UNIVERSITY OF KWAZULU NATAL

DIFFERENCES IN HOURLY EARNINGS ACROSS CAPE TOWN, JOHANNESBURG & DURBAN

By Sakhile Keith Mpungose
209516715

A dissertation submitted in partial fulfilment of the requirements for the degree of Master of Commerce (Economics)

School of Accounting, Economics and Finance
College of Law and Management Studies

Supervisor: Dr Colette Muller
November 2015

Supervisors permission to submit for examination

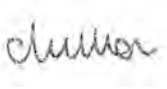
Date: 10 December 2015

Student Name: Sakhile Keith Mpungose
Student no.: 209516715
Dissertation Title: DIFFERENCES IN HOURLY EARNINGS ACROSS DURBAN, JOHANNESBURG & CAPE TOWN

As the candidate's supervisor, I believe that the work submitted is an accurate reflection of the student's ability. I agree to the submission of this dissertation for examination. I have also perused the Turn-it-in report and am of the view that any similarities between this dissertation and the work of others are either incidental or the result of the use of common phrases.

Finally, the above student has satisfied the requirements of English Language competency.

Name of Supervisor: Dr Colette Muller

Signature: 

DECLARATION

I, Sakhile Keith Mpungose declare that:

(i) The research reported in this dissertation, except where otherwise indicated, is my original research.

(ii) This dissertation has not been submitted for any degree or examination at any other university.

(iii) This dissertation does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.

(iv) This dissertation does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:

a) their words have been re-written but the general information attributed to them has been referenced; and

b) where their exact words have been used, their writing has been placed inside quotation marks, and referenced;

(v) This dissertation does not contain text, graphics or tables copied and pasted from the internet, unless specifically acknowledged, and the source being detailed in the thesis and in the reference section.

Name of student: **Sakhile Keith Mpungose**

Student number: 209516715

Signature:

ACKNOWLEDGEMENTS

First and foremost I would like to give praise to the almighty Lord for providing me an opportunity to live and prosper. He is the Almighty (Genesis 49:25), the Creator of heaven and earth (Genesis 14:19), God of all mankind (Jaremiah 32:27) and He is able to do more things than we can ask or even imagine (Ephesians 3:20). Secondly, I would like to acknowledge my family for the encouragement and support throughout the course. Last but not least, I acknowledge the School of Accounting, Economics and Finance for providing me an opportunity to partake in this research, particularly my supervisor Dr Colette Muller for providing supervision and guidance.

ABSTRACT

Researchers have long noticed the existence of wagedifferentialsacrossmetropolitan areas and researchers have documentedwagedifferentialsacrossmetropolitan areas to be due to differences in the employment and industrial structures of regionaleconomies. This paper aims to examine differences in hourly earnings across Johannesburg, Durban and Cape Town, which are the largest metropolitan areas of South Africa. In doing, the study uses pooled cross-sectional data from the Labour Force Survey, which is a household-based sample survey collected by Statistics South Africa and data is pooled from the years 2005 to 2007. The study estimates a standard wage equation through Ordinary Least Squares and by controlling for metro dummy variables only; Results suggest there are significant differences in hourly earnings across the three metropolitan areas. However by including observable characteristics, results suggest there are no significant differences in hourly earnings between Johannesburg and Cape Town whilst difference in earnings between Johannesburg and Durban still persist. The study uses the Oaxaca Blinder decomposition to decompose the mean difference in hourly earnings into a part that is explained by observable characteristics and into part that is unexplained. Oaxaca Blinder decomposition results suggest that the mean difference in hourly earnings between Cape Town and Johannesburg is 13.8 percent, the mean difference in hourly earnings between Johannesburg and Durban is 26.9 percent and the mean difference in hourly earnings between Cape Town and Durban is 44.4 percent. Overall, differences in earnings across the three respective metropolitan areas can be attributed to both varying returns to observable characteristics and discrimination.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURE

LIST OF ACRONYMS

SA : South Africa

US : United States

UK : United Kingdom

Jhb	:	Johannesburg
DBN	:	Durban
CT	:	Cape Town
StatsSA:		Statistics South Africa
BCEA	:	Basic Conditions of Employment Act 1997
O-B	:	Oaxaca Blinder Decomposition
LFS	:	Labour Force Survey
CSF	:	Census Sample File
CURF	:	Confidentialised Unit Records Files
OLS	:	Ordinary Least Squares
NUTS	:	Nomenclature of Territorial Units of Statistics
CPS	:	Current Population Survey
ABS	:	Australian Bureau of Statistics
SLM	:	Segmented Labour Markets
PSID	:	Panel Study in Income Dynamics
NUTTS:		Nomenclature of Territorial Units of Statistics

CHAPTER ONE: INTRODUCTION

Metropolitan areas can be viewed as areas in which business and residential areas are densely concentrated as a result of growth posed by such areas (Glaeser and Kahn, 2004). According to Statistics South Africa's (StatsSA) 2011 Census, Johannesburg (Jhb), Durban (DBN) and Cape Town (CT) respectively, are the largest metropolitan areas in South Africa in terms of population. Together they account for approximately 22.41 percent of the overall population of the country. Within Jhb, DBN and CT, the majority of individuals (approximately 70 percent) are of working age (between 15 and 64 years old). Metropolitan areas attract people of working age as they provide a wide range of labour opportunities for individuals due to their capacity as engines of economic growth (Bettencourt *et al.,* 2007). Further to this, metropolitan areas attract high-skilled workers and high-tech job opportunities because of the variety in the location of workers and households. This high human capital presents an attendant divergence in the economic fortunes of regions (Beeson and Groshen, 1991). Metropolitan areas present an element of attraction for people in the sense that workers with skills are attracted to these areas rather than rural areas (Glaeser and Resseger, 2009). According to this reasoning, there is a logical complimentary relationship between cities and skills. This complimentary relationship further suggests the existence of a connection between the size of the metropolitan area, the per-worker productivity levels and the earnings of workers (Glaeser and Gottlieb, 2009). However, this complimentary relationship may vary with respect to each metropolitan area, thus causing wage differentials across metropolitan areas.

Whilst research on wage differentials across geographical areas is found in the United States (US) and the United Kingdom (UK), there is a lack of such research in South Africa. The main aim of this dissertation is therefore to extend this literature by specifically exploring the earnings differences in South Africa (SA). The research will concentrate on the Johannesburg, Durban and Cape Town metropolitan areas. These three metropolitan areas have been chosen because they are the largest metropolitan areas in SA both in terms of their overall population and the number of individuals within working age. These metropolitan areas comprise intense labour market activities, thus giving the study enough bases to examine the wage differences across them.

