. The statistics are reported for male and female separately including a further distinction by population group. The existing literature has shown the persisting effects of the extreme discrimination under the apartheid regime, particularly in the labour market, which can be observed in the vast differences in participation and employment across population groups.

Table 4.3: Participation Rates of Working Age Population

	1993		1995		2001		2003		2010		2012	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
Overall	72,55%	51,05%	66,93%	49,80%	72,29%	64,17%	72,13%	63,71%	68,73%	54,87%	69,21%	56,88%
By Population group												
African	70,11%	49,52%	63,00%	48,01%	70,19%	64,48%	70,33%	64,53%	66,35%	53,76%	67,72%	56,55%
Coloured	74,17%	56,52%	75,27%	58,09%	78,64%	66,35%	76,49%	64,94%	75,22%	57,55%	72,31%	58,26%
Indian/Asian	78,77%	42,47%	78,63%	40,27%	79,50%	52,99%	77,04%	49,37%	75,85%	49,49%	72,12%	43,79%
White	82,10%	57,30%	79,51%	56,24%	78,74%	62,95%	79,13%	60,44%	79,27%	62,94%	77,70%	62,19%

Source: Author's calculations based on weighted PALMS Data (1993 - 2012)

According to Table 4.3, overall participation has increased for women but decreased for men between 1993 and 2012. Labour market participation is defined as all individuals who are economically active no matter whether they are employed or unemployed according to the expanded definition of unemployment. Female labour market participation was at 51.05% in 1993 and increased to about 64% in 2001 and 2003, only to drop to 56.88% in 2012. It is important to note that this increase in participation could in part be attributed to the increased capturing of marginal workers noted by Wittenberg (2016a). Table 4.3 indicates that the shift in these two years is mostly driven by an increase in participation by Africans who are more likely to be found in marginalized employment. Male participation is consistently higher than female participation, even though the difference between male and female participation decreased from 21 percentage points to just under 13 percentage points over the period under observation. In 1993, 72.55% of men chose to participate in the labour market. This participation rate decreased to 69.21% in 2010 and 2012.

Assessing participation rates not only across gender but also across population groups shows a significantly lower participation rate of African men compared to men from all other population groups. The participation rate of African men was at 70.11% in 1993 and there has been a drop in the participation of African males in 2010 so that participation was at 67.72% in 2012. The participation rates of men from the Coloured, Indian/Asian and White population groups are consistently higher but go through a similar pattern as those of African men. White men generally have a higher chance of participating except for 2001

where Indian/Asian men reported the highest participation rate compared to all other population groups. White male participation was the highest right at the end of the apartheid regime with 82.1% in 1993 and has decreased to 77.7% in 2012 at which point it is still highest across population groups.

Table 4.4: Employment Rates Conditional on Participation

	1993		1995		2001		2003		2010		2012	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
Overall	78.99%	65.68%	78.26%	61.84%	66.02%	53.43%	65.34%	52.96%	70.50%	64.31%	69.88%	63.17%
By Population group												
African	74.71%	59.88%	72.43%	53.66%	59.33%	46.55%	58.61%	45.79%	66.08%	58.52%	65.49%	57.58%
Coloured	75.10%	58.01%	83.28%	72.34%	73.62%	65.19%	74.02%	66.93%	73.40%	73.67%	72.35%	72.76%
Indian/Asian	93.38%	85.39%	90.21%	78.93%	81.64%	68.85%	82.47%	74.22%	89.31%	85.46%	88.75%	89.46%
White	96.31%	95.04%	96.87%	91.76%	93.55%	88.10%	93.52%	90.22%	91.63%	92.15%	93.93%	92.76%

Source: Author's calculations based on weighted PALMS Data (1993 - 2012)

Across population groups, participation trends are very different for females. The lowest participation rates by far are reported for Asian and Indian women. Their participation fluctuates between 40% and 53% over the period under observation. However, it is important to note that the sub-sample of this group is rather small which may lead to distortions in the summary statistics. Coloured women reported the highest participation rates in 1995, 2001 and 2003 at 58%, 66% and 65% respectively. From 2001 onwards, African women participated in the labour market to a larger degree and report participation rates of about 64.5% in 2001 and 2003 but their participation rates decrease to 57% in 2012. In 2010 and 2012, white women have the highest participation rates at 63% and 62% respectively.

Casale (2004) identified a "feminisation" of the labour force, a period in which female participation increased significantly. Table 4.3 shows that during Casale's period of observation between 1995 and 2001, there was an immense increase in female participation. In the extended period of observation of this chapter, there is still an increase in female participation observed, however, to a smaller degree. Over the same period of time, there has been a decrease in male participation.

Table 4.4 reports the employment rates of those individuals that chose to participate in the labour market. Employment of males is significantly higher than that of females across all years. However, employment rates decreased after they peaked at 79% in 1993. The lowest rate of male employment was in 2003 at 65% and stabilized around 70% in 2010 and 2012. Female employment was also highest in 1993 at 66%. The lowest rates of female employment were reported in 2001 and 2003 at around 53% for both years. It increased to about 63% in 2012.

When employment rates are broken down by population group, one trend becomes increas-

ingly obvious. Africans are consistently less likely to be employed, and African women tend to be the most marginalized. This is in accord with the findings by Ntuli and Wittenberg (2013). In 1993, the employment rates of white men and women were the highest at 96% and 95% respectively compared to about 75% and 60% for African men and women. Casale's (2004) findings that the increased participation particularly of African women has mainly led to an increase in unemployment becomes obvious when the employment rates of African women dropped to 47% and 46% in 2001 and 2003. Their employment rates increase to about 58% in 2010 and 2012, however, these are still significantly lower than employment rates of 73% for Coloured, 89% for Asian and Indian and 93% for White women in 2012.

Finally, Table 4.5 reports some distributional statistics of real earnings between 1993 and 2012. The PALMS data set contains an earnings variable that is adjusted for inflation.[2] This variable forms the basis of the decomposition approach pursued in this chapter. Therefore, Table 4.5 reports the Gini coefficient as a standard measure of inequality and the mean and median of earnings for men and women. The distribution of earnings can only be assessed for individuals who report earnings. Therefore, the Gini has been estimated conditional on obtaining employment and reporting earnings.

Table 4.5: Distributional Statistics of Real Earnings

	1993		1995		2001		2003		2010		2012	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
Median	1 822.44	996.73	1 728.51	1 364.26	1 412.43	753.30	1 224.49	693.88	1 821.75	1 214.13	1 863.87	1 315.79
Mean	3 387.86	1 919.08	3 089.60	1 965.42	2 450.14	1 642.20	2 132.14	1 547.75	4 080.88	2 689.00	4 243.84	3 023.62
Gini	0.591	0.583	0.563	0.511	0.550	0.583	0.542	0.593	0.609	0.582	0.617	0.600

Source: Author's calculations based on weighted PALMS Data (1993 - 2012)
All earnings values are presented in 2000 Rand.

At the end of the apartheid era, mean and median incomes were relatively low. 50% of men earned R1822 or less and the median earnings for women were even lower at R996. Average earnings were higher than this, and as such, the mean of earnings was R3388 for men and R1919 for women. The reason that the mean is higher than the median can be that there are some individuals with substantial earnings that drive up the average of earnings. The relatively large discrepancy between mean and median is a potential indicator for income inequality. However, it is a much less accurate measurement of inequality than the Gini coefficient. In 1995, the median earnings for women had improved significantly and were at R1364. At the same time, the mean of female earnings had increased as well, albeit to a much lesser degree. For men, mean and median of earnings stood at R3089 and R1729 respectively. The decrease of these moments of the distribution continued for both, men and women, until 2003. At this point, the median of earnings was at the lowest at R1224 for men and R694 for women. This means that in 2003, 50% of women earned less than

[2]The base year for the inflation adjusted earnings variable is 2000.

R700 per month. Average earnings were slightly higher at R1548 for women and R2132 for men. By 2010 the mean and median earnings had improved. Women earned a median income of R1214 and an average income of R2689. Men's median income was slightly higher at R1822 and the average earnings for men increased significantly to R4081. This increase in mean earnings without a corresponding increase in the median is the reason that the Gini coefficient increased so significantly in this period. It means that for 50% of men, earnings improved only to a certain degree while for some high earning individuals, earnings increased significantly. This development continued in 2012 were the mean of male earnings was R4244. At the same time, 50% of men reported earnings of R1864 or less. Women's median earnings also increased in 2012 and were at R1316. However, average earnings increased significantly for women and a mean of R3024 can be observed.

In 1993, the Gini coefficient of earnings was at 0.591 for men and 0.583 for women. These are relatively high levels of earnings inequality even though the distribution of earnings was slightly more equal for women than for men. In 1995, levels of earnings inequality improved slightly. The Gini coefficient for women fell to 0.51, which was its lowest at the end of apartheid. For men, the level decreased to 0.56 but improved further in the following years. The lowest Gini coefficient reported for men was in 2003 where the levels of earnings inequality were estimated at 0.54. For women, the level of inequality in earnings had worsened by then and was measured at 0.593, a level that is even higher than inequality at the end of apartheid. The global financial crisis of 2007/2008 certainly had its effects on the South African economy. This may be the reason why a significant increase in earnings inequality for men is observed by 2010. The Gini coefficients of earnings for men jumped from 0.54 to 0.61. For women, inequality actually decreased slightly, although at 0.58, levels of earnings inequality remained high. The more significant increase in earnings inequality for women followed in 2012, however. By then, women's earnings reported a Gini coefficient of 0.6 compared to a mild increase of the Gini for men at 0.62.

The underlying reasons for these changes in the earnings distributions will be analysed in the remainder of this chapter. For this, a set of counterfactual earnings distributions will be constructed and the contribution of changes in different components to the changes in overall earnings inequality will be estimated. The various components that determine the earnings distribution will be discussed in Section 4.4.1. Section 4.4.2 then describes the different steps of the applied methodology for which the results are reported in Section 4.5.

4.4 Methodology

The methodology in this chapter draws from a micro-simulation approach developed by Bourguignon et al. (2008). Due to the distortion effects of high levels of unemployment,

this chapter adapts a modification to this approach first introduced by González-Rozada and Menendez (2006). The details of this methodology are discussed in this section.

In order to study the differences in wage distributions over time, this chapter analyses the effect of changes in the participation rate, employment rate, occupational structure, observed characteristics, the returns to these observed characteristics and the change in unobservables on the wage distribution. Through micro-simulations, each argument will be modified to assess how changes in each factor affect changes in overall labour income. The impact of each argument will be estimated by comparing measurements of inequality of initial labour income with inequality measurements after each argument has been changed.

Following González-Rozada and Menendez (2006), the distribution of labour income YL in period t can be represented as

$$YL_t = f(P_t, E_t, O_t, X_t, R_t, \varepsilon_t), \tag{4.1}$$

which is a function f of P_t and E_t, the participation and employment rates at period t, the occupational structure O_t, a vector of observed characteristics X_t, the returns to these observed characteristics R_t and an error term ε_t. At each step of the estimation, one of the actual arguments of this function will be replaced with a counterfactual one. The functional forms of each of the different arguments underlying labour market income will be discussed separately and in detail below. The function of labour market income is modified in this way to attempt to capture what the distribution of YL would have been given the change in each argument. Because this chapter is estimating changes over time, each actual argument of time t will be replaced with a counterfactual argument of time $t+l$ for $l \neq 0$. As such, a first counterfactual distribution of labour income would be $YL_t^* = f(P_{t+l}, E_t, O_t, X_t, R_t, \varepsilon_t)$ where the actual participation rate in time t has been replaced with the participation rate observed in period $t + l$. By comparing inequality measurements of the labour income distribution YL_t and the counterfactual distribution YL_t^*, the effect of the change in participation rate between period t and period $t + l$ can be estimated. By changing one argument in YL_t after another, the effect of changes in the labour market conditions on inequality can be measured between two given years. Each step of the estimation will be done with period t as the baseline and replacing each argument with counterfactual levels from $t + l$ and vice versa, such that average effects can be estimated.

4.4.1 Simulating Labour Market Outcomes

As outlined above, the technique applied in this chapter relies on changing each component that contributes to the determination of wages one step at a time. For this to be possible, these components need to be estimated for each year under observation. This section

presents the methodology and results of the estimations for each component before Section 4.4.2 reviews the individual steps of the counterfactual simulations.

Estimating Individual Labour Market Outcomes

The first step for an individual to earn a wage is the decision to participate in the labour market followed by the hiring decision of the employer. Mohanty (2010) summarized this as the "seek and hire theory". This chapter models this two-step decision-making process as a probit model with sample selection. It is necessary to correct for sample selection because the hiring decision of the employer will only be observed for individuals who selected into labour market participation.

Labour market participation is presented by an indicator P_i that equals one if the individual's job preference function y_i^P is greater than zero and zero otherwise. The second decision is presented by E_i, another binary indicator that equals one if employer's preference function y_i^E is greater than zero and zero otherwise. This bivariate employment model is based on

$$y_i^P = X_i^P \beta^P + \varepsilon_i^P \tag{4.2}$$

$$y_i^E = X_i^E \beta^E + \varepsilon_i^E \tag{4.3}$$

where X_i^P present the characteristics that determine an individual's participation decision and X_i^E are those characteristics that determine the employment decision. The coefficients β^P and β^E report the effect of characteristics X_i^E and X_i^P on the probability of participation and the probability of employment respectively. The binary labour market outcome equations are then represented by

$$Prob(P_i = 1) = \Phi(X_i^P \beta^P + \varepsilon_i^P) \tag{4.4}$$

$$Prob(E_i = 1) = \Phi(X_i^E \beta^E + \varepsilon_i^E), \tag{4.5}$$

where $\Phi(\cdot)$ are the probability functions for participation and employment and the errors of the participation decision (ε_i^P) and the errors of the latent employment offer (ε_i^E) are jointly normally distributed, as such

$$\begin{pmatrix} \varepsilon_i^P \\ \varepsilon_i^E \end{pmatrix} \sim N \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right\}. \tag{4.6}$$

However, the employment offer will only be observed if an individual chose to participate. Hence, $E(E_i|P_i = 1, X_i^E)$. There is a chance that those who did not choose to participate are less likely to have a positive employment outcome. Thus, any estimation of the prob-

ability of employment will be biased. Heckman (1979) identified this sample selection bias and recommended a correction term that addresses what is essentially an omitted variable problem. While Heckman (1979) developed a method to correct for sample selection when estimating wages, the procedure can be applied to probit estimations as well (see Van de Ven and Van Praag, 1981 or Greene, 2002). Therefore, if it can be assumed that the choice to participate depends on X_i^P and ε_i^P and not upon X_i^E, then

$$\begin{aligned} &E(E_i|P_i = 1, X_i^E) = E(E_i|X_i^E X_i^P \varepsilon_i^P) = X_i^E \beta^E + E(\varepsilon_i^E|X_i^E X_i^P \varepsilon_i^P) \\ &\text{becomes} \\ &E(E_i|P_i = 1, X_i^E) = X_i^E \beta^E + E(E_i|P_i = 1) = X_i^E \beta^E + E(\varepsilon_i^E|\varepsilon_i^P > -X_i^P \beta_i^P). \end{aligned} \tag{4.7}$$

In this instance, $(\varepsilon_i^E|\varepsilon_i^P > -X_i^P \beta_i^P)$ is essentially the 'omitted variable'. Heckman's contribution was that this omitted variable can be modeled using the inverse Mill's ratio and this chapter will follow the Heckman modification in estimating a bivariate probit regression with selection bias. Using the employment status provided in the PALMS data, individuals that are employed or unemployed are classified as participating, any individual that is out of the labour force is classified as non-participating. Table 4.6 reports the results of the joint estimation of the probability of being employed conditional on the probability of participation including the correction for sample selection bias.

The results presented in Table 4.6 are those of a bivariate probit in which the selection equation is the participation choice and the selection-corrected estimation is the employment status. The Heckman model (1979) requires so-called exclusion restrictions, which in this case are variables that affect participation but not the substantive problem of interest, the employment status. For this reason, the number of adults and the number of children in a household are included in the participation estimation but not in the employment status. Number of adults and number of children in a household are associated with socio-economic standing of an individual and are theorized to influence the participation decision. In most literature that accounts for the selection process, employment is modeled as a binary 'choice' of being employed or unemployed with the underlying assumption being that individuals may choose unemployment voluntarily (Ntuli, 2007; Chamberlain and van der Berg, 2002). In the context of high and structural unemployment prevalent in South Africa, this assumption simply does not hold (Kingdon and Knight, 2000). Instead, the participation choice of an individual may or may not result in employment. To account for these constraints in the labour market, this analysis assumes a demand-driven model of the labour market. Therefore, the number of adults and the number of children is not theorized to have an effect on the employment decision made by the employer. It is important to note, however, that this exclusion restriction has its limitations, as it may not hold in a supply-side model where it can be argued that the number of adults in the household can influence the likelihood of

Table 4.6: Bivariate Probit Results with Selection Correction

	1993		1995		2001		2003		2010		2012	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
Employed												
Age	-0.041	-0.043	-0.019**	-0.030**	0.041	0.134**	0.047**	0.020	0.013	0.072**	-0.011	-0.008
	(0.03)	(0.03)	(0.01)	(0.01)	(0.03)	(0.06)	(0.02)	(0.02)	(0.01)	(0.03)	(0.01)	(0.02)
Age2	0.001**	0.001**	0.000**	0.001***	-0.000	-0.001	-0.000	0.000	0.000	-0.000	0.000*	0.001**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Years of Education	-0.030	-0.042	-0.053***	-0.056***	-0.076***	-0.076***	-0.077***	-0.090***	-0.108***	-0.116***	-0.105***	-0.104***
	(0.02)	(0.03)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
(Years of Education)2	0.004**	0.005**	0.006***	0.007***	0.007***	0.008***	0.007***	0.008***	0.009***	0.010***	0.008***	0.009***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Married	0.493***	0.288***	0.807***	0.171***	0.732***	-0.044	0.696***	0.069**	0.452***	0.104**	0.533***	0.156***
	(0.08)	(0.07)	(0.03)	(0.02)	(0.04)	(0.05)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)
Western Cape	0.407***	0.224**	0.096*	0.010	0.204***	0.206***	0.146**	0.237***	0.137**	0.314***	0.108**	0.293***
	(0.12)	(0.11)	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)	(0.06)	(0.06)	(0.05)	(0.05)
Eastern Cape	-0.670***	-0.127	-0.445***	-0.231***	-0.111**	0.047	-0.110**	0.005	-0.050	0.159**	0.015	0.034
	(0.08)	(0.08)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)
Northern Cape	0.520**	-0.406	0.010	-0.187**	0.026	-0.056	0.032	-0.106*	0.049	0.004	0.102*	0.082
	(0.25)	(0.27)	(0.06)	(0.06)	(0.06)	(0.07)	(0.06)	(0.06)	(0.07)	(0.06)	(0.06)	(0.06)
Free State	0.071	0.139*	0.055	0.074*	0.067	0.014	0.046	-0.051	0.075	0.127**	0.034	-0.055
	(0.08)	(0.08)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)
KwaZulu-Natal	-0.433***	0.061	-0.189***	-0.071*	-0.180***	0.017	-0.052	0.093**	0.228***	0.372***	0.311***	0.288***
	(0.07)	(0.08)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.05)
North West	-0.107	0.097	-0.094**	-0.099**	0.018	-0.188***	-0.047	-0.110**	0.079	0.118**	0.069	-0.061
	(0.09)	(0.09)	(0.05)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.05)	(0.06)	(0.05)	(0.05)
Gauteng	*omitted*											
Mpumalanga	-0.071	0.159*	-0.133**	-0.146***	0.034	0.084*	0.053	0.071	0.044	0.153**	0.148**	-0.022
	(0.09)	(0.09)	(0.05)	(0.04)	(0.05)	(0.05)	(0.04)	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)
Limpopo	-0.575***	0.007	-0.359***	-0.275***	-0.238***	-0.191***	-0.358***	-0.183***	0.163**	0.333***	0.210***	0.142**
	(0.10)	(0.09)	(0.05)	(0.05)	(0.05)	(0.04)	(0.05)	(0.04)	(0.06)	(0.06)	(0.05)	(0.05)
African/Black	-0.895***	-0.985***	-0.900***	-0.941***	-0.930***	-0.990***	-0.938***	-1.157***	-0.538***	-0.953***	-0.723***	-0.877***
	(0.11)	(0.13)	(0.05)	(0.05)	(0.06)	(0.10)	(0.06)	(0.05)	(0.07)	(0.07)	(0.07)	(0.07)
Coloured	-1.249***	-0.765**	-0.609***	-0.429***	-0.730***	-0.552***	-0.642***	-0.651***	-0.446***	-0.588***	-0.656***	-0.602***
	(0.23)	(0.24)	(0.06)	(0.05)	(0.07)	(0.10)	(0.07)	(0.07)	(0.08)	(0.08)	(0.08)	(0.08)
Indian/Asian	0.079	-0.206	-0.341***	-0.147*	-0.406***	-0.528***	-0.390***	-0.430***	-0.084	-0.301**	-0.265**	-0.110
	(0.16)	(0.17)	(0.08)	(0.08)	(0.09)	(0.10)	(0.09)	(0.10)	(0.11)	(0.14)	(0.13)	(0.14)
White	*omitted*											
urban	-0.357***	0.106	-0.189***	-0.132***	-0.037	0.054*	-0.124***	0.030	0.137***	0.199***	0.243***	0.168***
	(0.07)	(0.08)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.04)
_cons	2.228***	1.617**	1.767***	1.418***	0.166	-2.346*	0.101	0.100	0.456	-0.931	0.962**	0.646*
	(0.46)	(0.62)	(0.18)	(0.27)	(0.56)	(1.25)	(0.43)	(0.43)	(0.32)	(0.80)	(0.30)	(0.39)
Participation												
Age	0.365***	0.272***	0.365***	0.289***	0.347***	0.316***	0.343***	0.321***	0.338***	0.285***	0.367***	0.303***
	(0.01)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
Age2	-0.005***	-0.003***	-0.005***	-0.004***	-0.004***	-0.004***	-0.004***	-0.004***	-0.004***	-0.004***	-0.005***	-0.004***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Years of Education	-0.040	-0.019	-0.038***	-0.053***	-0.027**	-0.022**	-0.024**	-0.007	0.008	-0.039***	-0.005	-0.038**
	(0.03)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
(Years of Education)2	0.001	0.005**	0.003***	0.006***	0.004***	0.006***	0.005***	0.006***	0.004***	0.009***	0.004***	0.008***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Married	0.540***	-0.379***	0.612***	-0.291***	0.484***	-0.275***	0.492***	-0.306***	0.550***	-0.296***	0.623***	-0.273***
	(0.08)	(0.06)	(0.03)	(0.02)	(0.04)	(0.02)	(0.03)	(0.02)	(0.04)	(0.02)	(0.04)	(0.02)
Western Cape	0.112	-0.205**	0.032	-0.094**	-0.021	-0.144***	0.038	-0.128**	-0.044	-0.012	-0.039	-0.081*
	(0.10)	(0.08)	(0.05)	(0.03)	(0.05)	(0.04)	(0.05)	(0.04)	(0.06)	(0.05)	(0.05)	(0.04)
Eastern Cape	-0.392***	-0.393***	-0.348***	-0.329***	-0.150***	-0.207***	-0.168***	-0.237***	-0.281***	-0.214***	-0.209***	-0.185***
	(0.08)	(0.06)	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)	(0.03)	(0.05)	(0.04)	(0.05)	(0.04)
Northern Cape	0.021	-0.528**	-0.120**	-0.280***	-0.130**	-0.198***	-0.087	-0.194***	-0.276***	-0.134**	-0.158**	-0.114**
	(0.18)	(0.20)	(0.06)	(0.04)	(0.06)	(0.05)	(0.06)	(0.05)	(0.06)	(0.05)	(0.06)	(0.05)
Free State	-0.280***	-0.140**	-0.054	-0.023	-0.123**	-0.075*	-0.087*	-0.001	-0.172***	-0.108**	-0.261***	-0.133**
	(0.08)	(0.06)	(0.04)	(0.03)	(0.04)	(0.04)	(0.05)	(0.04)	(0.05)	(0.04)	(0.05)	(0.04)
KwaZulu-Natal	-0.060	0.015	-0.141***	-0.229***	-0.116**	-0.164***	-0.101**	-0.094**	-0.291***	-0.226***	-0.257***	-0.162***
	(0.08)	(0.06)	(0.04)	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)
North West	-0.079	-0.114	-0.053	-0.249***	-0.079*	-0.194***	-0.095**	-0.210***	-0.142**	-0.290***	-0.102*	-0.182***
	(0.09)	(0.07)	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.05)	(0.04)
Gauteng	*omitted*											
Mpumalanga	-0.029	-0.069	-0.080**	-0.244***	-0.158***	-0.101**	-0.093**	0.020	-0.030	0.034	0.199***	0.198***
	(0.08)	(0.07)	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.05)	(0.04)
Limpopo	-0.486***	-0.316***	-0.520***	-0.452***	-0.286***	-0.062	-0.301***	-0.050	-0.344***	-0.317***	-0.228***	-0.180***
	(0.09)	(0.07)	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.05)	(0.04)
African/Black	-0.112	0.219**	-0.128***	0.224***	0.120**	0.450***	0.193***	0.552***	0.100*	0.302***	0.125**	0.342***
	(0.09)	(0.08)	(0.04)	(0.03)	(0.05)	(0.04)	(0.04)	(0.04)	(0.06)	(0.05)	(0.06)	(0.05)
Coloured	-0.272	0.441**	0.126**	0.311***	0.259***	0.405***	0.214***	0.504***	0.169**	0.181**	0.127*	0.283***
	(0.18)	(0.17)	(0.05)	(0.03)	(0.06)	(0.05)	(0.06)	(0.05)	(0.07)	(0.06)	(0.07)	(0.05)
Indian/Asian	-0.088	-0.446***	0.234***	-0.253***	0.200**	-0.166**	0.093	-0.238***	0.025	-0.180**	-0.130	-0.380***
	(0.16)	(0.11)	(0.06)	(0.05)	(0.09)	(0.07)	(0.08)	(0.07)	(0.10)	(0.08)	(0.10)	(0.08)
White	*omitted*											
urban	-0.017	0.160**	-0.058**	0.175***	-0.079**	0.114***	-0.053**	0.058**	0.183***	0.204***	0.210***	0.192***
	(0.07)	(0.06)	(0.02)	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
No. of Adults	-0.099***	-0.035**	-0.080***	-0.047***	-0.036***	-0.018	-0.035***	-0.039***	-0.058***	0.003	-0.068***	-0.015**
	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
No. of Children	-0.022	-0.061***	-0.036***	-0.031***	-0.026**	-0.013	-0.045***	-0.042***	-0.032***	-0.053***	-0.028**	-0.063***
	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Constant	-4.599***	-4.394***	-5.118***	-4.635***	-5.180***	-5.323***	-5.244***	-5.468***	-5.543***	-5.409***	-5.985***	-5.664***
	(0.23)	(0.22)	(0.09)	(0.08)	(0.13)	(0.15)	(0.13)	(0.11)	(0.14)	(0.12)	(0.14)	(0.12)
Rho	-0.818***	-0.868***	-0.757***	-0.741***	-0.285	0.108	-0.443**	-0.533***	-0.619***	-0.296	-0.752***	-0.672***
	(0.15)	(0.24)	(0.06)	(0.09)	(0.18)	(0.37)	(0.14)	(0.12)	(0.10)	(0.20)	(0.10)	(0.12)
No. of Obs.	10 238	12 414	36 583	42 779	30 601	34 684	29 265	32 709	24 488	29 257	23 976	28 251

Source: Author's calculations based on weighted PALMS Data (1993 - 2012)

finding employment or that the number of children may influence the reservation wage.

Starting with the participation choice, age has a positive effect on the probability of an individual to participate. Age-squared has a negative effect which indicates diminishing marginal expected returns on age. Increased years of education have a small but negative impact on the probability of participation and is jointly significant with the square of years of education across most years. The coefficients on years of education squared are small and positive which indicates that only above a certain threshold are returns to education positive. In 1995, the first year where years of education and its square are jointly significant, this turning point is at 12.77 years of education for men and at 8.93 years of education for women. Being married has continuously positive effects of participating for men. For women, the effect is reversed. This may indicate the traditional gender roles in which married women are more likely to stay at home and choose not to participate. However, marital status has a different effect on the employment outcome. For men, being married increases the chance of employment significantly across years. Contrary to participation choice, married women are also more likely to be employed even though the coefficients are smaller than those for men.

The discussion of descriptive statistics in Section 4.3.1 has shown that Gauteng is the province with the largest proportion of working age adults in South Africa for most of the period under observation. In Table 4.6, Gauteng forms the baseline against which the probability of participating is compared for all other provinces. Holding everything else constant, the coefficients for the participation equation are negative for most provinces. This means that relative to Gauteng, being in another province results in a lower probability of participation. There are few exceptions with provinces reporting positive coefficients, however, most of those coefficients are insignificant and hardly different from zero. A notable exception is Mpumalanga. For this province, the relative probability of participating was positive and significant for both, men and women, in 2012. Additionally, men are less likely to participate if they stay in urban areas than their rural counterparts up until 2003. For the years 2003, 2010 and 2012, men in urban areas are more likely to participate in the labour market. For all years under observation, women in urban areas are more likely to participate than in rural areas. This is not surprising as individuals with the intention to enter the labour market might choose urban areas in the hopes of increased chances of finding a job.

As far as population groups are concerned, Table 4.6 shows that Coloured women are more likely to participate than White women *ceteris paribus*. This is unsurprising seeing the relatively high participation rates for Coloured women described in Table 4.3. Coloured men are also more likely to participate, holding all else constant, except in 1993. African

men and women both report mostly positive coefficients, except for the early years after apartheid, where the probability of participation for African men was negative compared to their White counterparts. It is important to point out that African women are significantly more likely to participate relative to White women. Lastly, Asian and Indian men and women tend to have a lower probability of participating relative to White men and women. This is particularly true for women of this population group.

Finally, both, number of adults and the number of children, have a negative impact on the probability of participating for both, men and women. Across all years of observation, the coefficients are small but significant. It makes sense that the coefficient for number of children is negative. Children are defined as under 15 years old and child care may lead an individual not to participate in the labour market. Furthermore, the fact that the coefficient for number of adults in a household is negative suggests that with an increased household size, the need to participate in the labour market might be lower for each adult because of other employed adults in the household.

The reasons for choosing a bivariate probit estimation that corrects for selection bias following Heckman (1979) have been explained above. The tetrachoric correlation coefficients (rho) show that for most of the years under observations the participation choice and the employment decision are highly dependent processes. Rho reports coefficients between -0.28 and -0.86 for all years except for 2001 where the correlation between error terms of the participation and the employment equation is small and statistically insignificant. Nonetheless, the fact that the coefficients are significant across most years implies the necessity for the Heckman correction. A simple probit estimation of the employment outcome would result in inconsistent estimates. The fact that rho is negative and significant in most years reflects the low absorption rate of the South African labour market. The discussion of the descriptive statistics of employment rates and participation rates suggested that higher participation rates do not necessarily lead to higher employment rates in South Africa. The negative rho supports this finding. Characteristics that lead to the choice to participate in the labour market do not necessarily translate into a successful outcome in the employment probit.

Moving to the employment equation, age and age square are not jointly significant in determining the employment outcome as many years report insignificant coefficients. Age squared is statistically significant but close to zero for most years. Including a square for years of education as well shows, that education itself has a small but significant and negative effect on the probability of employment. However, the square of years of education is positive indicating that only after a certain turning point, higher levels of education lead to a higher probability of employment. For example, in 2012 where the coefficients of years of education and its square are jointly significant for both, men and women, the turning point

is at 13 years of education for men and at 12 years of education for women. Therefore, only once an individual has obtained at least a matric, the returns to education are positive. Marital status is another important factor for a positive employment outcome. Other than with the participation choice, being married increases the probability of being employed for both, men and women. However, the coefficient is significantly smaller for women.