In analysing earnings differences across the three main metropolitan areas in South Africa the dissertation utilises both descriptive and multivariate analysis of data from selected Labour Force Surveys to identify the correlates of earning differences across the Johannesburg, Durban and Cape Town metros. Furthermore, the study considers whether differences in earnings originate from geographical differences in worker characteristics or differences in how certain characteristics are rewarded across the metros.

The starting point of analysis for this study is that earnings in a perfectly competitive labour market are assumed to be equal amongst workers across labour markets and regions. This is based on the fact that in a perfectly competitive labour market, it is assumed that firms and workers are homogenous such that wages are said to equalise across labour markets (Dumond, Hirsch and Macpherson, 1999). However, various labour market wage differential theories deviate from the neoclassical view that wages equalise amongst workers across labour markets. These wage differential theories include the:

- Human Capital Theory (Mincer, 1974; Chiswick, 1974; Hirsch 1978; Chiswick and Mincer, 1972),
- Wage Discrimination Theory (Pager *et al.,* 2009),
- Theory of Compensating Wage Differentials (Hamernesh and Wolfe, 1990),
- Efficiency Wage Theory (Weiss, 1980; Shapiro and Stiglitz, 1984), and
- Trade Unions (Lewis, 1986 and Booth, 1995).

Wage discrimination theory, for example, suggests that the distribution of earnings amongst workers may vary systematically between demographic groups with workers being paid at different rates based on a demographic trait (Pager *et al*., 2009). This deviates from the neoclassical view that wages equalise amongst workers

Further to this, differences in earnings at aggregate level across regions and metropolitan areas are best explained by the Theory of Segmented Labour Markets and the Theory of Polarized Development. These theories and methods were developed by labour economists and they counter the view that wages amongst workers and labour markets equalise (Booth, 2014) and suggest instead that there are differences in earnings across labour markets, regions and metropolitan areas.

Labour economists have studied wage differentials amongst workers across metropolitan areas for various reasons (Fuchs, 1967). Firstly, wage differences amongst workers across

metropolitan areas have implications for policy implementation. There is the possibility that wage differences may occur because one metropolitan area may attract workers with higher human capital compared to another metropolitan area. Wage differences may also occur because the returns on observable characteristics are higher in one metropolitan area than in other metropolitan areas. Secondly, examining wage differences is critical as it enables researchers to test economic theories of wage differential such as Wage Discrimination Theory and Institutional Rigidities in the form of unions. To achieve this, wage differences across metropolitan areas may be caused by discrimination in the sense that varying differences will persist if one metropolitan discriminates more than other metropolitan areas. Alternatively, one metropolitan area may be highly unionised whilst other metropolitan areas are not unionised. Thirdly, wage differences across metropolitan areas can be used as an indication of prevailing wage inequalities across metropolitan areas. Wage inequalities across metropolitan areas prevail when a metropolitan area pays higher hourly earnings than other metropolitan areas. Wage inequalities across metropolitan areas suggest the existence of inefficiencies. These inefficiencies become problematic when workers in the low paying metropolitan areas are worse off, not because of their skills levels, but just because they are located in a low paying metropolitan area. When returns on observable characteristics are higher in one metropolitan area but lower in another metropolitan area for workers with identical skills and work regardless of the human capital investment, then the prevailing wage differentials across metropolitan areas become problematic, particularly to a government that advocates for investments in human capital or a government that advocates for the elimination of inequalities. Therefore, examining wage differences across metropolitan areas is critical as it enables the study to assess inefficiencies that may arise, particularly when metropolitan areas reward observable characteristics differently based upon location.

The remaining chapters are structured as follows:

Chapter Two examines some of the existing literature pertaining to wage differences across labour markets, regions and metropolitan areas in order to understand some concepts and arguments that explain differences in earnings. Chapter Two also reviews published journals on wage differences across metropolitan areas. The aim is to show how wage differential theories such as Human Capital Theory, Wage Discrimination and Polarized Development can be linked with differences in earnings amongst workers across metropolitan areas. Chapter

Three discusses the data used in this study and presents the results of the descriptive analysis of the distribution of the sample in terms of demographics and labour market outcomes across Johannesburg, Durban, and Cape Town. Chapter Four outlines the econometric methodology adopted by the study and presents the results of the multivariate analysis showing whether earnings differences exist across the three metros. The magnitude of these differences is also described. In this chapter, the Oaxaca-Blinder Decomposition is discussed and the results of the decomposition which identifies what portion of the gap in earnings between the metros can be explained by observable differences in individual characteristics between metros, and what portion is unexplained by these differences are presented. Lastly, Chapter Five provides the conclusions and recommendations of the study.

CHAPTER TWO: LITERATURE REVIEW

2.1 Introduction

Researchers have long recognised the existence of wage differentials across metropolitan areas and have documented such. According to Angel and Mitchell (1991) wage differentials across metropolitan areas are due to differences in the employment and industrial structures of regional economies. The starting point of the analysis for this study is that wages should be equal amongst homogenous workers but research by Fuchs (1967) showed that wages within labour markets are not equal and this can be explained by, amongst others, wage differential theories, differences in the returns on observable characteristics, and differences in the industrial structures of metropolitan areas (Hanushek, 1973).

This chapter examines both the theoretical and empirical literature on earnings differences amongst workers across metropolitan areas. In doing so, the study provides a starting point for the analysis by providing a brief theoretical discussion on how wages are determined and distributed in both perfect and imperfect labour markets. The discussion on wage differential theories is followed by a discussion on the causes of wage differences across metropolitan areas. Some metropolitan areas attract workers with high value human capital compared to other metros. In addition, metropolitan areas comprise industrial structures that pay higher earnings compared to other metros. The possibility exists that differences in earnings occur as a result of the cost of living including transport costs being higher in some metros, forcing businesses in those metros paying higher wages as a form of compensation compared to other low paying metropolitan areas. Finally, the study reviews some of the existing empirical evidence on wage differences across regions or metropolitan areas.