When estimating the probability of participation, most provinces had lower probabilities for an individual to participate relative to Gauteng. With regards to the probability of employment, this relationship seems to be reversed. Many provincial dummies report small but positive signs indicating that the probability of being hired is slightly higher outside of Gauteng, this is true particularly from 2001 onwards. A notable exception is Limpopo which reports significant and lower probabilities of employment compared to Gauteng. It seems that for most years, the higher participation probability in Gauteng results in relatively lower chances of finding employment.

In 1993 individuals in urban areas were less likely to be employed than individuals in rural areas. The coefficients for the urban dummy are small but significant and negative until 2003. From 2010 onwards, the probability of being employed was higher in urban areas than in rural areas for both, men and women.

Regarding population groups, individuals from all population groups have a lower probability of being hired compared to white individuals. For most years, the probability to be employed is lowest compared to white men and women for Africans. The only exception are Coloured men which had an even smaller probability of employment relative to White men. Over the period of observation, this relationship has improved slightly and particularly for African men, the probability of being employed is slightly higher than for African women by 2012 relative to their white counterparts. The coefficients just discussed are used in the counterfactual estimations for participation and employment explained in Section 4.4.2.

Estimating Individual Wages

Table 4.7 reports the results of a Mincer type regression of log wages (Mincer, 1974) for all years under observation. The wage regression is represented by

$$\ln\ w = \alpha + \beta_1 \cdot age + \beta_2 \cdot age^2 + \gamma_1 \cdot yrsedu + \gamma_2 \cdot yrsedu^2 + \delta X + \lambda + \varepsilon, \tag{4.8}$$

where $\ln\ w$ is the log of real earnings, α is a constant and age, age squared, the years of education an individual has obtained ($yrsedu$) and the square of ($yrsedu^2$) as well as the vector X are a set of individual characteristics that can be assumed to influence the wage function. λ is the inverse Mill's ratio derived from the bivariate probit estimation of partici-

pation and employment to reflect selection into earnings. It is important to note that while the estimation corrects for selection bias, the possibility for endogeneity through which some of the chosen covariates may bias the estimation remains. However, data constraints do not allow for further controls. This should be taken into account when interpreting the results. The coefficients estimated in this step, $\alpha, \beta_1, \beta_2, \gamma_1, \gamma_2$ and the set of coefficients associated with the characteristics in vector X, δ, as well as λ and the regression residual ε are used in the counterfactual wage estimations in Section 4.4.2. All regressions are estimated using a cluster-robust variance estimator. The results of the regression of log wages are presented in Table 4.7 for all years.

Table 4.7 shows that, for men, earnings are weakly increasing with age. Age and age squared are jointly significant in most years but very close to zero, age squared shows small diminishing returns for age for men. For women, the relationship between age and earnings is statistically insignificant in 2010.

Schooling measured in years of education has a positive relationship with earnings up until 2003. In 2010 and 2012, it is small but negative. Years of education and the square of this variable are jointly significant for most of the period under observation. However, the square of years of education is zero for most years and in 2012, the significant coefficients show that for women, the turning point at which more years of education have a positive effect on wages is very low.

Provinces have a jointly significant effect for most years. The effect for Gauteng is consistently more positive for women. Additionally, being in an urban area is positively associated with earnings and significant across all years. Women in urban areas report a stronger relationship with earnings than men in 1993 and then again in 2010 and 2012.

As far as occupation is concerned, the coefficients are jointly significant across all years. The strongest relationships can be observed for managers and professionals, both occupations have an increasing effect on earnings. Elementary occupations have a positive and significant effect on earnings in 1993 for men. In 1995, the effects were negative but significant and stronger for men than for women. By 2010, elementary occupation had a stronger negative effect for women. In 2012, the effect was negative for both genders and slightly stronger for men.

A variable accounting for hours worked in the last month is statistically significant but the coefficients are close to zero. The dummy variables for population group are jointly significant and belonging to any population group other than white has a decreasing effect on earnings. The coefficients are consistently higher for men than for women but have de-

Table 4.7: Results of Wage Regression (in logs)

	1993		1995		2001		2003		2010		2012	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
Age	0.14***	0.06***	0.09***	0.05***	0.07***	0.02**	0.06***	0.03***	0.03***	0.00	0.03***	0.02**
	(0.01)	(0.02)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Age2	-0.00***	-0.00**	-0.00***	-0.00***	-0.00***	-0.00**	-0.00***	-0.00***	-0.00***	-0.00	-0.00***	-0.00**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Years of Education	0.04*	0.04*	0.03***	0.03**	0.04***	0.02**	0.01*	0.03**	-0.05***	-0.04***	-0.04**	-0.05***
	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
(Years of Education)2	0.00**	0.00	0.00***	0.00***	0.00***	0.00***	0.01***	0.00***	0.01***	0.01***	0.01***	0.01***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Western Cape	-0.31**	-0.06	0.12***	0.26***	0.09**	0.14**	-0.02	0.11**	-0.01	-0.13**	0.08	0.03
	(0.11)	(0.22)	(0.03)	(0.03)	(0.04)	(0.05)	(0.04)	(0.05)	(0.04)	(0.05)	(0.05)	(0.04)
Eastern Cape	-0.57***	-0.06	0.11***	0.14***	-0.02	-0.14**	-0.19***	-0.10*	-0.03	-0.12**	-0.02	-0.05
	(0.13)	(0.24)	(0.03)	(0.04)	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)
Northern Cape	*omitted*											
Free State	-0.59***	-0.41	-0.26***	-0.27***	-0.05	-0.14**	-0.18***	-0.12**	-0.18***	-0.30***	-0.05	-0.04
	(0.12)	(0.27)	(0.03)	(0.04)	(0.05)	(0.06)	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)
KwaZulu-Natal	-0.48***	0.01	0.25***	0.33***	0.12**	0.07	0.02	0.02	-0.01	-0.12**	-0.02	-0.09**
	(0.12)	(0.25)	(0.03)	(0.04)	(0.05)	(0.05)	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
North West	-0.09	0.29	0.12***	0.20***	0.14**	0.13**	0.09**	0.05	0.26***	0.01	0.25***	0.12**
	(0.14)	(0.28)	(0.03)	(0.04)	(0.05)	(0.06)	(0.05)	(0.06)	(0.05)	(0.05)	(0.06)	(0.06)
Gauteng	-0.38***	0.19	0.29***	0.48***	0.18***	0.27***	0.09**	0.24***	0.14**	0.04	0.17***	0.09*
	(0.10)	(0.26)	(0.03)	(0.04)	(0.04)	(0.05)	(0.04)	(0.05)	(0.04)	(0.05)	(0.05)	(0.05)
Mpumalanga	-0.10	0.05	0.26***	0.41***	0.14**	-0.02	0.04	0.03	0.24***	-0.03	0.15**	0.04
	(0.13)	(0.27)	(0.03)	(0.04)	(0.05)	(0.06)	(0.05)	(0.06)	(0.05)	(0.06)	(0.06)	(0.05)
Limpopo	-0.39**	0.10	0.38***	0.54***	0.09*	0.10*	0.13**	0.12**	0.00	-0.21***	0.03	-0.04
	(0.14)	(0.26)	(0.04)	(0.05)	(0.05)	(0.06)	(0.05)	(0.06)	(0.06)	(0.06)	(0.07)	(0.06)
Urban	0.41***	0.51***	0.32***	0.30***	0.22***	0.17***	0.26***	0.23***	0.17***	0.24***	0.15***	0.21***
	(0.07)	(0.07)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Manager	1.44***	1.42***	-0.01	0.49**	0.90***	1.02***	0.60***	1.21***	0.62***	0.50***	0.38***	0.73***
	(0.14)	(0.15)	(0.07)	(0.16)	(0.06)	(0.14)	(0.08)	(0.18)	(0.11)	(0.15)	(0.11)	(0.16)
Professionals	1.25***	1.78***	-0.29***	0.41**	0.68***	0.81***	0.42***	0.99***	0.44***	0.44**	0.19*	0.57***
	(0.11)	(0.13)	(0.07)	(0.15)	(0.06)	(0.13)	(0.08)	(0.18)	(0.10)	(0.14)	(0.11)	(0.16)
Clerk	0.95***	1.07***	-0.58***	0.10	0.48***	0.58***	0.23**	0.70***	0.23**	0.22	0.05	0.23
	(0.11)	(0.12)	(0.07)	(0.15)	(0.06)	(0.13)	(0.08)	(0.18)	(0.11)	(0.14)	(0.11)	(0.16)
Services	0.91***	0.44***	-0.63***	-0.12	0.27***	0.08	0.01	0.19	-0.02	-0.18	-0.20*	-0.14
	(0.10)	(0.12)	(0.07)	(0.15)	(0.06)	(0.13)	(0.08)	(0.18)	(0.10)	(0.14)	(0.11)	(0.16)
Agriculture	*omitted*											
Artisan	0.87***	0.93**	-0.53***	-0.17	0.35***	0.09	0.05	0.15	0.12	-0.04	-0.11	-0.16
	(0.10)	(0.31)	(0.07)	(0.16)	(0.05)	(0.13)	(0.08)	(0.18)	(0.10)	(0.15)	(0.11)	(0.16)
Machine Operator	1.02***	0.99***	-0.64***	-0.03	0.39***	0.33**	0.12	0.36**	0.11	-0.10	-0.18*	-0.06
	(0.10)	(0.12)	(0.07)	(0.15)	(0.05)	(0.13)	(0.08)	(0.18)	(0.10)	(0.15)	(0.11)	(0.16)
Elementary	0.58***	0.29**	-1.04***	-0.56***	-0.03	-0.24**	-0.33***	-0.15	-0.20**	-0.36**	-0.46***	-0.37**
	(0.09)	(0.12)	(0.07)	(0.15)	(0.05)	(0.12)	(0.08)	(0.17)	(0.10)	(0.14)	(0.10)	(0.15)
Hours worked	0.00**	0.00***	0.00***	0.00***	0.00**	0.00***	0.00***	0.00***	0.00**	0.00***	0.00*	0.00***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
African/Black	-0.91***	-0.58***	-0.65***	-0.33***	-0.59***	-0.22***	-0.52***	-0.14**	-0.54***	-0.28***	-0.52***	-0.25***
	(0.07)	(0.10)	(0.02)	(0.05)	(0.03)	(0.05)	(0.03)	(0.04)	(0.04)	(0.05)	(0.04)	(0.05)
Coloured	-1.23***	-0.69***	-0.52***	-0.17***	-0.32***	-0.10**	-0.29***	-0.08**	-0.39***	-0.22***	-0.36***	-0.15**
	(0.11)	(0.17)	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)	(0.05)	(0.04)	(0.05)
Indian/Asian	-0.49***	-0.47***	-0.36***	-0.17***	-0.31***	-0.10*	-0.35***	-0.04	-0.16**	-0.01	-0.29***	-0.02
	(0.10)	(0.11)	(0.03)	(0.03)	(0.04)	(0.05)	(0.04)	(0.05)	(0.06)	(0.07)	(0.06)	(0.06)
White	*omitted*											
Inverse Mill's Ratio	1.52**	-0.23	-1.07***	-0.14	-1.53***	-2.41***	-2.18***	-2.60***	-2.10***	-1.78***	-1.16***	-2.22***
	(0.65)	(0.96)	(0.16)	(0.44)	(0.17)	(0.38)	(0.17)	(0.31)	(0.35)	(0.43)	(0.29)	(0.44)
Constant	4.37***	5.35***	7.04***	5.96***	6.60***	7.68***	7.44***	7.27***	8.22***	8.19***	8.04***	8.10***
	(0.43)	(0.65)	(0.13)	(0.26)	(0.18)	(0.39)	(0.18)	(0.30)	(0.25)	(0.39)	(0.23)	(0.32)
No. of Obs.	4178	2972	18306	12458	12912	10294	12024	9515	10719	9586	10473	9271
R-Squared	0.64	0.66	0.67	0.66	0.60	0.67	0.64	0.69	0.47	0.50	0.45	0.54

Source: Author's calculations based on weighted PALMS Data (1993 - 2012)

creased over the period under observation for both, men and women. Finally, accounting for selection bias through the inverse Mill's ratio results in mostly significant coefficients, especially in the later years. This implies that without the selection correction, the estimates would have shown a downward-bias. The coefficients presented in Table 4.7 will be used in the estimations of counterfactual wages for individuals who have been assigned a new occupational choice as described in Section 4.4.2.

Estimating Individual Occupational Allocation

Another determinant of wages is an individual's occupation. Therefore, occupational dummies are included in the wage regression presented in Table 4.7. As individuals will be selected into the counterfactual employment sample who were not employed previously, occupational allocation needs to be imputed. Therefore, the actual allocation into different occupations needs to be estimated. Previously, Table 4.1 reported the distribution across occupations for all years under observation. Table 4.8 below shows the results of the ordered logit regression for the allocation into different occupations. The ordered logit approach has been chosen following Bourguignon, Ferreira and Lustig (2005). Alternative approaches use a multinomial logit regression (see for example Inchauste et al., 2014). It is possible to use an ordered logit regression in this instance as occupations are implicitly ranked by skill level in the underlying data. The ordered logit regression estimates a probability for each individual to take on a certain occupation based on the variables discussed below.

Age is negatively correlated to the allocation into occupations and age squared is not correlated even though the variables are jointly significant. The effect of years of education is small and jointly significant with the square of years of education which is negative but small for all years under observation. The effects are more substantial for women, particularly after 1995. Furthermore, being married has a negative effect and is highly significant for men and women. In 2010 and 2012, the effect of being married is smaller for women than for men. In all other years, the effect is vice versa.

The urban indicator is negatively associated with occupational allocation and significant for most years. The coefficients are not statistically significant for women in 2001, 2003 and 2010. Provincial dummies are jointly significant and are mostly positive relative to Gauteng province. Similarly, the set of indicator variables on population groups are jointly significant. The only population group that reports a number of statistically insignificant coefficients are Asians and Indians. This may be caused by the overall small size of this population group. Again, the omitted variables represent the baseline variables for province and population group respectively.

The cut-off points reported in Table 4.8 present the estimated thresholds of the latent

Table 4.8: Ordered Logistic Regression of Occupational Choice

	1993		1995		2001		2003		2010		2012	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
Age	-0.09**	-0.10**	-0.10***	-0.08***	-0.06***	-0.07***	-0.07***	-0.08***	-0.02	0.02	-0.04**	0.02
	(0.03)	(0.03)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Age2	0.00**	0.00**	0.00***	0.00***	0.00**	0.00**	0.00**	0.00**	0.00	-0.00**	0.00*	-0.00**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Years of Education	0.26***	0.36***	0.09***	0.06**	0.11***	0.13***	0.06***	0.14***	0.13***	0.17***	0.18***	0.19***
	(0.05)	(0.05)	(0.01)	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)
(Years of Education)2	-0.04***	-0.04***	-0.02***	-0.03***	-0.02***	-0.03***	-0.02***	-0.03***	-0.02***	-0.04***	-0.03***	-0.03***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Married	-0.28**	-0.15	-0.08**	-0.09**	-0.16**	-0.19***	-0.15**	-0.20***	-0.26***	-0.14**	-0.20***	-0.16**
	(0.14)	(0.12)	(0.04)	(0.04)	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)
Urban	-0.21	-0.27**	-0.64***	-0.30***	-0.43***	-0.07	-0.24***	-0.12	-0.34***	-0.06	-0.25***	-0.15**
	(0.17)	(0.12)	(0.04)	(0.05)	(0.05)	(0.06)	(0.06)	(0.08)	(0.06)	(0.06)	(0.06)	(0.07)
Western Cape	0.54***	1.03***	0.20**	0.20**	0.19**	0.30**	0.24**	0.43***	0.26**	0.18*	0.17**	0.14
	(0.16)	(0.22)	(0.06)	(0.07)	(0.09)	(0.09)	(0.10)	(0.10)	(0.08)	(0.09)	(0.08)	(0.09)
Eastern Cape	1.06***	1.40***	-0.06	-0.20**	0.06	-0.27**	0.06	-0.11	0.15	-0.15	0.19**	-0.14
	(0.19)	(0.20)	(0.06)	(0.07)	(0.09)	(0.08)	(0.08)	(0.09)	(0.09)	(0.10)	(0.09)	(0.09)
Northern Cape	0.46	0.56	0.51***	0.35***	0.65***	0.07	0.44***	0.27*	0.22**	-0.05	0.15	0.14
	(0.34)	(0.93)	(0.08)	(0.10)	(0.10)	(0.15)	(0.12)	(0.14)	(0.10)	(0.11)	(0.11)	(0.11)
Free State	0.41**	-0.16	0.40***	0.22**	0.17**	-0.08	0.14*	0.02	0.45***	0.04	0.17**	0.16*
	(0.16)	(0.14)	(0.06)	(0.07)	(0.07)	(0.09)	(0.08)	(0.10)	(0.08)	(0.09)	(0.08)	(0.09)
KwaZulu-Natal	0.67***	0.86***	-0.04	-0.18**	0.02	-0.15*	0.07	-0.11	0.10	0.07	0.04	0.04
	(0.15)	(0.15)	(0.05)	(0.07)	(0.07)	(0.08)	(0.07)	(0.09)	(0.07)	(0.08)	(0.07)	(0.08)
North West	0.45**	-0.32**	-0.04	-0.27**	0.05	-0.22**	0.05	-0.11	0.07	-0.18*	0.23**	-0.06
	(0.18)	(0.15)	(0.07)	(0.08)	(0.07)	(0.10)	(0.08)	(0.10)	(0.08)	(0.10)	(0.09)	(0.11)
Gauteng	*omitted*											
Mpumalanga	0.19	-0.10	-0.11*	0.13	0.07	-0.09	0.10	0.08	0.13	0.13	0.18**	0.08
	(0.16)	(0.14)	(0.06)	(0.09)	(0.08)	(0.10)	(0.08)	(0.10)	(0.09)	(0.09)	(0.08)	(0.10)
Limpopo	-0.14	0.39**	-0.60***	-0.64***	-0.24**	-0.26**	0.04	-0.11	-0.06	-0.02	-0.02	-0.01
	(0.19)	(0.18)	(0.08)	(0.11)	(0.09)	(0.10)	(0.09)	(0.11)	(0.10)	(0.11)	(0.09)	(0.10)
African/Black	0.52**	1.27***	0.61***	0.56***	0.92***	1.24***	0.93***	1.13***	1.21***	1.27***	1.34***	1.20***
	(0.19)	(0.21)	(0.05)	(0.06)	(0.07)	(0.07)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
Coloured	0.81**	1.40**	0.56***	0.35***	0.90***	0.60***	0.81***	0.37***	0.95***	0.53***	0.92***	0.43***
	(0.34)	(0.49)	(0.06)	(0.06)	(0.10)	(0.09)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.09)
Indian/Asian	-0.05	0.52*	0.07	0.09	0.08	0.40**	0.16	0.23	0.10	0.05	0.32**	-0.06
	(0.22)	(0.27)	(0.08)	(0.09)	(0.15)	(0.14)	(0.14)	(0.14)	(0.14)	(0.14)	(0.15)	(0.17)
White	*omitted*											
1st cut-off	-6.31***	-6.18***	-7.12***	-9.15***	-5.27***	-8.19***	-5.85***	-8.00***	-4.41***	-5.88***	-4.43***	-5.50***
	(0.84)	(0.78)	(0.22)	(0.28)	(0.30)	(0.37)	(0.32)	(0.43)	(0.33)	(0.36)	(0.33)	(0.36)
2nd cut-off	-4.50***	-3.84***	-5.75***	-6.55***	-3.85***	-5.23***	-4.42***	-5.29***	-2.98***	-3.28***	-2.87***	-3.04***
	(0.77)	(0.67)	(0.22)	(0.26)	(0.29)	(0.33)	(0.31)	(0.39)	(0.32)	(0.35)	(0.33)	(0.36)
3rd cut-off	-3.76***	-2.28***	-5.21***	-5.04***	-3.39***	-3.88***	-3.96***	-4.03***	-2.51***	-1.99***	-2.46***	-1.85***
	(0.78)	(0.68)	(0.22)	(0.26)	(0.29)	(0.33)	(0.31)	(0.38)	(0.32)	(0.35)	(0.33)	(0.35)
4th cut-off	-2.36**	-0.28	-4.52***	-4.17***	-2.62***	-2.88***	-3.27***	-3.20***	-1.66***	-1.07**	-1.63***	-0.91**
	(0.78)	(0.67)	(0.21)	(0.25)	(0.29)	(0.33)	(0.31)	(0.38)	(0.32)	(0.35)	(0.33)	(0.35)
5th cut-off	-2.11**	-0.14	-4.39***	-4.14***	-2.38***	-2.81***	-3.19***	-3.17***	-1.60***	-1.03**	-1.60***	-0.90**
	(0.78)	(0.67)	(0.21)	(0.25)	(0.29)	(0.33)	(0.31)	(0.38)	(0.32)	(0.35)	(0.33)	(0.35)
6th cut-off	-1.63**	-0.05	-3.33***	-3.84***	-1.32***	-2.49***	-2.15***	-2.89***	-0.62*	-0.84**	-0.63*	-0.73**
	(0.77)	(0.67)	(0.21)	(0.25)	(0.29)	(0.33)	(0.31)	(0.38)	(0.32)	(0.35)	(0.32)	(0.35)
7th cut-off	-0.63	0.39	-2.34***	-3.54***	-0.36	-2.28***	-1.20***	-2.68***	0.14	-0.67*	0.10	-0.58
	(0.78)	(0.67)	(0.21)	(0.25)	(0.29)	(0.33)	(0.31)	(0.38)	(0.32)	(0.35)	(0.32)	(0.35)
No. of Obs.	4240	3009	18626	12861	12944	10332	12032	9521	10770	9636	10608	9390

Source: Author's calculations based on weighted PALMS Data (1993 - 2012)

occupation variable that differentiates the first occupation category (managers) from the second category (professionals) and the second from the third and so on. The coefficients reported in Table 4.8 will be used in the estimations of counterfactual occupational choices described in Section 4.4.2.

4.4.2 Counterfactual Micro-Simulations

The previous section described the different steps of estimating individuals' wages. For this, the probability of participation was estimated. Furthermore, the probability of employment was predicted while accounting for selection bias. For both steps, an individual's probability to participate or be employed in period t as well as the coefficients used to estimate these probabilities were stored for all periods t. Additionally, Table 4.7 presented the results of a wage regression run on all employed individuals. Finally, Table 4.8 examined the probability of each individual to choose a particular occupation in period t and the probabilities as well as coefficients used for the prediction were stored. This section will discuss how to derive a counterfactual scenario for each of these steps, using the estimators generated in Section 4.4.1.

In the outline of the micro-simulation approach by Bourguignon et al. (2008), the authors assume perfect clearing conditions for the labour market. If the labour market is in equilibrium, each individual will be allocated into counterfactual categories based on a utility-optimizing choice with regards to labour force participation and employment. However, González-Rozada and Menendez (2006) note that high levels of structural unemployment distort these equilibrium conditions in such a way that individuals will not always be allocated into their optimal selection choice. To simulate the effects of such distortions, random error terms will be added in the selection into labour force participation as well as into employment. The details of this randomization will be discussed below.

Participation Effect

As a logical first step, the counterfactual simulations start with modifying the participation rate. The procedure of simulating the effect of a change in participation will be discussed in detail below. The rest of the arguments will be changed accumulatively by replacing one argument at a time.

In this first step, the wage distribution of a counterfactual participation rate will be estimated. That is, this step estimates the counterfactual wage distribution of

$$YL_t^{1*} = f(P_{t+l}, E_t, O_t, X_t, R_t, \varepsilon_t). \tag{4.9}$$

To assess the effect of a counterfactual participation rate, it is necessary to simulate a counterfactual labour force. For this purpose, the probability of participating will be estimated for each individual at time t as if they were confronted with the circumstances of period $t + l$. This is done by applying the coefficients that determine participation in $t + l$ to individuals' characteristics in period t. The coefficients associated with the choice of participating are estimated for each period following the bivariate probit model discussed above. Using the coefficients β^P_{t+l}, a linear prediction index S^*_p of individuals in time t is estimated where $S^*_p = \beta^P_{t+l} X_t$. Following this procedure would result in a probability of participating which is dependent on characteristics observed and the assumed preferences they represent. However, the non-clearing nature of the labour market prohibits such an allocation into labour force participation or non-participation. In order to simulate the distortions caused by structural unemployment, the allocation into participation will be randomized as follows.

The randomization process applied in this chapter follows Garlick (2016) in a modification to the method by González-Rozada and Menendez (2006). For each member of the actual sample in time t, a random number will be drawn from a standard uniform distribution $U(0, 1)$. The random numbers ξ_1 are then multiplied by the standard deviation of the counterfactual prediction index S^*_p before they are added to S^*_p. Rescaling the random numbers in this way is necessary to ensure equal variance between the deterministic prediction index and the random component. The result of this procedure is a semi-random transformation of the linear prediction index which can be used to estimate the counterfactual probability of participating:

$$P^*_p = \frac{exp(\beta^P_{t+l} X^P_t + \xi_1)}{1 + exp(\beta^P_{t+l} X^P_t + \xi_1)}. \tag{4.10}$$

Following this step, the number of participating individuals in the counterfactual, N^*_p, is determined by applying the $t + l$ participation rate to the sample at time t such that $N^*_p = N_t \times P_{t+l}$. Once the size of the counterfactual sample has been established in this manner, individuals are ordered by the semi-random counterfactual probability of participating from highest probability to lowest and N^*_p individuals with the highest P^*_p values are selected into the counterfactual sample of participation.

Introducing an element of randomness produces a counterfactual sample that will include some individuals whose assigned participation status differs from their actual status in that period. This imitates the effect that high structural unemployment and other labour market imperfections have in preventing individuals from obtaining their optimal participation outcome. This is a more accurate assumption in the reality of the South African labour market than the perfect allocation Bourguignon et al. (2008) presumed.

González-Rozada and Menendez (2006) use a similar randomization method, however, Gar-

lick (2016) identifies a problematic feature of the sequential Poisson sampling technique used by the first two authors. The selection term ε_p estimated by González-Rozada and Menendez (2006) divides a random number ξ_p drawn from a standard uniform distribution $U(0,1)$ by the estimated probability $P^0 = \beta_{t+l}X_t$. The selection term is therefore defined as $\varepsilon_p = \xi_p/\beta_{t+l}X_t$. However, this can lead to extreme values as the probability P^0 in the denominator might take values very close to zero. This can eliminate the advantages of randomization as the assumption of asymptotically normal distribution might fail, rendering the results open to doubt. Therefore, the additive randomization approach is preferable.

Once the counterfactual participation sample has been identified in this manner, labour income is estimated for the individuals identified in the sample and the Gini coefficient is estimated as a measurement of inequality. The only factor that has been changed for this counterfactual distribution of labour income is the participation rate.

Employment Effect

As a second argument, the counterfactual employment sample is drawn. Similarly to the participation effect, the allocation into employed and unemployed will be randomized as imperfections in the labour market prohibit individuals to be allocated optimally. As such, this step estimates

$$YL_t^{2*} = f(P_{t+l}, E_{t+l}, O_t, X_t, R_t, \varepsilon_t). \tag{4.11}$$

Of the counterfactual participation sample established in the first step, a number $N_e^* = N_p^* \times E_{t+l}$ will be employed. The size of the counterfactual employment sample N_e^* is established by multiplying the employment rate of period $t+l$ with the number of individuals in the counterfactual participation sample N_p^*. As discussed in the determination of the counterfactual participation sample, the coefficients estimated in Section 4.4.1 are used to estimate a counterfactual prediction index S_e^*. This counterfactual prediction index is the result of simulating an individual's employment outcome as if they were in period $t+l$. This is achieved through applying the coefficients from period $t+l$ to the characteristics in period t. A random number ξ_2 is drawn from a uniform distribution and the semi-random probability of employment P_e^* is estimated as follows

$$P_e^* = \frac{exp(\beta_{t+l}^E X_t^E + \xi_2)}{1 + exp(\beta_{t+l}^E X_t^E + \xi_2)}. \tag{4.12}$$

Individuals will be ranked in a descending order and the N_e^* individuals with the highest counterfactual probability of being employed will be selected into the counterfactual employment sample.

Once the counterfactual employment sample has been identified, labour income is estimated for the individuals selected into the sample and the Gini coefficient is estimated as a measurement of inequality. The only factors that have been changed to simulate period $t+l$ for this counterfactual distribution of labour income is the participation rate and the employment rate.

Occupational Allocation Effect

Changes in the occupational structure of the labour force can have significant effects on the distribution of overall labour income. Therefore, Table 4.8 estimated the probability of an individual to be allocated into a specific occupation. In this step of the counterfactual estimation, equation

$$YL_t^{3*} = f(P_{t+l}, E_{t+l}, O_{t+l}, X_t, R_t, \varepsilon_t) \tag{4.13}$$

will be estimated.

In order to simulate individuals selecting into an occupation in period t as if they were acting in period $t+l$, the coefficients from $t+l$ will be applied to the characteristics in period t. Using the linear prediction scores resulting from this, each individual in the counterfactual employment sample established above will be given a counterfactual probability of choosing a particular occupation. The number of individuals in each occupational category is estimated by multiplying the counterfactual number of people in the employment sample, N_e^*, with the fraction of individuals in each occupation category in $t+l$ such that for each occupation c: $N_{o,c}^* = N_e^* \times F_{o,c}$, where $F_{o,c}$ is the fraction of individuals in a specific occupation category c in period $t+l$. To establish the counterfactual occupation, individuals will be ordered by the highest probability of being in a specific occupational category and those $N_{o,c}^*$ individuals with the highest probability will be allocated into that category. Unlike the allocation into participation and employment, there is no element of randomness assumed to interfere with the allocation into an occupation once an individual has entered the labour market.

Using the counterfactual dummies of occupation, labour income and inequality measures are estimated for the counterfactual employment sample. Other characteristics such as age or years of education have not been changed at this stage.

Endowment Effect

Over time, the socio-demographic characteristics of individuals in the labour market change. Through assessing the 'endowment effect' or 'population effect', this step analyses how changes in the socio-demographic structure of individuals in the labour market affect labour market income.

Table 4.1 described changes in the demographic structure of the working age population across all years under observation. These changes will be accounted for when estimating

$$YL_t^{4*} = f(P_{t+l}, E_{t+l}, O_{t+l}, X_{t+l}, R_t, \varepsilon_t). \tag{4.14}$$

It is not simply possible to apply the characteristics from period $t+l$ to period t. Therefore, the methodology in this chapter will follow a re-weighting approach laid out by Inchauste et al. (2012). In a rank-preserving transformation, the percentiles in the distribution of demographic variables such as age and years of education in period t will be replaced with the percentiles of these variables observed in period $t+l$. That means that the first percentile of age in year t will be replaced with the mean age of the first percentile in year $t+l$. The same will be done for years of education. Additionally, all dummy variables such as geographical, provincial and population group dummies, will be re-weighted in such a way that they reflect the socio-demographic structures observed in period $t+l$. This is done by applying the ratio of changes in means between period t and period $t+l$ to each dummy variable. Lastly, the weighting variable will be adjusted to represent the population growth between period t and $t+l$. Again, a ratio will be applied to adjust the weighting variable in period t for the change in population size by time $t+l$. The wage distribution that will be estimated following these adjustments will account for the following changes accumulatively: the change in participation, employment and occupational structure as well as for the change in characteristics of individuals and population growth in the counterfactual employment sample.