2.2 Wage Differential Theory

The labour market is distinct from other markets. Labour comprises several features that distinguishes it from other business inputs but is driven by the same dynamics as markets for goods and services (Marshall, 1980). In particular, workers (suppliers of labour) within the labour market are the owners of human capital. Workers in some labour markets (depending on what labour does) must be present for the delivery of their skills. This indirectly implies

that workers must be located in proximity to their workplace. Within the context of labour markets, the perfectly competitive labour market does not exist as it is a theoretical construct. Booth (2014) describes a perfectly competitive labour market as a market that comprises infinite firms selling identical goods. In this scenario firms make zero profits as prices charged are equal to the cost of producing an item. Based on a neoclassical perspective, this implies that there is homogeneity in firms, workers and the type of job performed within the workplace. Firms and workers are assumed to comprise perfect information about wages and job conditions so that firms can perfectly substitute workers and workers can perfectly substitute firms.

A contrary view is that the labour market is imperfect due to varying employment conditions relating to both the employer and the employee. Stigler (1962) stipulates that the imperfect state of the labour market is a result of imperfect information in the labour market. The imperfect information results in an imperfect labour market and is a contributing factor to the imperfect substitutability between workers and between employers (Manning, 2010). Manning (2010) further suggests that many aspects of the labour market can be best analysed from the perspective that there is some degree of imperfection within such markets given that workers and employers cannot be perfectly substitute as a result of imperfect and asymmetric information. In an imperfect labour market, both the employee (worker) and employer (owners of the firm) receive rent based on the existing employment relationship. On the one hand this employment relationship implies that employers will be worse off if the worker decides to leave the workplace because the rent received by the employer is gained from the services and productivity of workers. On the other hand, workers gain rent from supplying their skills within the workplace as they are rewarded for their services. The loss of his current employment will make the worker worse off as a job identical to the current one cannot be found at a zero cost to the worker.

An interesting feature of the labour market observed in recent years is the heterogeneity of jobs, workers and firms within the labour market. Within the context of workers, this implies that workers comprise different human capital, different productivity levels, different jobs and effort levels when executing the intended job. This then also implies different wages paid to each individual worker (Becker, 1975).

According to Human Capital Theory, earnings within the labour market are determined by human capital. Differences in human capital imply differences in earnings received by workers (Samuel, Gintis and Osborne, 2001). The stock of human capital can be viewed as the skills accumulated by an individual worker which enables the worker to obtain higher earnings. The stock of human capital held by each worker differs according to the worker's investment decisions. Because of these varying decisions towards the investment in human capital, workers become heterogeneous in their skills and income levels. The extent to which workers are entitled to higher earnings is a result of investment plans that seek to improve human capital. Therefore, there appears to be a clear relationship between investment in human capital and earnings received by workers. In simplifying this relationship, the stock of human capital comprises features such as the education and experience obtained by an individual worker within the workplace. Education forms the stock of human capital on the basis that an investment in education is an increasing function of earnings. However, investment decisions about education are reliant upon time. There is a trade-off between current period income (earnings without the skill) and future period income (higher earnings with skill). Workers who invest in education forego earnings that could have been obtained in the current period (without education) in order to obtain higher earnings in future periods (by investing in education).

Apart from investing in education, the work experience gained by an individual worker contributes to the stock of human capital. However, the relationship between work experience and earnings received by workers is explained by an inverted u-shape. This inverted u-shape suggests earnings rise and rise quite rapidly with time spent in employment. It peaks at a late middle age thereafter falls off somewhat towards retirement. As a result, the relationship between experience and education is non-linear. The differences in investments towards human capital by workers are arguably amongst the leading factors contributing to differences in earnings between workers in the workplace. Human Capital Theory can be used to help explain differences in earnings amongst workers (Schultz, 1974: Mincer, 1974; Chiswick, 1974; Hirsch, 1978 and Mincer and Chiswick, 1972). Given the relationship between earnings, education and work experience, the Mincerian Equation can be used to investigate the determinants of wages. The Mincerian Equation is the cornerstone of empirical labour economics as it is most commonly used to estimate the returns on education, and the

measure of the effect experience has upon hourly earnings (Heckman, Loncher and Todd, 2003).

Wage differences amongst workers can also be caused by discrimination within the workplace (Pager *et al*., 2009). Discriminatory treatment within the workplace is one of the major sources of differences in earnings amongst workers. There are two sources of discrimination, namely wage discrimination and job/occupational discrimination (Darity and Mason, 1998; Appleton *et al.*, 1999 and Green, 2003). Wage Discrimination Theory suggests that discrimination within the workplace causes wages earned by equally skilled workers to vary systematically for different demographic groups (Becker and Gary, 1957). This is to say, differences in earnings within the workplace may arise amongst equally skilled workers as compensation for non-productive characteristics such as gender, race, or sexual orientation characteristics. In wage disparities amongst workers, discrimination occurs at the point of wage-setting decisions whereby, apart from productivity characteristics, a specific group earns relatively higher than another group. This is despite the assumption that both groups have equal marginal products. Wages differ as the employer values the productivity of one group (former) and devalues the productivity of another group (latter).

Labour market discrimination can also be present in the form of occupational discrimination. Looking at discrimination from a gender/wage differential perspective, traditional and cultural factors combined have unfavourable influences on the distribution of employment opportunities for female workers. Women tend to be less exposed to employment opportunities than men (Oaxaca, 1973). Furthermore, a specific group of workers might be crowded into certain occupations resulting in reduced wages in those crowded occupations. Occupational crowding causes the supply of labour to exceed the demand, thus lowering the wages received by workers (Darity and Mason, 1998). This discrimination component results in differences in wages between those in crowded occupations (unfavourable) and those in non-crowded occupations (favourable).

Apart from wage and occupational discrimination, preferences by workers and clients also contribute towards discrimination within the workplace which results in differences in earnings amongst workers (Green, 2003). Employee discrimination occurs when a majority group is compensated more by the employer for working with a minority group in a case

where the employer demands both sets of groups within the workplace. Customer discrimination occurs in instances where customers have an aversion for sellers or workers belonging to a specific group. In cases where customers have direct contact with sellers, firms may face situations where customers receive disutility by being served by a disfavoured group. The firm compensates the disutility derived by lowering its prices to those customers and paying lower wages to the disfavoured group (Holzer and Ihlanfeldt, 1998).

Wage differences amongst workers can also occur as a form of compensating wage differentials which occur when workers supply their hours of work in risky conditions of employment (Hamermesh and Wolfe, 1986). The theory of compensating wage differentials suggests that jobs in risky conditions of employment are subject to higher wages, whilst other factors remain perfectly equal or tending towards equality (Smith, 1979). Compensating wage differentials occur in situations where jobs performed by workers are heterogeneous in the sense that they differ systematically in terms of speed, working environment, risk of accident and social prestige of the work (Garen, 1988). When analysing compensating wage differentials, it is worth noting that safety is a normal good and wealthy workers will chose to work in safe working condition rather than risky conditions (Viscusi, 1978). Consistent with this view, this implies that workers with a great stock of human capital (higher income) will work in safe working conditions compared to those with less stock of human capital. Firms compensate workers for the disutility they derived from performing risky jobs and this compensation creates wage disparities amongst workers as some worker will be reluctant to work in risky conditions (Hwang *et al.,* 1992).

Further to this, efficiency wages paid by firms provide an explanation for persistent and large wage differences amongst workers with similar observable characteristics within the workplace across labour markets (Stiglitz, 1976). Employers pay a wage that is higher than the market clearing wage in an attempt to attract a productive workforce and this market clearing wage is considered as an efficiency wage (Katz, 1986). The basic hypothesis underlying the Efficiency Wage Theory is that an individual's productivity is reliant upon the wage rate. Efficiency wages are coupled with high effort levels, decreased shirking, a productive labour force, facilitated teamwork, low turnover costs, increased morale and loyalty by workers towards the firm (Dunlop, 1985). This implies that the utility derived from working hard becomes greater than the utility derived from shirking.

Given that workers are heterogeneous within the workplace, attributes coupled with efficiency wages are based on adverse selection and incentives (Weiss, 1980; Shapiro and Stiglitz, 1984). In the Efficiency Wages Theory, adverse selection reinforces the relationship between productivity and wages due to the existence of asymmetry of information. Workers are heterogeneous in the labour market and, supposing that ability is positively correlated with wages, then higher wages will attract higher ability workers (Stiglitz, 1976b; Weiss, 1980; Malcolmson, 1981). If firms provide higher wages based on ability then workers who are willing to work for less than the going wage, reinforces the firm's presumption that the worker is a low ability worker. Low ability workers lose their jobs more frequently as they are not productive, thus remedying the adverse selection problem. Adverse selection as a result of asymmetric information explains differences in wages amongst workers due to the varying distribution of ability and acceptance of wages. However, firms can avoid adverse selection by using certain screening and selection models which induce workers to reveal their true productivity characteristics (Yallen, 1984).

Incentives also play a meaningful role in achieving productivity within the workplace. Paying higher wages or performance bonuses induces workers to supply more output within the firm (Lazear, 2000). Traditional models of efficiency wages and incentives developed by Shapiro and Stiglitz (1976) suggest that in equilibrium, differences in wages amongst homogenous workers persist as firms who find shirking to be costly, tend to offer higher wages than other firms as an attempt to eliminate shirking. This model implies that in such cases higher wages play a dual role. They allocate productive employment and provide incentives for workers to put more effort in the production process. This dual effect allows for wage dispersion amongst workers and across firms.

Another factor contributing to differences in wages amongst individuals is that of union membership. Union members earn a wage premium relative to their non-union counterparts (Lewis, 1986 and Booth, 1995). It is worth noting that some labour markets are non-unionised while other labour markets are highly unionised. The wage differential between union and the non-union labour market is higher for low-skilled workers than for high-skilled workers (Hirsch and Schumacher, 1998). Historically, trade unions have served the purpose of exerting an equalising effect upon the distribution of earnings across workers within the labour market through collective bargaining (Freeman, 1980). Unions provide a stabilising effect on the

distribution of earnings among union members and without the existence of unions, wages amongst workers within a unionised labour market will diverge (Freeman and Medoff, 1984). Therefore, theory suggests that the existence of unions eliminates the dispersion of wages amongst workers but the impact is not clear-cut (Bell and Pitt, 1998). Labour economists and researchers have long striven to examine this ambiguity in the relationship between unions and wage inequalities amongst workers across labour markets (Card *et al.,* 2004). The ambiguity stems from the basis that although unions increase the wages of their members (relative to non-union members) this generates horizontal wage inequalities between workers supplying their hours in unionised and non-unionised labour markets (Kahn, 1980; Blanchflower, 2002 and Budd, 2005).

Finally yet importantly, are the Segmented Labour Market (SLM) Theory and the Theory of Polarized Development which move away from the neoclassical view of equalising wage differentials. These theories suggest that wage differences arise because of institutional frameworks, productive structures, the level of development amongst regions and access to technology (Azzoni and Servo, 2002).

The SLM theory divides the labour market into two parts: the primary labour market which is characterised by skilled workers, higher wages and longer tenure at the workplace, and the secondary labour market which is characterised by poor tenure and lower wages. These are non-competing segments and returns on observable characteristics vary, thus implying variation in earnings amongst workers working in both segments (Leontaridi, 1998). Some metropolitan areas may comprise primary labour market concentration, whilst other metropolitan areas may comprise secondary labour market concentration. Given that both labour market segments yield varying earnings, this would then have an effect on the aggregate distribution of earnings for both metropolitan areas. For example, a metropolitan area with a concentration of primary labour markets would have a higher aggregate distribution in earnings compared to a metropolitan area with secondary labour market concentration. These varying aggregate distributions in earnings result in differences in earnings across metropolitan areas.

The Theory of Polarized Development refers to the poles of economic growth and development that is concentrated in a particular geographical place. Economic development is

associated with rising earnings, therefore geographical areas comprising poles of economic growth are said to positively affect earnings (Antonczyk *et al.*, 2010). Poles of economic growth may vary geographically. This implies that poles of economic growth may affect earnings differently depending upon geographic locations resulting in varying earnings across geographical areas affected by polarised development. This implies that if one metropolitan area is affected by polarised development whilst another metropolitan area is not, then the former will have higher aggregate distribution in earnings compared to the latter. Both these theories speak to the aggregate differences in wages across regions or metropolitan areas.

From the discussion presented thus far, it is clear that wages amongst workers are unlikely to equalise because workers are heterogeneous and the distribution of earnings amongst workers across labour markets will vary. In the next section, the study will examine the effect that metropolitan areas have on wages and also examine some of the explanations that try to understand how earnings differ across metropolitan areas.