Price Effect

An important determinant of the distribution of labour market income is how the characteristics of individuals are being rewarded. The Oaxaca-Blinder decomposition method acknowledged this fact by deriving not only the endowment effect but also the so-called 'price effect' which is determined by the change in returns at constant mean endowment. The methodology applied in this chapter is able to determine the 'price effect' or 'returns effect' not only for mean endowments but by carrying out this exercise for the entire distribution of the counterfactual employment sample. Therefore, this step estimates

$$YL_t^{5*} = f(P_{t+l}, E_{t+l}, O_{t+l}, X_{t+l}, R_{t+l}, \varepsilon_t). \tag{4.15}$$

Using the coefficients estimated in the wage regression in Section 4.4.1, the change in returns can be simulated by using the coefficients from period $t+l$ in the estimation of counterfactual wages. As the different effects are added one after another, the price effect is estimated for the counterfactual employment sample that is allocated into counterfactual occupations after

socio-demographic characteristics have been re-weighted. That means that the coefficients β_1, β_2, γ and δ from period t in equation (4.8) are replaced with coefficients from period $t+l$. Therefore, the wage equation for the counterfactual employment sample is transformed in this way:

$$\begin{aligned} \ln\ w &= \alpha + \beta_1 \cdot age + \beta_2 \cdot age^2 + \gamma \cdot yrsedu + \delta X + \lambda + \varepsilon, \\ \ln\ w &= \alpha_{t+l} + \beta_{1,t+l} \cdot age^*_{t+l} + \beta_{2,t+l} \cdot age^{2,*}_{t+l} + \gamma_{t+l} \cdot yrsedu^*_{t+l} + \delta_{t+l} X^*_{t+l} + \lambda_{t+l} + \varepsilon, \end{aligned} \tag{4.16}$$

Since counterfactual individual characteristics have been estimated in the previous step, indicated with $*$ and $t+l$, they will be included in the estimation of the price effect. Therefore, the Gini coefficient is estimated for the earnings distribution resulting from the estimation of equation (4.16).

Effect of the Error Term

The final component that is being changed in the estimation of the counterfactual wage distribution is the regression residual. The residual captures unobserved characteristics such as ability. Through re-scaling the predicted residuals of the earnings regression, the effect of changes in unobservable factors can be estimated. However, this estimation is not a straightforward procedure. Instead, it is necessary to perform a rank-preserving exercise on the residuals. Assuming that the error terms are normally distributed with a zero mean, it is possible to approximate the effect of changes in the error term by re-scaling the residuals of period t with the ratio of standard deviations in period t and $t+l$. Following Inchauste et al. (2012), the final counterfactual is defined as

$$YL^*_t = f(P_{t+l}, E_{t+l}, O_{t+l}, X_{t+l}, R_{t+l}, (\hat{\sigma}^{t+l}/\hat{\sigma}^t)\varepsilon_t), \tag{4.17}$$

where $\hat{\sigma}^t$ and $\hat{\sigma}^{t+l}$ are the variances of the error term in period t and $t+l$ respectively.

The effects that have been discussed in this section include the effect of changes in the participation choice, the effect of changes in the employment decision, changes in the occupational structure, a re-weighting process to assess the endowment effect as well as changes due to population growth, the price effect and finally, the effect of changes in the regression residual. However, the disadvantage of this sequential decomposition approach is that the effects are not assessed individually while holding all other components constant. Except for the first step of the decomposition, the measurement of the participation effect, each effect is measured when more than one factor has been changed. As such, the observed change in the distribution after measuring the effect of changes in employment also captures the interaction between the participation effect and the employment effect. In order to correct for that particular issue, the Heckman correction is employed. However, in the current

sequential estimation of the different effects of the wage regression, there are interactions between the change in price and the change in unobservables for example that have not been captured. Every following effect includes the interaction between itself and all previously estimated effects. Notably, the price effect is estimated after the endowment effect has been measured and the effect of unobservables is measured after both, the price effect and the endowment effect. Bouillòn et al. (2013) offer a solution to this problem.

According to Bouillòn et al. (2013), the following transformation can be used to decompose the total change in the joint earnings distribution into the following parts:

$$(-)\Delta D = \underbrace{D_\beta(X_{t+l}, \epsilon_t)}_{return} + \underbrace{D_X(\beta_t, \epsilon_t)}_{population} + \underbrace{D_\epsilon(X_{t+l}, \beta t + l)}_{unobserved} + \underbrace{[D_\epsilon(\beta_t, X_{t+l}) - D_\epsilon(\beta_t, X_t)]}_{remainder} \quad (4.18)$$

where the total change in the earnings distribution can be disaggregated into the sum of the returns effect, the population effect (endowment effect), the effect of unobservables and a remainder term that captures the interactions of the returns effect and the effect of the unobservables. Therefore, this remainder will be included in the discussion of the results of the counterfactual wage estimation in Section 4.5.

4.5 Results

The counterfactuals discussed in Section 4.4.2 give rise to a set of results. Table 4.9 presents a synopsis of the counterfactual simulations that will be estimated. At each step, one further component is added to the counterfactual. Furthermore, all counterfactual estimations will be simulated with period t as starting period $t0$ and period $t + l$ as final period $t1$ and vice versa, with period $t + l$ as starting period and period t as final period. As such, there will be an estimation where the year 1993 is the initial year of estimation, $t0$, and the year 2012 is the final year of estimation, $t1$, but there will also be an estimation where the year 2012 is the initial year $t0$ and the year 1993 is the final year $t1$. The effects are estimates both ways in order to account for path-dependence. The results are still subject to path-dependency as the effects are added cumulatively, however, due to the nature of the labour market, it is not possible to counteract effects of path-dependence in the way that Azevedo et al. (2013) have done for household income. In their micro-simulations of the effect of different household sources and household demographics on changes in inequality, the authors were able to calculate Shapely-Shorrocks values to limit possible effects of path dependence. Shapely-Shorrocks values are estimated by calculating the effects of the different components across all possible paths.[3] This approach has been employed in Chapter 2. However, given the consecutive choices of individuals in the labour market, not all paths are

[3]See for example Shorrocks (2013).

possible. It is not logical to assess the effect of changes in the employment structure before estimating the effect of changes in the participation choice. Estimating the average effects of using both periods as the initial year is therefore the only logical remedy to the issue of path-dependency in the counterfactual estimations discussed below.

Figure 4.1 and Figure 4.2 report the average decomposition effects for women and men respectively. The figures show the results of the counterfactual simulations performed using both, period t as the initial year and period $t+l$, as the initial year. For the discussion of the results of the counterfactual simulations, it is important to note that the simulated contribution of the different effects estimated to the overall change in earnings inequality is not intended to explain what caused those changes. Instead, the estimated counterfactuals provide empirical evidence on the effect of trends in the labour market and the resulting changes in earnings inequality.

Table 4.9: Counterfactual Estimations

Counterfactual	**Description**
$YL_t^* = f(P_{t+l}, E_t, O_t, X_t, R_t, \varepsilon_t)$	Individuals' participation choice in period t is simulated as if the choice was made in period $t+l$. The counterfactual participation sample takes size N_P^*. Allocation into participation is semi-random.
$YL_t^* = f(P_{t+l}, E_{t+l}, O_t, X_t, R_t, \varepsilon_t)$	Additional to the above, the employment outcome in period t is simulated as if in period $t+l$. The counterfactual employment sample takes size N_E^*. Allocation into employment is semi-random.
$YL_t^* = f(P_{t+l}, E_{t+l}, O_{t+l}, X_t, R_t, \varepsilon_t)$	Additional to the above, individuals are allocated into counterfactual occupations. No randomization.
$YL_t^* = f(P_{t+l}, E_{t+l}, O_{t+l}, X_{t+l}, R_t, \varepsilon_t)$	Additional to the above, socio-demographic characteristics are transformed in a rank-preserving exercise. The population weight is re-scaled.
$YL_t^* = f(P_{t+l}, E_{t+l}, O_{t+l}, X_{t+l}, R_{t+l}, \varepsilon_t)$	Additional to the above, returns to characteristics from period $t+l$ are applied to the counterfactual employment sample.
$YL_t^* = f(P_{t+l}, E_{t+l}, O_{t+l}, X_t, R_t, (\hat{\sigma}^{t+l}/\hat{\sigma}^t)\varepsilon_t)$	Additional to the above, the distribution of the residuals is re-scaled.

Additionally, the changes to the Gini coefficients are presented in the order of effects that are being added to the counterfactual estimation, moving from the actual Gini coefficient observed in $t0$ to the actual Gini observed in $t1$ and vice versa. The nature of the estimation of the endowment effect allows to separate out the effect of population growth and the Gini coefficient of the rescaled population weight which are reported separately. Furthermore, all estimations are done separately for men and women.

The earlier discussion of Table 4.5 showed that there was an overall increase in the Gini coefficient of real earnings between 1993 and 2012 for men and women. The Gini coefficient for men peaked in 2012 at 0.62, an increase from previously 0.583 in 1993. For women, the Gini coefficient increased consistently between 1995 and 2012 after a slight decrease between 1993 and 1995. In 1995, the Gini coefficient for female real earnings was at 0.511, in 2012 the inequality measure had increased to 0.6 which is slightly lower than the Gini coefficient for men in the same year. Figure 4.1 and Figure 4.2 decompose how the different effects contributed to the changes in the Gini coefficient across years under observations for men and women respectively. The sums underlying Figure 4.1 and Figure 4.2 are shown in

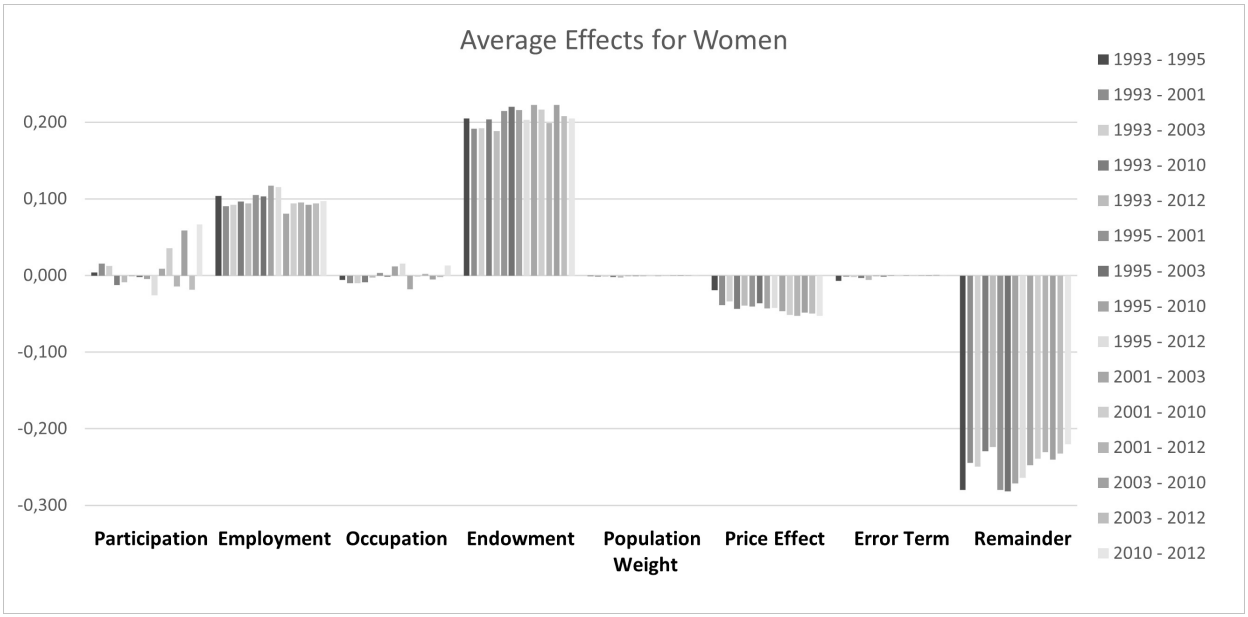


Figure 4.1: Decomposition of Average Effects for Women 1993 - 2012

Tables 4.11 and 4.12 in Appendix 4.A. All Gini coefficients associated with the discussion of the changes in this section can be found in Table 4.10 in Appendix 4.A.

There are a number of facts that stand out in the presentation of Figure 4.1. Firstly, a large increasing effect on changes in inequality across all years under observation is the employment effect. For all years, the changes in employment rates as well as in the expected returns to employment increased inequality significantly. Table 4.4 showed that employment rates of women first decreased until 2003 and increased in the years after. These changes affected the distribution of earnings significantly. Furthermore, the low absorption rate of the South African labour market (see for example Leibbrandt, Levinsohn and McCrary, 2010) probably contributed to the employment effect driving inequality upwards.

The largest increasing factor to inequality in earnings is the endowment effect. The changes in observed characteristics increased inequality by up to 0.2 points in the Gini coefficient particularly between 2001 and 2003 as well as between 2003 and 2010. The endowment effect measures the effect of changes in the age and educational structures as well as in the geographical structure. Of these factors, the unequal opportunities that characterize education in South Africa (Garlick, 2016; Branson et al., 2012) are most likely causing such a large endowment effect. However, geographical marginalization in rural areas will contribute to increases in the Gini coefficient as well. The changes in these characteristics do not take into account the change in population size. Therefore, the weighting variable is adjusted to account for population growth by population group.

Table 4.4 disaggregated the changes in employment rates by population group. The effects

were small but significant. In order to account for the changes in overall population size as well as changes in the different population groups, the weighting variable is re-calibrated. However, the effects of these changes are rather small and mostly close to zero.

Changes in occupational allocations had comparatively small effects that changed signs in different comparisons. In all comparisons to 1993 as well as between 2001 and 2003, the changes in occupational allocation reduced the changes in the Gini coefficient. In subsequent comparisons, the effect is mostly increasing.

Two effects had mostly decreasing effect on the Gini coefficient of earnings for women. Those are the price effect and the interaction effect. The effect of changes in returns to characteristics was relatively small below -0.1. It is essential to keep in mind that the effect of changes in returns was measured after the endowment effect. That means that the effect of changes in characteristics has already been absorbed at the point of measuring the effect of changes in their returns. While the effect of changes in characteristics has been large and positive, the changes in returns are negative across all years under observation.

The decomposition of the changes in earnings as displayed in the simplified equations (4.9) to (4.17) allows the separating out of the effects discussed thus far. However, there is an additional interaction that is captured in the remainder that follows from the transformation captured in equation (4.18) following Bouillòn et al. (2013). The effect of the changes in these interaction terms is strongly negative on the inequality measured in the distribution of earnings of women between 1993 and 2012. The largest effects are between 1993 and 1995 and 1995 and 2003. The effects are negative but smaller in size in all other decompositions. As the name suggests, the interaction effect captures possible interactions between the effects discussed so far that result from the fact that the effects cannot be estimated across all possible paths. A change in the price effect might affect employment but because these steps have to be modeled after one another, such effects are lumped together in the interaction term. Therefore, it is difficult to say what exactly caused such a large negative interaction effect but it is possible that the remainder captured a number of changes that happened particularly at the time of the regime change in South Africa which is why the term is particularly large at earlier points in time.

Additionally, the average effect of changes in the error term is reported in Figure 4.1. This effect captures changes in the distribution of unobserved characteristics, for example ability, and the remuneration associated with these characteristics. However, the decomposition shows that the effect in these unobserved characteristics is close to zero across all possible paths.

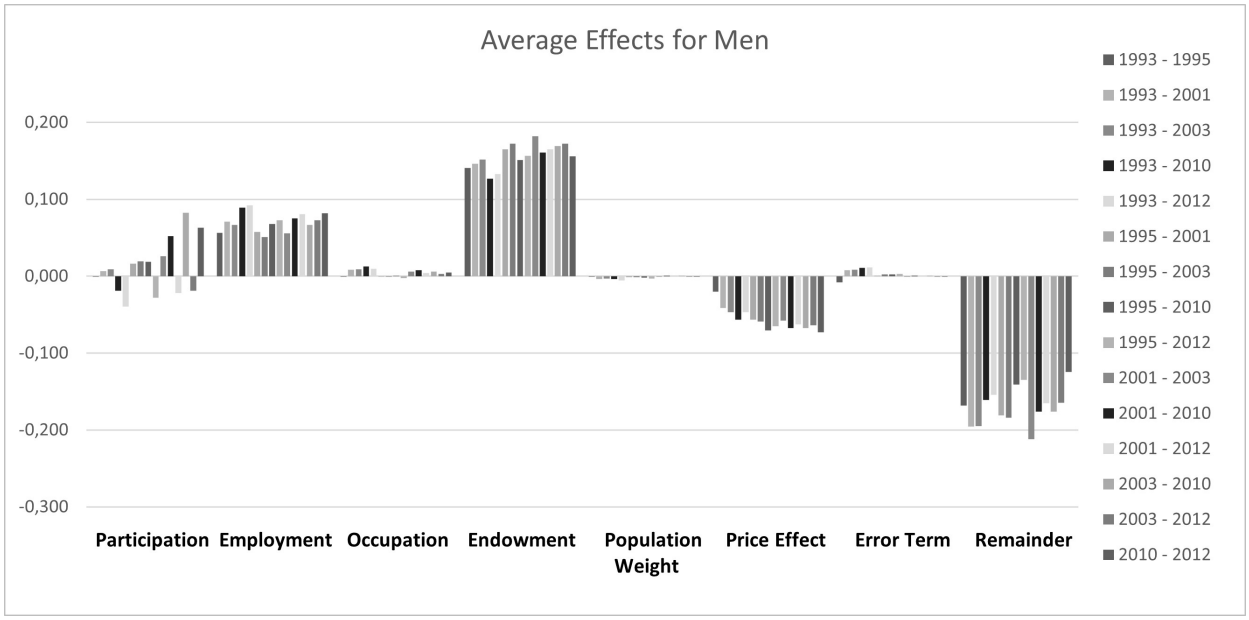


Figure 4.2: Decomposition of Average Effects for Men 1993 - 2012

Figure 4.2 displays the decomposition of changes in the Gini coefficient of earnings for men between 1993 and 2012. The effects shown have generally the same signs as the effects discussed in Figure 4.1.