2.3 Metropolitan Area's Effect on Wages

Metropolitan areas are regarded as engines of economic growth (Lucas, 1988). Workers tend to organise themselves around metropolitan areas because these areas offer better amenities and help to facilitate the rate of interaction between workers who are less skilled and workers who are highly skilled and thereby enhance the accumulation of human capital (Glaeser and Maré, 2001). Metropolitan areas attract highly educated workers as a result of industrialisation and the ongoing process of economic development within such cities (Storper and Scott, 2009). Metropolitan areas consist of superior facilities and they are better developed. As a result, they attract highly educated workers because highly educated workers expect to earn more and wages in metropolitan areas are higher than in other areas. This is explained by the high level of human capital that characterises the people who live in metropolitan areas (Glaeser and Maré, 2001; Glaeser and Resseger, 2009; Storper and Scott 2009; and Snow and Pavan, 2012). Therefore, the agglomeration of economic activities biased towards cities has been argued to be a result of localised information and knowledge spill-overs that are based on the accumulation of human capital (Black and Henderson, 1999).

The distribution of earnings differs across the metropolitan areas depending upon the geographic characteristics and economic characteristics of a metropolitan area. The larger the economic landscape of a metropolitan area, the higher the earnings of an individual worker residing within that metropolitan (Becker and Murphy, 1992). Wages within metropolitan areas are said to be subjected to factors such as varying compositions of the labour force, amenity levels in different regions and varying costs of living. Geographic characteristics explain the structure of the city in terms of size, the effect of city amenities upon production and the structure of earnings amongst workers within a labour market (Roback, 1988). Differences in the real wages of homogenous workers across metropolitan areas can be attributed to amenities (Carlino, 1986). This is on the basis that if workers care about the amenities of life provided by the metro, then they may be prepared to accept lower wages in high amenity metros. On the other hand, workers can also be compensated with higher wages if they are working in a low amenity area. Therefore, differences in wages amongst workers across metropolitan areas may exist to compensate for differences in amenities and dis-amenities. Without mobility constraints, in the long-run wages would equalise for workers with identical preferences and skills located in labour market with similar amenities (Dumond, Hirsch and Macpherson, 1999).

If a production factor such as labour is flexible (assuming there is nothing constraining the worker i.e. family responsibility) and can move from one region to another, then workers would not settle for lower wages in their respective regions or metropolitan areas if they were able to earn higher wages in other regions, labour markets or metropolitan areas (Azzoni and Servo, 2002). Metropolitan areas vary in terms of size such that some metropolitan areas will be relatively large in terms of industrial structure, population, and workforce compared to other metropolitan areas. Variation in the size of metropolitan areas has implications on the distribution of wages as it implies that wages will vary across metropolitan areas depending upon the size of each metro (Beeson and Groshen, 1991). There appears to be a relationship between the size of a metropolitan area and the distribution of earnings. This relationship implies that the larger the metropolitan area, the higher the wages relative to smaller metropolitan areas. This view is consistent with the view that high paying metropolitan areas have higher costs of living. The larger the metropolitan area is, the larger the wage distribution and the higher the cost of living including transport costs compared to smaller metropolitan

areas. This variation is a clear indication that wage differences across metropolitan areas can be attributed to differences in the size of metropolitan areas and the cost of living. Therefore, the relationship between large metropolitan areas and higher earnings is monotonically positive as the larger the metropolitan area, the higher the distribution of earnings and the higher the costs of living associated to that metropolitan area (Baum-Snow and Pavan, 2012).

Differences in wages across metropolitan areas can also be attributed to the industrial structures of metropolitan areas. Metropolitan areas vary in terms of industrial structure. Some metropolitan areas are dominated by industries that pay high wages whilst other metropolitan areas comprise industries that pay lower wages. This variation in earnings amongst industrial structures across metros implies that wages do not equalise as they vary depending on the industrial structure of the metropolitan areas. This variation caused by the industrial structure within metropolitan areas may serve to compensate for the cost of living associated with the metropolitan area (Johnson, 1983). Furthermore, wages differ across metropolitan areas not because of skills alone, but also because of fiscal conditions and prices that workers pay for their goods and services (Dumond *et al., 2007).*

Apart from geographic characteristics, metropolitan area size, and industrial structure, differences in wages across metropolitan areas can also be attributed to differences in the average levels of market valued worker characteristics such as experience and education, and differences in the return to other observable characteristics (Dickie and Gerking, 1987). Various labour economists and researchers have supported the view that wage differentials across metropolitan areas are a source of varying worker characteristics (Gallaway, 1963; Hanushek, 1973; Goldfarb and Yezer, 1976; Sahling and Smith, 1983 and Krumm, 1984). Some researchers have also supported the view that wage disparities across metros are a result of varying returns on observable characteristics (Coelho and Ghali, 1971; Bellante, 1979; and Gerking and Weirick, 1983). Therefore, identical workers with identical worker characteristics, education, experience and other determinants of human capital can earn different hourly wages depending upon the city/region/location in which they supply their hours to work. This is a divergence from the neoclassical perspective that wages equalise amongst workers in the long-run as workers are heterogeneous in their skills and returns on observable characteristics differ across metropolitan areas.

2.4 Empirical Evidence on Wage Differences across Metropolitan Areas

Various studies have examined the wage differentials across metropolitan areas and the theories that cause wage differences amongst workers. However, the majority of empirical evidence relating to wage differences across metropolitan areas has been limited to the US, UK and Australia. Not much research has been conducted on the topic in SA.

Fuchs (1967) examined hourly earnings differentials in the United States by regions (South and the rest of the United States) and by the size of the city using data from the Census of Population for the year 1960. Differences in observable characteristics such as age, race, gender and education were identified amongst the factors that resulted in such differences across regions. The author aimed to present new estimates of regional wage differentials which are based on average hourly earnings of non-agricultural persons.

Results presented by the author suggested that earnings were significantly lower in the South relative to the rest of the United States and these differences earnings appeared to be greater for non-white workers than white workers. This is because non-white workers possessed lower skill levels in both regions and the regional differences widened with lower skills levels regardless of the race. Results further suggested that average earnings in the rest of the United States were 25 percent higher relative to the South region. Apportioning this difference, one third of this difference could be attributed to observable characteristics, one third of the differential to the size of the city and one third of the difference remained unexplained after controlling for the composition of labour and the size of the city. This paper supports the study's theoretical explanation that differences and returns on observable characteristics have an effect on differences in earnings across regions and metropolitan areas.

Kiefer and Smith (1977) examined the impact of institutional rigidities in the form of unions and wage discrimination across South and North regions of the United States. In executing this study, the authors used data from the May 1973 Current Population Survey. The authors argued that differences in ways of life, traditions, institutions, and history may have differently affected the distribution of wages and prices across regions. Workers may have performed similar work but the distribution of earnings varied amongst workers living in the South and North regions of the United States. Furthermore, institutional rigidities in the form of unions

and discrimination by sex or gender may have had different effects on earnings across regions. Given these views, the authors sought to examine the effect of unions and discrimination on wages. They also examined whether or not the effect varied across the South and North regions of the United States.

In doing so, the authors estimated the wage equation and their results suggested that there was variation in earnings across the South and North regions. The impact of unions and discrimination by race or gender varied across regions. Union estimates suggested that the wage differential amongst workers was higher in the South than in the North and this was because the North region was highly unionised. Amongst white workers, the effect of unions on wage differential was minimal, reported to be one and a half percent. Discrimination estimates suggested the effect of discrimination was deeper in the South regions than the North regions. However, within each region, discrimination appeared to be prominent and discrimination was more often practiced against female's workers than male workers.

Contrary to the above two papers, Gerking and Weirick (1983) examined compensating differences and interregional wage differentials in the South and North regions of the United States. Their research was similar to that of Kiefer and Smith (1977) except that they used the Theory of Compensating Differences in Wages. The authors used data from the Panel Study in Income Dynamics (PSID) for the year 1976. They argued that under perfect competition conditions, wages amongst workers with similar observable characteristics, perfect mobility, similar preferences, who worked and lived in the same location, would be identical in the labour market. If this was not the case, then this would contradict the Theory of Compensating Wage Differentials under perfect competition.

In executing this study, the authors estimated the hedonic wage equation using Ordinary Least Squares (OLS) as a model of estimation. Findings presented by them were consistent with the Theory of Compensating Differences such that after controlling for the cost of living, workers with the same amount of human capital, the same work environment, and the same personal characteristics earned approximately the same regardless of where they lived. This result differed from the expectations and with the previously examined paper mainly because of the treatment of cost of living differences across regions, the treatment of hour supplied by workers and the completeness of the wage equation. However, the authors suggest that

interregional wage differences are the result of varying relative endowments of various heterogeneous labour types. This paper challenged the conventional view that wages did not equalize across regions or metropolitans and this was based upon the authors' treatment of variables that affect hourly earnings.

Johnson (1983) examined the nature and causes of wage differentials between large metropolitan cities within the US. In executing his study, he first examined the area's effect on the nominal wage for different types of workers which included male and female full-time workers, hourly and salaried. Johnson (1983) used data from the May Current Population Survey (CPS) for 1973-1976. He emphasised that earnings in the South of the US are lower relative to other areas in terms of nominal wages but that there were no significant differences in terms of real wages. Johnson (1983) also found that there were significant differences in earnings in for example an individual white worker residing in Boston who was a non-union member earned ten percent less than a white worker residing in Detroit. He sought to examine the cause of these differences and based his research question on whether wage differences in the US were a result of a disequilibrium phenomenon: whether differences in earnings were caused by unionisation or the compensating differentials of city amenities.

Johnson (1983) proposed that the wage equation of an individual worker was assumed to depend upon skills, compensating, discrimination and rent variables. He allowed controls for variables such as union membership, the human capital variable, race, and public sector employment. On the one hand, the results suggested that wage differentials in real terms amongst women were less than the differences in nominal terms. On the other hand, wage differences in real terms amongst male workers were found to be greater than wage differences in nominal terms. It was concluded that individuals were more likely to accept non-union employment at a low wage in an area that was largely unionised because the long-term income prospects were greater if they resided in areas with a greater probability of yielding high paying jobs. Results further showed that there was a variation in earnings across metropolitan areas of the U.S and that union spill-over effects had significant impact on wages. This showed that institutional rigidities can be contributing factors in the differences in earnings amongst workers depending upon the metro. This paper clearly suggested that earnings differences across metros depended largely on the conventional explanations for wage differences such as gender, unionisation and compensating wage differentials.

Azzoni and Servo (2002) examined the effect of education, cost of living and regional wage inequalities in Brazil using data from the years 1992, 1995 and 1997 collected by the Brazilian official Statistics agency. In their paper, they argued that Brazil was one of the developing countries that suffered from wage inequalities. Differences in earnings across regions were high and there were minimal prospects that disparities would be eliminated in the foreseeable future. The authors aimed to analyse wage differences and the extent to which control variables such as personal and labour market factors were able to explain these differences in earnings across 10 Brazilian metropolitan areas in the years 1992, 1995 and 1997 respectively. In doing so, the aim was to verify whether or not wage differences across the ten metropolitan areas were significantly lowered when controls for personal, cost of living and labour characteristics were introduced into the model. If workers were perfectly mobile or flexible to move from region to region, then why would they accept a lower wage in their respective region if they could earn a higher wage in other regions? The same question was raised by the authors with regards to capital in the sense that if capital was perfectly mobile to move from high wage regions into low wage regions, then wages would equalise within the labour market. However, as suggested by theory which has been outlined in this study, wages do not equalise as a result of the imperfectness of the labour market.

The authors estimate the Mincerian equation, where the dependent variable is the logarithm of hourly earnings and the independent variables are a host of personal, educational and labour market factors. They estimated the model using the Restricted Least Squares. Results suggested that wage differentials were slightly reduced when including control variables but inequalities across the ten metropolitan areas remained high. The magnitude of the difference between the two extremes was approximately 42 percent. This is a clear indication that rigidities in the form of unions in the Brazilian labour market were still prominent enough to suggest such prevailing wage differences for more than two decades.

Recent empirical work by Motellon *et al.,* (2011) on regional wage differentials suggested that wages in real terms do not differ significantly in macro-regions in the US, but rather average real wages differed due to the heterogeneity of worker attributes. Estimates found by Farber and Newman (1987), showed that differences in returns are as important as regional wage differentials given the heterogeneity of workers within the labour-market. Attention in these studies has primarily been focussed on the situation in the US due to severity of the issue and

the availability of micro data in such countries. This does not necessarily imply that wage differentials are negligible in other regions. The authors suggested that in most European countries, wage differentials were found to be more stable than the US.