The effect of changes in participation had no clear sign in the decomposition of changes in female earnings. The same is true for men, except that the participation effect show slightly higher magnitudes for men.[4] These changes have the largest positive effect between 2003 and 2010 as well as 2010 and 2012, between 1993 and 2012, the largest negative effects can be observed.

After assessing the participation effect, the changes in employment had an increasing effect on inequality across all year of observation. The effects are slightly less strong than those for women with the change in the Gini coefficient of below 0.1 points. The strongest effects are estimated between 1993 and 2010 as well as 1993 and 2012, indicating that over the entire period under observation changes in the employment rate increased the Gini coefficients. This is to be expected given that unemployment increased over this period of time.

Changes in the occupational structure indicate a small but increasing effect on inequality. This is different to the observation of changes in the female earnings distribution. The effect is the highest between 1993 and 2010.

[4]Figures 4.1 and Figure 4.2 are scaled exactly the same to ensure comparability.

Table 4.1 showed that the age and educational structure of men of working age changed significantly between 1993 and 2012. Therefore, the effect of these changes in characteristics is estimated. Overall, the endowment effect had positive signs and was highest between 2001 and 2010. Additionally, the effect of changes in the size and make up of the population is estimated by re-calibrating the weighting variable. The effect captured by the population weight remains close to zero.

A strong negative effect on the distribution of earnings for men was the change in returns to observed characteristics. This effect is captured by the price effect which is highest during the later years under observation. While the effect was negative for women as well, it is stronger in the distribution of male earnings.

Unobserved characteristics such as ability are captured in the error term of the wage regression. The changes in such unobserved characteristics can have an impact on the overall distribution of wages. However, the decomposition in Figure 4.2 shows that for all years under observation, that effect was close to zero. Even though between 1993 and 2010 as well as 1993 and 2012 respectively, changes in unobserved characteristics slightly increased the Gini coefficient. Additional to the explained effects, there are interactions between the changes in the different components of the wage equation. Following Bouillòn et al. (2013), these interactions are captured in the remainder. For women, these interactions were quite large and negative. For men the scale of these effects are slightly smaller but remain negative, meaning they decrease the Gini coefficient.

Overall, the most significant effect on the changes in the earnings distribution for men was the interaction effect which had a decreasing effect on the inequality of earnings. As such, the interactions between the different effects counterbalanced some of the large increasing effects of changes in endowment and employment. It is unfortunate that this large decreasing factor on inequality cannot be sensibly decomposed further due to the issue of path dependency when estimating wages.

4.6 Conclusion

Income inequality is a severe problem in the South African society and has been shown to be driven by high levels of inequality in the labour market. Table 4.5 showed that inequality of real earnings as measured by the Gini coefficient increased between 1993 and 2012. Earnings inequality reported a Gini coefficient of 0.591 for men in 1993. For women, that measurement stood at 0.583. Over the period under observation, the Gini coefficient increased to 0.617 for men; inequality worsened for women as well, for whom the Gini coefficient was at 0.6 in 2012. In order to understand the empirical regularities underlying these changes, this

chapter applied advanced micro-simulation techniques.

The data used for the analysis in this chapter stems from the Post-Apartheid Labour Market Series (PALMS) and data sets chosen span from 1993 to 2012. Due to a change in instruments that affect the underlying earnings distribution, more recent years of the study cannot sensibly be included in this analysis. The micro-simulations employed in this chapter follow an approach by Bourguignon et al. (2008) with a modification to account for the non-clearing nature of labour markets that feature high levels of unemployment by González-Rozada and Menendez (2006).

In order to identify the principles underlying the changes in earnings inequality, the determinants of wages need to be estimated. These include the prediction of labour market participation, employment outcome, occupational choice as well as the prediction of wages. The bivariate probit estimation of labour market participation and employment outcome that accounted for selection bias showed that all other provinces had lower probabilities of participation than Gauteng but many reported higher chances of employment. Furthermore, the estimation showed lower chances of employment for all population groups relative to white South Africans. The effect is particularly high for African women. The data confirms previous findings of increased female participation in the labour market, particularly in the early years after apartheid (Bhorat and Leibbrandt, 1999a; Casale, 2004 and Van der Westhuizen et al., 2007). The increase in the participation rate of women exceeded even the increased participation rate of men over the same period. However, for women, this did not translate into equal improvements in the rate of employment leading to increased unemployment particularly for African women. African women have a higher probability to participate than white women and even report higher coefficients than African men despite the fact that higher participation probabilities do not translate into higher employment probabilities. One reason for this development is the fact that rather than being pulled into the labour market through increased demand for labour, women are being pushed into the labour market due to economic need (Casale and Posel, 2002). This leads to increased self-employment and informal labour for women. Section 4.3 reported that there has been a significant increase in capturing marginal workers in the PALMS data (Wittenberg, 2016a), including women in the informal market. However, Bhorat et al. (2016) support that capturing the complexities of vulnerability in South African survey data is difficult. Therefore, at least some of the increase in unemployment for women can be attributed to the fact that informal labour is regularly underestimated in labour market surveys and female employment is therefore not fully captured (Casale and Posel, 2002).

The micro-simulation approach by Bourguignon et al. (2008) aimed at extending the Oaxaca-Blinder decomposition. Through counterfactual micro-simulations, a whole new

distribution is estimated which can be assessed through counterfactual Gini coefficients. The coefficients of the bivariate probit estimation as well as the estimation of the probability to choose a particular occupation together with the coefficients of the wage regression are used to construct counterfactual wages that account for the changes in different determinants of the wage equation. This analysis estimated not only the 'price effect' and 'endowment effect' for men and women across the whole earnings distribution but further decomposed changes in earnings inequality into a participation effect, employment effect, the effect of changes in the occupational structure, population weight and the effect of the error term. Furthermore, a remainder that accounts for interaction between these effects was reported.

The decomposition has shown that the changes in endowments mainly drove the increase in earnings inequality for women and men. This is despite the fact that mean years of education have increased for both, men and women, between 1993 and 2012. However, changes in the characteristics of working age individuals affect different parts of the counterfactual distribution. Firstly, they affect the probability of participation, secondly the probability of employment and thirdly the wage outcome upon successful employment. Garlick (2016) finds that by affecting labour market outcomes, changes in education increased earnings inequality in wage employment even though through affecting remuneration (i.e. the price effect), changes in education decreased wage inequality. This supports the dichotomy of findings presented in the figures in Chapter 4.5. Endowments increase earnings inequality and the price effect reduces it. The fact that increased average education can have such opposing effects traces back to the structure of educational attainment in South Africa. The increase in mean years of education over the post-apartheid era has left a majority of working age individuals without complete secondary education (Branson et al., 2012). Furthermore, the quality of the education received differs widely across South Africa (Chamberlain and van der Berg, 2002). Over the period under observation, average years of education increased but so did the share of African working age individuals. The National Development Plan (2013) recognizes that "The quality of school education for most black learners is poor." This supports the findings by Chamberlain and van der Berg (2002) that there exist racial disadvantages that can be traced back to high inequality in the quality of education received.

Another increasing factor for both, women and men, was the changes in employment. These effects were partially counteracted by the remainder. This 'interaction term' captures the effects of interaction between all the other estimated effects that remain unobserved. This interaction term had a decreasing effect on female earnings inequality. Due to path dependency in the estimation of counterfactual wages, certain interactions between the different effects cannot be sensibly drawn out. Estimating average effects is the only attempt to try

and make up for this short-coming. However, seeing the large effect of the interaction terms on the female and male earnings distribution, it would certainly be desirable to analyse the underlying interactions more closely.

The implementation of micro-simulations offers in-depth explanations of trends that can be observed in the data. The decomposition highlighted the increasing effect of changes in employment on earnings inequality, particularly for women. Some of this discrepancy between employment and participation may be attributed to the difficulty in capturing informal labour. Therefore, labour force surveys should improve the tools to capture different aspects of vulnerability in the labour market. However, this cannot explain the entire increasing effect of changes in employment on inequality. Some of this effect may trace back to the differential treatment of women without any obvious explanations in the hiring decision of the employers. Bhorat and Goga (2013) as well as Kwenda and Ntuli (2018) find evidence of discrimination based on gender in the South African labour market. Furthermore, the discussion of wage determinants such as age, education and marital status have shown that there are few significant differences between men and women, especially in more recent years. Nonetheless, the results of Chapter 3 indicated that women and Africans are more likely to earn no taxable income. This chapter showed that women select primarily into lower paying jobs in elementary occupations. Thus, it is necessary to address the underlying reason for this. If women are pushed into the labour market due to economic need and have to make work for themselves because they cannot find (formal) employment opportunities (Casale and Posel, 2004), then it is necessary to establish a social security net that supports these women such that they can take the time to search for more stable employment opportunities without facing poverty due to lack of income. Overall, it will be necessary to create not only jobs but also ensure stable employment in order to lift households and individuals out of poverty in the long run. This however, may result in a trade-off for the government between the creation of a more flexible labour market by stimulating an increase in informal workers that can lead to more job growth, or the creation of less but more stable employment which can reduce inequality in the long term. The National Development Plan (2013) is acutely aware of the need to adapt policies for the needs of informal livelihoods and therefore suggests social security mechanisms that cover those risks associated with informal employment. This could solve the mentioned trade-off, if implemented correctly. The envisioned policy changes are a tall order for a government faced with many inequalities but are necessary for a more equal society.

Furthermore, in order to increase expected returns from female labour, the skills of women, particularly at the bottom of the earnings distribution, should be improved. This can be done by addressing the inequalities prevalent in education. However, resources are scare and there are large differences in schools' ability to convert those resources into positive outcomes

(van der Berg, 2008). Therefore, pro-poor spending becomes more important. Schools that perform particularly bad need to be given more resources and guidance than those that perform well. Teachers and school principals need appropriate training and bureaucratic structures should be established such that they support teachers' needs rather than complicate processes. Some of these policies have been attempted since the end of apartheid with very poor results. Therefore, those areas that perform the least should be identified and addressed vigorously in order to reduce the continuously high level of inequality in education.

Similarly, the marginalization, particularly of African women, despite policy interventions and affirmative action laws such as the Employment Equity Act of 1998 requires a revision of these policies and their implementations. The National Development Plan (2013) already offers several strategies against income inequality as well as gender inequality. These plans include expanding public employment, particularly for women and youth, addressing educational barriers as well as the underlying social, cultural and religious obstacles women may face when entering the job market. This indicates that the government understands the complexities of the structural disadvantages faced by women and Africans, however, the South African government is yet to prove that it can translate this plan into good policies.

The use of micro-simulations in the analysis of the evolution of earnings inequality in post-apartheid South Africa has highlighted several areas for additional research and has offered analytical foundations for policy interventions to reduce earnings inequality in South Africa. Chapter 5 will summarize the findings of this and the other chapters of this thesis before presenting the contributions to international and South African literature made by the various analyses discussed in this thesis.

4.A Counterfactual Results

All Gini coefficients associated with the decomposition of changes in the earnings distribution are presented in 4.10. The Gini coefficients are shown in sequential order moving the Gini coefficients from $t0$ to $t1$ and are ordered according to years in $t0$. The table presents the results of all counterfactuals with the results for the estimation of the female counterfactual wage estimation reported in the second row for every year.

Table 4.10: Counterfactual Gini Simulations

t0	Actual Gini t0	Participation	Employment	Occupation	Endowment Effect	Population Weight	Price Effect	Error Term	Actual Gini t1	t1
t0=1993										
1993	0.591	0.579	0.645	0.659	0.78	0.781	0.72	0.759	0.563	1995
1993	0.583	0.605	0.691	0.657	0.851	0.852	0.816	0.838	0.511	1995
1993	0.591	0.581	0.64	0.65	0.782	0.781	0.701	0.75	0.550	2001
1993	0.583	0.606	0.686	0.651	0.851	0.853	0.806	0.821	0.583	2001
1993	0.591	0.58	0.645	0.66	0.786	0.786	0.692	0.753	0.542	2003
1993	0.583	0.607	0.699	0.666	0.857	0.86	0.824	0.831	0.593	2003
1993	0.591	0.579	0.659	0.681	0.794	0.794	0.693	0.757	0.609	2010
1993	0.583	0.603	0.689	0.663	0.856	0.858	0.799	0.828	0.582	2010
1993	0.591	0.577	0.654	0.677	0.788	0.788	0.691	0.762	0.617	2012
1993	0.583	0.604	0.688	0.662	0.854	0.856	0.814	0.827	0.600	2012
t0=1995										
1995	0.563	0.574	0.622	0.606	0.767	0.765	0.786	0.732	0.591	1993
1995	0.511	0.497	0.619	0.64	0.856	0.855	0.852	0.816	0.583	1993
1995	0.563	0.574	0.619	0.616	0.775	0.777	0.715	0.718	0.550	2001
1995	0.511	0.503	0.613	0.629	0.856	0.859	0.827	0.818	0.583	2001
1995	0.563	0.574	0.62	0.618	0.776	0.778	0.71	0.727	0.542	2003
1995	0.511	0.503	0.621	0.638	0.857	0.861	0.838	0.824	0.593	2003
1995	0.563	0.572	0.617	0.616	0.772	0.775	0.701	0.714	0.609	2010
1995	0.511	0.495	0.619	0.639	0.857	0.862	0.826	0.829	0.582	2010
1995	0.563	0.573	0.619	0.618	0.772	0.775	0.698	0.72	0.617	2012
1995	0.511	0.498	0.619	0.639	0.856	0.862	0.838	0.829	0.600	2012
t0=2001										
2001	0.550	0.573	0.656	0.662	0.823	0.817	0.816	0.782	0.591	1993
2001	0.583	0.591	0.691	0.706	0.888	0.883	0.853	0.835	0.583	1993
2001	0.550	0.572	0.643	0.645	0.815	0.811	0.759	0.758	0.563	1995
2001	0.583	0.589	0.689	0.68	0.882	0.878	0.829	0.836	0.511	1995
2001	0.550	0.573	0.636	0.642	0.816	0.817	0.756	0.767	0.542	2003
2001	0.583	0.596	0.68	0.668	0.88	0.881	0.837	0.834	0.593	2003
2001	0.550	0.572	0.638	0.647	0.816	0.82	0.758	0.767	0.609	2010
2001	0.583	0.589	0.68	0.674	0.882	0.885	0.833	0.843	0.582	2010
2001	0.550	0.571	0.64	0.65	0.812	0.816	0.747	0.764	0.617	2012
2001	0.583	0.592	0.683	0.678	0.874	0.878	0.835	0.837	0.600	2012
t0=2003										
2003	0.542	0.571	0.639	0.643	0.821	0.813	0.814	0.769	0.591	1993
2003	0.593	0.594	0.686	0.699	0.892	0.887	0.855	0.844	0.583	1993
2003	0.542	0.57	0.626	0.626	0.813	0.808	0.758	0.746	0.563	1995
2003	0.593	0.597	0.686	0.666	0.887	0.882	0.832	0.843	0.511	1995
2003	0.542	0.571	0.621	0.626	0.816	0.816	0.761	0.75	0.550	2001
2003	0.593	0.598	0.676	0.652	0.885	0.884	0.834	0.838	0.583	2001
2003	0.542	0.569	0.62	0.626	0.813	0.816	0.757	0.754	0.609	2010
2003	0.593	0.592	0.679	0.663	0.887	0.89	0.838	0.852	0.582	2010
2003	0.542	0.569	0.62	0.628	0.806	0.809	0.745	0.748	0.617	2012
2003	0.593	0.594	0.679	0.663	0.881	0.885	0.841	0.847	0.600	2012
t0=2010										
2010	0.609	0.583	0.682	0.687	0.827	0.82	0.807	0.764	0.591	1993
2010	0.582	0.537	0.644	0.653	0.867	0.86	0.832	0.796	0.583	1993
2010	0.555	0.584	0.674	0.676	0.822	0.816	0.749	0.74	0.563	1995
2010	0.528	0.535	0.646	0.651	0.864	0.858	0.808	0.806	0.511	1995
2010	0.501	0.584	0.669	0.675	0.828	0.825	0.753	0.744	0.550	2001
2010	0.474	0.539	0.636	0.64	0.864	0.861	0.81	0.8	0.583	2001
2010	0.447	0.585	0.668	0.673	0.825	0.823	0.746	0.749	0.542	2003
2010	0.420	0.539	0.637	0.642	0.864	0.861	0.817	0.804	0.593	2003
2010	0.393	0.584	0.662	0.671	0.815	0.816	0.734	0.74	0.617	2012
2010	0.366	0.536	0.633	0.642	0.855	0.856	0.814	0.802	0.600	2012
t0=2012										
2012	0.617	0.552	0.66	0.656	0.811	0.8	0.803	0.755	0.591	1993
2012	0.600	0.562	0.665	0.687	0.872	0.865	0.828	0.804	0.583	1993
2012	0.617	0.552	0.651	0.647	0.807	0.798	0.744	0.729	0.563	1995
2012	0.600	0.562	0.671	0.683	0.872	0.865	0.803	0.811	0.511	1995
2012	0.617	0.552	0.644	0.643	0.81	0.806	0.749	0.733	0.550	2001
2012	0.600	0.563	0.663	0.673	0.875	0.871	0.809	0.808	0.583	2001
2012	0.617	0.552	0.646	0.644	0.81	0.806	0.744	0.74	0.542	2003
2012	0.600	0.562	0.665	0.676	0.874	0.871	0.815	0.811	0.593	2003
2012	0.617	0.552	0.638	0.638	0.806	0.806	0.742	0.735	0.609	2010
2012	0.600	0.563	0.661	0.677	0.874	0.874	0.811	0.821	0.582	2010