Farber and Newman (1987) aimed to provide more insight and literature on the issue of regional wage differentials for the autonomous communities and cities (Nomenclature of Territorial Units of Statistics (NUTS II)) of Spain using micro data from a matched employer-employee dataset for the 1995 and 2002 waves. In doing so, they decompose the difference in earnings between low-income and high-income regions into differences in region characteristics and into regional differences in their prices. They did not just examine the wage gap but rather attempted to report on the entire pattern of the distribution of wages regardless of the region. By analysing wage patterns regardless of the regions, the authors expected the effects such as the ones related to the agglomeration economies to stand out as this would take into account the economic performance. Furthermore, they controlled for wage determinants (personal, education and labour market factors) but separated the effects of these determinants into those which were caused by human capital and those caused by job-firm characteristics.

Results suggest that workers with more human capital were major contributors of wage differential between regions. Results suggested that wage differentials existed in Spain in terms of average wage and with regards to other wage distribution characteristics. The result further indicated that these wage differences were not a factor of compensating wage differentials or structural variation but rather were due to regional characteristics given that workers with identical jobs working in identical firms comprise different earnings depending upon the region in which they were located. Therefore, after controlling for job and firm characteristics, the results showed that the equalisation of human capital does not imply that wages would be equated across regions given that the returns on human capital differed across regions. If human capital had been compensated in regions with low levels of income, then a greater deal of wage disparities would have been eliminated. In this context, the authors suggested that the variation in the return to human capital across regions was a policy implication in the sense that the return to human capital in low-income regions was less than the return to human capital in high-income regions. This raises the question of policy interventions by government in promoting human capital investment in low-income regions.

The authors provided an alternative approach which generally implied a focus on the reward obtained by skilled workers in low-wage regions. Returns on human capital persistently differed across regions as a result of economies of density and the potential of regional markets. Overall, this study appealed to all countries to provide regional policies that aimed to promote investments in human capital in an attempt to eliminate wage disparities.

More recent work by Mallik *et al.,* (2014) examined whether or not employability and earnings returns produced similar outcomes in metropolitan and regional labour markets. Their arguments were that the probability of being employed and returns on earnings differed across metropolitan areas and regional labour markets. This was in support of earlier arguments raised by Motellon *et al*., (2011). However, they differed in policy implications as they suggested that the more diverse a regional labour market was internally, and by comparison with other types of labour markets, the more likely it was that policies would perform with greater success when they addressed labour market challenges. This line of argument was supported by the results presented by the authors.

The probability of being employed and returns on earnings is dependent upon a range of factors which are considered to be important determinants of both employability and the returns on earnings. The factors include personal, educational, labour market factors and migration factors. This paper uses the human capital developed by Mincer (1958 and 1974), Becker (1964), and Becker and Chiswick (1966) as a theoretical framework. This paper provides a more detailed and comprehensive analysis than is usually possible especially with regards to factors such as education and ethnic background by providing a regional element to the analysis of wage differentials.

Mallik *et al.,* (2014) used micro-level data from the 2006 Census Sample File (CSF), which is available through the Confidentialised Unit Records Files (CURF) released by the Australian Bureau of Statistics (ABS) for a range of regions (ABS, 2006). In estimating the probability of being employed, the authors used the standard Probit Model (given that the dependent variable is dichotomous) and when estimating wage equations, the authors used Ordinary Least Squares (dependent variable is continuous). With regards to the wage equation which was estimated via OLS, the authors acknowledged the possibility of bias results which comes with sample selection bias. As a result, they further estimated the wage equation using the

Heckman Model which corrects for sample selection bias. They further included a dummy variable for area to identify whether significant differences existed between the two areas. Estimating the model with a dummy variable for area type separated the metropolitan city and the rural area.

The results suggested that there were significant differences in earnings across metropolitan areas and regional labour markets. Results presented by the authors supported empirical evidence but varied across metros and regional labour markets particularly with respect to personal characteristics, educational and labour market factors. This variation suggested that the two labour markets were different in their demand and supply of labour. The authors further suggested that in New South Wales the response to labour market outcomes (employability and wage equation) were diverse, suggesting that policies implemented by government to eliminate labour market challenges would be effective. This argument was based on the idea that the more diverse a regional labour market internally, the more likely it was that policies would perform with great success. However, place-based policies need to vary when catering for internal and external differences.

2.5 Conclusion

The neoclassical paradigm views the labour market as being perfectly competitive, thus implying wages across labour markets equalise. However, multiple wage differential theories deviate from the neoclassical view that wages equalise amongst workers across labour markets/regions and metropolitan areas. Therefore, the idea that wages will be equalised under perfect competition has proven to be just an illusion. The labour market is made up of heterogeneous firms and workers with no evidence that the labour market is imperfect in actual terms.

Apart from this view, wages differ across metropolitan areas for various reasons. Some metropolitan areas have the capacity to attract workers with high value human capital compared to other metropolitan areas. Returns on observable characteristics across metropolitan areas vary as some metropolitan areas have higher returns on earnings whilst others have lower returns on observable characteristics. Another factor that may lead to differences in earnings across metropolitan areas is the size of the metropolitan area and the industrial structure within each metropolitan area. Large metropolitan areas are associated

with higher earnings whilst small metropolitan areas are associated with lower earnings. This can therefore be a contributing factor to differences in earnings. Industrial structures may pay differently across metropolitan areas and industrial structures within metropolitan areas may pay high hourly earnings relative to other industrial structures in other metropolitan areas. This may result in differences in earnings across metropolitan areas.

Empirical evidence presented in this study suggests that wages differ across labour markets/ regions or metropolitan areas. These differences are explained by varying returns on observable characteristics that differ with respect to labour markets, regions or metropolitan areas. Recent empirical evidence presented by Azzoni and Servo (2002), Motellon *et al*., (2011) and Mallik *et al.* (2014) is consistent with this view. However, the economic implications of wage differences across labour markets, regions or metros drawn by these authors differ. Azzoni and Servo (2002) suggested that the presence of differences in earnings across labour markets or regions is a clear condition showing the effect that rigidities have on the structure of earnings. The second implication presented by Motellon *et al.* (2011) is that the varying regional returns on investments in human capital alter government policies in low income regions which promote the investment in human capital. The last implication is presented by Mallik et al. (2014) which suggests that the more diverse labour markets and regions or metros are, the more likely it is that government policy will be effective in the existence of labour market challenges.