Source: Author's calculations based on weighted PALMS Data (1993 - 2012)

The tables below show the sums underlying Figure 4.1 and Figure 4.2 which result from estimating the average effects from Table 4.10.

Table 4.11: Decomposition of Average Effects for Women 1993 - 2012

t0	Participation	Employment	Occupation	Endowments	Population Weight	Price Effect	Error	Interaction	t1
1993	0.004	0.104	-0.006	0.205	0.000	-0.019	-0.007	-0.280	1995
1993	0.015	0.090	-0.010	0.191	-0.001	-0.039	-0.002	-0.245	2001
1993	0.013	0.092	-0.010	0.192	-0.001	-0.034	-0.002	-0.250	2003
1993	-0.013	0.097	-0.009	0.204	-0.002	-0.044	-0.004	-0.229	2010
1993	-0.009	0.094	-0.003	0.189	-0.002	-0.039	-0.006	-0.224	2012
1995	-0.001	0.105	0.003	0.215	-0.001	-0.040	-0.001	-0.280	2001
1995	-0.002	0.103	-0.001	0.220	-0.001	-0.036	-0.001	-0.282	2003
1995	-0.004	0.117	0.012	0.216	-0.001	-0.043	0.001	-0.271	2010
1995	-0.026	0.116	0.015	0.203	-0.001	-0.043	-0.001	-0.264	2012
2001	0.009	0.081	-0.018	0.222	0.000	-0.047	0.000	-0.248	2003
2001	0.036	0.094	-0.001	0.216	0.000	-0.052	0.000	-0.239	2010
2001	-0.014	0.096	0.002	0.199	0.000	-0.053	0.000	-0.231	2012
2003	0.059	0.092	-0.005	0.223	0.000	-0.048	0.000	-0.240	2010
2003	-0.019	0.094	-0.002	0.208	0.000	-0.050	0.001	-0.232	2012
2010	0.066	0.097	0.013	0.205	0.000	-0.053	-0.001	-0.220	2012

Source: Author's calculations based on weighted PALMS Data (1993 - 2012)

Table 4.12: Decomposition of Average Effects for Men 1993 - 2012

t0	Participation	Employment	Occupation	Endowments	Population Weight	Price Effect	Error	Interaction	t1
1993	0.000	0.057	-0.001	0.141	0.000	-0.020	-0.008	-0.168	1995
1993	0.006	0.071	0.008	0.147	-0.003	-0.041	0.008	-0.195	2001
1993	0.009	0.067	0.009	0.152	-0.003	-0.047	0.008	-0.195	2003
1993	-0.019	0.089	0.013	0.127	-0.004	-0.057	0.011	-0.161	2010
1993	-0.040	0.092	0.010	0.133	-0.005	-0.047	0.011	-0.154	2012
1995	0.017	0.058	0.000	0.165	-0.001	-0.057	0.000	-0.181	2001
1995	0.019	0.051	-0.001	0.172	-0.002	-0.059	0.003	-0.184	2003
1995	0.019	0.068	0.000	0.151	-0.002	-0.070	0.002	-0.141	2010
1995	-0.028	0.073	-0.002	0.157	-0.003	-0.065	0.003	-0.135	2012
2001	0.026	0.056	0.006	0.182	0.000	-0.058	0.000	-0.212	2003
2001	0.052	0.075	0.008	0.161	0.000	-0.067	0.000	-0.176	2010
2001	-0.022	0.081	0.004	0.165	0.000	-0.063	0.000	-0.165	2012
2003	0.082	0.067	0.006	0.169	0.000	-0.068	0.000	-0.176	2010
2003	-0.019	0.073	0.003	0.172	-0.001	-0.063	0.000	-0.165	2012
2010	0.063	0.082	0.005	0.156	0.000	-0.073	0.000	-0.125	2012

Source: Author's calculations based on weighted PALMS Data (1993 - 2012)

Chapter 5

Conclusion

5.1 Key Findings and Results

This thesis set out to interrogate three distinct issues relating to income inequality in South Africa. Investigating dynamic rather than static questions allowed for gaining a deeper understanding of the drivers of income and earnings inequality, and access to a unique data set on income tax records made it possible to obtain new insights into the evolution of the income distribution in the post-apartheid era.

Chapter 2 examined how the dynamics of micro-determinants explain changes in aggregate income inequality. The increase in income inequality between 1993 and 2008 was primarily driven by changes in labour market income although, through successful targeting of government grants, an even stronger increase in inequality could be offset. The more refined analysis using micro-simulations showed that particularly between 2008 and 2014 changes in household composition drove inequality upwards, whereas all income sources report an equalizing effect. The nuances of these effects would be overlooked when relying on the traditional static decomposition approach. Overall, this chapter found a small decrease in the Gini coefficient between 1993 and 2014 despite an increase in inequality measured in 2008.

Chapter 3 analysed the trends in the income distribution in household survey data compared to tax administration data. The comparison highlighted a potential weighting issue for top earners in the household survey that caused an upward bias in the Gini coefficient in the later year of the panel survey. Otherwise, NIDS captured taxable incomes relatively well and highlighted a large fraction of individuals earning no taxable income. This supports the result of Chapter 2 that many individuals rely on non-taxable incomes, such as government grants, as income sources. Using tax administration data to account for incomes at the top of the income distribution, the results show that inequality decreased between 2011 and 2014.

Finally, Chapter 4 took a closer look at the labour market and used micro-simulations to examine how changes in the determinants of wages explain changes in earnings inequality over time. While actively accounting for the effects of structural unemployment on participation and selection into employment, the results show that changes in observed characteristics as well as the discrepancy between participation and employment were strong contributors to increases in earnings inequality between 1993 and 2012. On the other hand, changes in the returns to observed characteristics decreased earnings inequality. Changes in occupational choices mostly decreased earnings inequality over time for women while these changes led to small increases in earnings inequality for men. This chapter highlighted the various channels through which changes in wage determinants contributed to the increase in earnings inequality for women and men between 1993 and 2012.

There are various potential avenues of research that can expand the key findings of this thesis. The results of Chapters 2 and 3 showed a decrease in household income inequality and individual taxable income before 2014 whereas Chapter 4 showed that earnings inequality increased between 1993 and 2012. This discrepancy certainly warrants further investigation even though the comparability of the results is limited due to the different years and different surveys adopted for this investigation. Furthermore, due to the change in survey instruments after 2012 in the PALMS data, more recent analysis of the earnings distribution is desirable but should be done with caution. Given the finding of mis-measurement of high earners in the later years of NIDS, it would be of interest to compare more recent household survey data to tax administration data. If SARS would make such data available over a longer period of time, the evolution of high earners and their representation in household survey data could be tracked.

The strong effect of changes in the labour market, particularly of unemployment, played a predominant role in the discussion of the results of the different chapters. When investigating the labour market and changes in the earnings distribution more closely, Chapter 4 highlighted some differences in the effects for women and men. Therefore, a more detailed analysis of the differences between female and male earnings using micro-simulations would be interesting. Additionally, the micro-simulation approach to decomposing earnings over time highlighted strong interaction effects between the different determinants of wages assessed. Therefore, there is room to investigate which interactions are driving the results and in what way. Chapter 4 also showed that the increase in average education did not lead to a decrease in inequality but rather that the inequality in education potentially leads to increased earnings inequality. The quality and inequality of education is a prevalent topic in South Africa which already prompts many researchers to investigate the subject. The varying effects of changes in education on participation, employment, choice of occupation and determining wages should be studied in more detail.

Chapters 2 and 3 highlight the strong reliance on government grants as a (non-taxable) income source and as an instrument to counteract increases in inequality. A valuable lesson to be learned from these results is that, given the macro-economic state of the country, South Africa can no longer rely on social grants in this way. There is little scope for fiscal policy to expand government grants. Additionally, the redistributive effects of grants may be exhausted. Furthermore, Chapter 4 concluded that a reliable social security net can assist women in securing incomes while looking for formal employment. One possibility could be to improve the efficiency of government grants by adjusting their administration and design. However, a more long-term solution that has even higher redistributive potential would mean that policy makers focus on developing instruments that allow individuals to

earn sustainable incomes as well as create revenue for the state. The closer examination of the labour market in Chapter 4 showed that the endowment effect contributed to increasing earnings inequality. Therefore, a critical starting point would be the education and training of (potential) workers. South Africa still relies heavily on labour-intensive trades that require low skills. Therefore, skills development and education should be a priority in order to move the labour market into the 21st century. It is critical to improve on the quality of education for all South Africans, not just a privileged few, and eliminate high inequality in education to aid workers in their progression in the labour market. Furthermore, employment or wage subsidies could be implemented in order to reduce employment costs and increase labour demand. Not only can this affect women who are particularly vulnerable to high unemployment and informal labour but this can also contribute positively to GDP growth in order to get closer to the goals set out in the NDP. Overall, active labour market policies and a commitment to creating training opportunities as well as improving the quality of primary and secondary education should be the focus of policy makers who want to build an inclusive and more equal South Africa.

5.2 Contributions to Understanding Income Inequality in South Africa

Income inequality can be studied in many different ways, and as such, this thesis offers various contributions to international and South African literature on income and earnings inequality. The different chapters present a detailed, up-to-date analysis of income and earnings inequality over the post-apartheid period, that, through the use of dynamic decompositions of changes in incomes, has highlighted some new and important aspects of inequality in South Africa.

Extending and moving away from the classic decomposition methods made it possible for Chapter 2 to analyse the contribution of income sources that make up a rather small share of overall household income, such as government grants or investment income. Furthermore, Chapter 2 has been able to show the dynamic contribution of social grants within a rigorous analytic framework for the first time. Static decompositions cannot do this and other simulation work on grants has focused on studying densities or growth incidence curves with and without grants. The micro-simulations approach offers a more substantial understanding of income dynamics than these kind of 'after-the-fact'-type of analyses. Additionally, the use of micro-simulations accounted for demographic changes (household composition, structure of the population) and estimated their importance in the changes of income and earnings inequality. The thesis has shown that there have been massive changes in these demographic factors (including labour participation and migration) and so their explicit

inclusion is important.

Given that the analysis in this thesis is based on income data from household surveys, mis-measuring income dynamics at the top end is potentially very important. This thesis has taken this concern seriously and, through careful work with tax administration data, contributed to the understanding of dynamics at the top end that did not come out of the income source decomposition work and indeed, was hidden by the decomposition work. Due to lack of appropriate data, previous research was not able to thoroughly investigate the make-up of earners at the top of the income distribution or to accurately evaluate the changes in inequality for that part of the income distribution. Detailed comparative analysis of information from household survey data and income tax records showed the strengths and weaknesses of each type of data set and found that for the ongoing panel of the household survey, concerns about mis-measuring at the top end were justified. This thesis found a remedy for this by combining the inequality indices from both data sources to measure inequality across the entire distribution.

This thesis was able to show that, while labour income is always dominant and always a driver of increased inequality, this role has not been constant over time (as the static decompositions seem to suggest). The dynamic decompositions show labour income to be strongly dis-equalizing for most of the period, but less so recently. As these dynamics are very important, this thesis extended the research into labour market dynamics through a second detailed micro-simulation exercise. Chapter 4 modeled the factors driving the changes in earnings inequality by accounting for the various avenues through which the different determinants of wages affect earnings while accounting for South Africa's most important obstacle in the labour market, structural unemployment. For the first time, existing distortions caused by structural unemployment could be accounted for by introducing an element of randomness into the participation and employment outcome. Changes in the underlying characteristics that determine wages work not only through affecting the participation and employment outcome but also through changes in the occupational structure as well as changes in the returns to those characteristics. By meticulously accounting for changes in each of these different channels, this thesis drew attention to the potential dichotomy of effects caused by changes in the determinants of wages which need to be taken into account when addressing the issue of earnings inequality.

Thinking about changes in income inequality as being driven by changes in labour market income, income from government grants and demography, this thesis has added to the understanding of each of these three. Furthermore, this thesis provided comprehensive insights into the developments at the top end of the income distribution, an area that was previously not possible to assess due to the lack of available tax administration data. The

various determinant of wages work through different channels. By investigating the effects of a variety of determinants on these different channels, this thesis has deepened the understanding of the labour market in South Africa. Overall, this thesis has laid substantive groundwork for further investigation into each of the different fields of analysis.

Data Sources

[1] SALDRU (Southern Africa Labour and Development Research Unit). Project for Statistics on Living Standards and Development - PSLSD [dataset]. *Southern Africa Labour and Development Research Unit* [producer], *DataFirst* [distributor], Cape Town (2012). Version 2.0.

[2] SALDRU (Southern Africa Labour and Development Research Unit). National Income Dynamics Study 2008, Wave 1 [dataset]. *Southern Africa Labour and Development Research Unit* [producer], *DataFirst* [distributor], Cape Town (2015). Version 6.1.

[3] SALDRU (Southern Africa Labour and Development Research Unit). National Income Dynamics Study 2010-2011, Wave 2 [dataset]. *Southern Africa Labour and Development Research Unit* [producer], *DataFirst* [distributor], Cape Town (2015). Version 3.1.

[4] SALDRU (Southern Africa Labour and Development Research Unit). National Income Dynamics Study 2014-2015, Wave 4 [dataset]. *Southern Africa Labour and Development Research Unit* [producer], *DataFirst* [distributor], Cape Town (2016). Version 1.1.

[5] Kerr, A., Lam, D. and Wittenberg, M. (2016), Post-Apartheid Labour Market Series [dataset]. *DataFirst* [producer and distributor], Cape Town (2016). Version 3.1.

[6] SARS (South African Revenue Service). Personal Income Tax Data 2010 - 2011 [dataset]. *South African Revenue Service*, Pretoria (2016).

[7] SARS (South African Revenue Service). Personal Income Tax Data 2013 - 2014 [dataset]. *South African Revenue Service*, Pretoria (2016).

Literature

[1] Alexander, P. Rebellion of the poor: South Africa's service delivery protests–a preliminary analysis. *Review of African Political Economy 37*, 123 (2010), 25–40.

[2] Alvaredo, F. A note on the relationship between top income shares and the Gini coefficient. *Economics Letters 110*, 3 (2011), 274–277.

[3] Alvaredo, F., and Atkinson, A. B. Colonial rule, apartheid and natural resources: Top incomes in South Africa, 1903-2007. *CPER Working Paper Series*, 8155 (2010).

[4] Alvaredo, F., and Londoño, J. High incomes and personal taxation in a developing economy: Colombia 1993-2010. *Commitment to Equity Working Paper*, 12 (2013).

[5] Atkinson, A. B. *Measuring top incomes: Methodological issues.* Oxford University Press, 2007, ch. 2, pp. 18–42.

[6] Atkinson, A. B., Piketty, T., and Saez, E. Top incomes in the long run of history. *Journal of Economic Literature 49*, 1 (2011), 3–71.

[7] Azevedo, J., Inchauste, G., and Sanfelice, V. Decomposing the recent inequality decline in Latin America,. *World Bank Policy Research Working Paper*, 6715 (2013).

[8] Azevedo, J., Sanfelice, V., and Nguyen, M. Shapley decomposition by components of a welfare measure. *World Bank* (2012).

[9] Barros, R. P. D., Carvalho, M. D., Franco, S., and Mendonça, R. Uma análise das principais causas da queda recente na desigualdade de renda brasileira. *Instituto de Pesquisa Econômica Aplicada (Ipea)*, 1203 (August 2006).

[10] Bhorat, H., and Goga, S. The gender wage gap in post-apartheid South Africa: A re-examination. *Journal of African Economies* (2013), 1–22.

[11] Bhorat, H., and Hodge, J. Decomposing shifts in labour demand in South Africa. *South African Journal of Economics 67*, 3 (1999), 155–168.

[12] Bhorat, H., and Leibbrandt, M. Correlates of vulnerability in the South African labour market. *DPRU Working Paper*, 99/27 (1999).

[13] Bhorat, H., and Leibbrandt, M. Modelling vulnerability and low earnings in the South African labour market. *DPRU Working Paper*, 99/32 (1999).

[14] Bhorat, H., Lilenstein, K., Oosthuizen, M., and Thornton, A. Vulnerability in employment: Evidence from South Africa. *DPRU Working Paper*, 04 (2016).

[15] Bhorat, H., and Oosthuizen, M. The Post-Apartheid Labour Market: 1995 - 2004. *DPRU Working Paper*, 06 (2006).

[16] Bhorat, H., Steenkamp, F., Rooney, C., Kachingwe, N., and Lees, A. Understandingand characterizing the services sector in South Africa: An overview. *DPRU Working Paper*, 201803 (2018).

[17] Blinder, A. S. Wage discrimination: reduced form and structural estimates. *Journal of Human Resources* (1973), 436–455.

[18] Bouillon, C. P., Legovini, A., and Lustig, N. Rising inequality in mexico: Household characteristics and regional effects. *Journal of Development Studies 39*, 4 (2003), 112–133.

[19] Bourguignon, F. The poverty-growth-inequality triangle. *Poverty, Inequality and Growth*, 125 (2004).

[20] Bourguignon, F., Ferreira, F. H., and Leite, P. G. Beyond Oaxaca-Blinder: Accounting for differences in household income distributions. *The Journal of Economic Inequality 6*, 2 (2008), 117–148.

[21] Bourguignon, F., Ferreira, F. H., and Lustig, N. *The microeconomics of income distribution dynamics in East Asia and Latin America*. World Bank Publications, 2005.

[22] Branson, N., Garlick, J., Lam, D., and Leibbrandt, M. Education and inequality: The south african case. *SALDRU Working Paper*, 75 (2012).

[23] Branson, N., Garlick, J., Lam, D., and Leibrandt, M. Education and inequalit y: The south african case. *SALDRU Working Paper 75* (2012).

[24] Burger, R., and Woolard, I. The state of the labour market in South Africa after the first decade of democracy. *Journal of vocational education and training 57*, 4 (2005), 453–476.

[25] Burkhauser, R. V., Feng, S., Jenkins, S. P., and Larrimore, J. Recent trends in top income shares in the United States: reconciling estimates from March CPS and IRS tax return data. *Review of Economics and Statistics 94*, 2 (2012), 371–388.

[26] Casale, D. What has the feminisation of the labour market 'bought' women in South Africa? Trends in labour force participation, employment and earnings, 1995–2001. *Journal of Interdisciplinary Economics 15*, 3-4 (2004), 251–275.

[27] Casale, D. M., and Posel, D. R. The continued feminisation of the labour force in south africa: An analysis of recent data and trends. *South African Journal of Economics 70*, 1 (2002), 156–184.

[28] Chisholm, L., and September, J. *Gender Equity in South African Education 1994 - 2004*. HSRC Press, Cape Town, 2005.

[29] Cowell, F. A. Measurement of inequality. *Handbook of Income Distribution Volume 1* (2000), 87–166.

[30] Crouch, L., Gustafsson, M., and Lavado, P. *Measuring Educational Inequality in South Africa and Peru*. Comparative Education Research Centre, 01 2009, pp. 461–484.

[31] Dell, F. Top incomes in Germany and Switzerland over the twentieth century. *Journal of the European Economic Association 3*, 2-3 (2005), 412–421.

[32] Diaz-Bazan, T. V. Measuring inequality from top to bottom. *World Bank Policy Research Working Paper*, 7237 (2015).

[33] DiNardo, J., Fortin, N. M., and Lemieux, T. Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. *Econometrica*, 5 (1996), 1001–1044.

[34] Donaldson, A. Enabling growth: redistribution priorities for South Africa. *Econ REDI 3x3* (2014).

[35] Ferreira, F., and de Barros, R. P. The slippery slope: explaining the increase in extreme poverty in urban Brazil, 1976-1996. *PUC-Rio Texto para discussão*, 404 (1999).

[36] Finn, A., and Leibbrandt, M. The evolution and determination of earnings inequality in post-apartheid South Africa. *WIDER Working Paper*, 83 (2018).

[37] Finn, A., Leibbrandt, M., and Levinsohn, J. Income mobility in South Africa: Evidence from the first two waves of the National Income Dynamics Study. *NIDS Discussion Paper*, 05 (2012).

[38] Firpo, S., Fortin, N., and Lemieux, T. Decomposition methods in economics. *Handbook of Labor Economics 4* (2011), 1–102.

[39] Firpo, S., Fortin, N. M., and Lemieux, T. Unconditional quantile regressions. *Econometrica 77*, 3 (2009), 953–973.

[40] Garbinti, B., Goupille-Lebret, J., and Piketty, T. Income inequality in France, 1900-2014: Evidence from Distributional National Accounts (DINA). *Journal of Public Economics 162* (2018), 63–77.

[41] Garlick, J. Changes in the inequality of employment earnings in South Africa. Master's thesis, University of Cape Town, 2016.

[42] González-Rozada, M., and Menendez, A. Why have urban poverty and income inequality increased so much? Argentina, 1991–2001. *Economic Development and Cultural Change 55*, 1 (2006), 109–138.

[43] Greene, W. *Econometric Analysis*, Fith Edition ed. Upper Saddle River, New Jersey, 2002.

[44] Heckman, J. J. Sample selection bias as a specification error. *Econometrica 47*, 1 (1979), 153–161.

[45] Hoogeveen, J. G., and Özler, B. Poverty and inequality in post-apartheid South Africa: 1995-2000. *Poverty and policy in post-apartheid South Africa* (2006), 59–94.

[46] Inchauste, G., Azevedo, J. P., Essama-Nssah, B., Olivieri, S., Nguyen, T. V., Saavedra-Chanduvi, J., and Winkler, H. *Understanding changes in poverty*. The World Bank, 2014.

[47] Inchauste, G., Lustig, N., Maboshe, M., Purfield, C., and Woolard, I. The distributional impact of fiscal policy in South Africa. *World Bank Policy Research Working Paper*, 7194 (2015).

[48] Inchauste, G., Olivieri, S., Saavedra-Chanduvi, J., and Winkler, H. What is behind the decline in poverty since 2000? Evidence from Bangladesh, Peru and Thailand. *World Bank Policy Research Working Paper*, 6199 (2012).

[49] Jenkins, S. P. Pareto models, top incomes and recent trends in UK income inequality. *Economica 84*, 334 (2017), 261–289.

[50] Jensen, R. T. Do private transfers 'displace' the benefits of public transfers? Evidence from South Africa. *Journal of Public Economics 88*, 1-2 (2004), 89–112.

[51] Jorda, V., and Niño-Zarazúa, M. Global inequality: How large is the effect of top incomes? *World Development 123* (2019), 104593.

[52] Kerr, A., and Wittenberg, M. A guide to version 3.1 of the Post-Apartheid Labour Market Series (PALMS). *Data First* (2016).

[53] Kingdon, G., and Knight, J. Unemployment in South Africa, 1995–2003: Causes, problems and policies. *Journal of African Economies 16*, 5 (2007), 813–848.

[54] Kwenda, P., and Ntuli, M. A detailed decomposition analysis of the public-private sector wage gap in South Africa. *Development Southern Africa 35*, 6 (2018), 815–838.

[55] Lakner, C., and Milanovic, B. Global income distribution: From the fall of the Berlin Wall to the Great Recession. *World Bank Economic Review 30* (2016).

[56] Leibbrandt, M., and Finn, A. Inequality in South Africa and Brazil - can we trust the numbers? *CDE Insight, Centre for Development an Enterprise* (2012).

[57] Leibbrandt, M., Finn, A., and Woolard, I. Describing and decomposing post-apartheid income inequality in South Africa. *Development Southern Africa* (2012).

[58] Leibbrandt, M., Levinsohn, J. A., and McCrary, J. Incomes in South Africa after the fall of apartheid. *Journal of Globalization and Development 1*, 1 (2010).

[59] Leibbrandt, M., Woolard, C., and Woolard, I. Poverty and inequality dynamics in South Africa: Post-apartheid developments in the light of the long-run legacy. *South African Economic Policy under Democracy. Oxford University Press, Oxford* (2009).

[60] Leibbrandt, M., Woolard, I., Finn, A., and Argent, J. Trends in South African income distribution and poverty since the fall of apartheid. *OECD Social, Employment and Migration Working Papers*, 101 (2010).

[61] Lemieux, T. Decomposing changes in wage distributions: A unified approach. *Canadian Journal of Economics/Revue canadienne d'économique 35*, 4 (2002), 646–688.

[62] Lerman, R. I., and Yitzhaki, S. Income inequality effects by income source: a new approach and applicationsto the United States. *The Review of Economics and Statistics* (1985), 151–156.

[63] Levy, B., Hirsch, A., and Woolard, I. Governance and inequality: Benchmarking and interpreting South Africa's evolving political settlement. *ESID Working Paper No. 51* (2015).

[64] Lustig, N. The missing rich in household surveys: Causes and correction methods. Commitment to Equity (CEQ) Institute, Tulane University, CEQ Working Paper 75 (forthcoming).

[65] Maitra, P., and Ray, R. The effect of transfers on household expenditure patterns and poverty in South Africa. *Journal of Development Economics 71*, 1 (2003), 23–49.

[66] Mincer, J., and Polachek, S. Family investments in human capital: Earnings of women. *Journal of Political Economy 82*, 2, Part 2 (1974), 76–108.

[67] Mohanty, M. S. Effects of positive attitude and optimism on employment: Evidence from the US data. *The Journal of Socio-Economics 39*, 2 (2010), 258–270.

[68] Morelli, S., Smeeding, T. M., and Thompson, J. P. Post-1970 trends in within-country inequality and poverty: Rich and middle income countries. *Center for Studies in Economics and Finance Working Papers*, 356 (2014).

[69] National Planning Commission. National Development Plan 2030. *Our future - make it work* (2011).

[70] National Treasury, and South African Revenue Service. Tax Statistics. *South African Revenue Service (SARS)* (2015).

[71] National Treasury, and South African Revenue Service. Tax Statistics. *South African Revenue Service (SARS)* (2017).

[72] NIDS. Wave 3 overview. Tech. rep., National Planning Comission, 2013.

[73] Ntuli, M., and Wittenberg, M. Determinants of black women's labour force participation in post-apartheid South Africa. *Journal of African Economies 22*, 3 (2013), 347–374.

[74] Oaxaca, R. Male-female wage differentials in urban labor markets. *International Economic Review* (1973), 693–709.

[75] Oosthuizen, M. Low pay in South Africa. *International Labour Review 151*, 3 (2012), 173–192.

[76] Orthofer, A. Wealth inequality in South Africa: Evidence from survey and tax data. *REDI 3x3*, 2016/15 (2016).

[77] Posel, D. Inter-household transfers in South Africa: prevalence, patterns and poverty. *SALDRU Working Paper 180* (2016).

[78] Posel, D., and Casale, D. *Internal labour migration and household poverty in post-apartheid South Africa*. HSRC Press, 2006, p. 351.

[79] Posel, D., and Rogan, M. Gendered trends in poverty in the post-apartheid period, 1997–2006. *Development Southern Africa 29*, 1 (2012), 97–113.

[80] Schiel, R., Leibbrandt, M., and Lam, D. *Assessing the impact of social grants on inequality: A South African case study.* Contemporary Issues in Development Economics. Palgrave MacMillan, New York, 2016, ch. 8, pp. 112 – 135.

[81] Shorrocks, A. F. Inequality decomposition by factor components. *Econometrica: Journal of the Econometric Society* (1982), 193–211.

[82] Shorrocks, A. F. Decomposition procedures for distributional analysis: A unified framework based on the Shapley value. *Journal of Economic Inequality* (2013), 1–28.

[83] South African Revenue Service. Guide on income tax and the individual. *South African Revenue Service (SARS)* (2013).

[84] Spaull, N. Poverty & privilege: Primary school inequality in South Africa. *International Journal of Educational Development 33*, 5 (2013), 436–447.

[85] Stark, O., Taylor, J. E., and Yitzhaki, S. Remittances and inequality. *The Economic Journal 96*, 383 (1986), 722–740.

[86] Statistics South Africa. Quarterly publication on Gross Domestic Product (GDP). *Statistics South Africa*, P0441 (1993-2017).

[87] Statistics South Africa. *Gender Series Volume II: Education and Gender, 2004 - 2014.* Stats SA, 2015.

[88] Statistics South Africa. Statistical Release: Quarterly Labour Force Survey Quarter 2/2018. *Statistics South Africa*, P0211 (2018).

[89] The International Bank for Reconstruction and Development, Ed. *South Africa Economic Update* , vol. 11. The World Bank, Washington, DC, 2018.

[90] Van de Ven, W. P., and Van Praag, B. M. The demand for deductibles in private health insurance: A probit model with sample selection. *Journal of econometrics 17*, 2 (1981), 229–252.

[91] Van der Berg, S. How effective are poor schools? poverty and educational outcomes in south africa. *Studies in Educational Evaluation 34*, 3 (2008), 145–154.

[92] Van der Berg, S., and Moses, E. How better targeting of social spending affects social delivery in South Africa. *Development Southern Africa 29*, 1 (2012), 127–139.

[93] Van Der Weide, R., Lakner, C., and Ianchovichina, E. Is inequality underestimated in Egypt? Evidence from house prices. *World Bank Policy Research Working Paper* (2016).

[94] van der Westhuizen, C., Goga, S., and Oosthuizen, M. Women in the South African labour market 1995-2005. *Development Policy Research Unit (DPRU)*, 07/118 (2007).

[95] Wittenberg, M. Wages and wage inequality in South Africa 1994–2011: Part 1–Wage measurement and trends. *South African Journal of Economics* (2016a).

[96] Wittenberg, M. Wages and wage inequality in South Africa 1994–2011: Part 2–Inequality measurement and trends. *South African Journal of Economics* (2016b).

[97] Woolard, I., Simkins, C., Oosthuizen, M., and Woolard, C. Tax Incidence Analysis. *Report for the Fiscal Incidence Study conducted for National Treasury* (2005).