CHAPTER THREE: DATA AND DESCRIPTIVE STATISTICS

3.1 Introduction

Discussing data and performing descriptive statistics before making any inferences is crucial because it enables the study to describe, analyse, and synthesise the structure of the data which constitutes the sample examined in this study. This chapter summarises the distribution of features relating to the study such as demographics, labour market factors and univariate analysis of the distribution in hourly earnings across Jhb, DBN and CT. In doing so, the chapter is divided into three parts. The first part provides a concise discussion on the data to be used in this study, how it is computed, how it is structured, what type of questions were asked, and how it is relevant in complying with the intended objectives of the study. The second part examines how the sample is spread out across the three metropolitan areas, how critical factors such as race, gender, and other labour market factors vary in proportions and how there might be potential changes over the periods across the three metropolitan areas. Descriptive statistics are used to provide analysis. The third part examines the distribution of hourly earnings across the three metropolitan areas by means of summary statistics and the use of a Kernel Density Plot. Findings provided by summary statistics are then compared to the Kernel Density Plot. The conclusion drawn in this last section lays the foundation for a multivariate analysis of Chapter Four.

3.2 Data

This study uses the Labour Force Survey (LFS), which is a household-based sample survey undertaken by Statistics South Arica (StatsSA). The LFS collects data relating to labour market activities of individuals aged 15 and above who live in South Africa. In relation to labour market activities, the LFS is useful to this study as it collects information about the level and pattern of unemployment as well as the industrial and occupational structure of the economy of SA. In capturing personal characteristics, respondents were asked questions about factors such as age, name, marriage status, location, race etc. When capturing educational characteristics, respondents were asked questions about their highest level of education obtained. Respondents were also asked questions about whether they were working part-time or full-time, the hours they usually worked, which sector/industry of employment they

were in, what are their monthly earnings. These questions helped to capture labour market factors to be examined in this study. The nature of this data makes it feasible for the study to make objective conclusions on the observed heterogeneity in hourly earnings across the three metropolitan areas. Therefore, the LFS data is relevant and consistent in helping the study in answering the research question and in carrying out the objectives of the study.

In previous years, the LFS data was collected as a biannual household survey but it is now collected on a quarterly basis. This study will, however, focus on the LFS of 2004, 2005, 2006 and 2007 which has been collected biannually. There are two major statistical controls imposed by the study that may lead to small sample size. Firstly, the study is restricted to respondents who fall within the working ages (15 - 65 years). Secondly, the study aims at examining differences in hourly earnings across Jhb, DBN and CT, so it is assumed that workers who live in any of the three respective metros also work in these metros. The above statistical controls imposed by the study based on the working age may result in a small sample size and poses serious implications as it can lead to models having low statistical power. There are three factors which can lead to unreliable findings in models with low statistical power. Firstly, the low probability of finding true effects, secondly the exaggerated estimate of the magnitude of the effect when a true effect is discovered and thirdly a low positive predictive value when an effect is claimed (Button *et al*., 2013). Given statistical properties to be imposed in this study, using only a single LFS dataset will have serious implications on the results presented by the study. The study pools biannual data from the 2004 LFS to the 2007 LFS. Each LFS to be pooled was designed as a rotating panel of dwelling units. Pooling data from LFS thus blends elements of both cross-sectional and time series data in a sense that cross-sectional data describes observations N and time-series describes the observation across time T to produce a set of $N \times T$ (Dielman, 1983). Pooling of data in this study will serve as a potential remedy for dealing with a small sample size as it accumulates or aggregates data, enabling the study to make objective comparisons in differences across populations (Verma, Gagliardi and Ferretti, 2009). Therefore, the study will have four waves, where wave one represents 2004, wave two represents 2005, wave three represents 2006 and wave four represents 2007.

In the next section descriptive statistics are presented to show how the data is dispersed across the three metropolitan areas in accordance with specific factors such as race, gender, industry representation, full/part-time employment and union membership. Furthermore, the

next section will show how hourly earnings are distributed across the three metros by means of summary statistics and the use of the Kernel Density Plot.

3.3 Descriptive Statistics

The descriptive statistics are weighted values calculated for employed individuals aged between 15 and 64 (belonging to the working age population), who were reported to live in Durban, Johannesburg and Cape Town at the time of the interview. The study assumes that individuals both live and work in their reported metropolitan area. Values presented in all the tables are in percentage terms. According to StatsSA's 2011 Census, SA is approximated to have a population of 51.8 million people. Of the total population, 22.41 percent live in the three respective metros. Amongst the three respective metros, Jhb has the highest population followed by CT and DBN respectively as approximately 8.6, 7.2 and 6.7 percent live in these respective metros. (StatsSA-Census, 2011). One potential reason for this is that the Jhb metro is South Africa's biggest and most economically vibrant area. It has been regarded as the New York of Africa as it is dominant within the continent in terms of scale, has the largest stock exchange, has corporate vibrancy, and has the largest financial service, media and culture (Rogerson, 2009). According to StatsSA (2011), the majority of individuals living in these three respective metros fall within the working age bracket of 15-64 years as 72.7, 69.6 and 70 percent of individuals living in the respective metros fall within the working age. Metropolitan areas have the potential to attract people of working age due to their economic structure with factors such as the industry structure, labour market characteristics, economic diversity and operating costs (Blumenthal *et al*., 2009).

3.3.1 Race and Gender Analysis

In this section, the study analyses the emergent descriptive statistics by examining how employed workers of the different four race groups are dispersed across the three metropolitan areas over the years under study. South Africa as a whole is a country highly populated predominantly by the African race and the expectation is that the African race will be intensely represented across the three metropolitan areas in all four waves. The reverse can be said with the White race group, as over the years it has been deemed as the minority race group in SA. On the one hand, the general expectation is that Indians will be highly popularised in DBN. This is linked to the 17th Century history of the immigrant Indians in South

Africa. They were in high demand as labourers in sugarcane fields because the European farmers regarded them to be cheap and reliable labour (Kuper, 1960). With the development of the city of Durban, these labourers left the fields and migrated to the city. The expectation is that CT is highly popularised by Coloureds due to the colonisation of the Dutch and inter-racial relationships that existed within the Western Cape Province (Seekings, 2011). This is in support with Brock's (1949) view that coloured individuals have their origins in CT primarily because slaves in the form of Hottentots, Bushmen and Europeans were brought to CT thus inter-racial relationships formed the coloured race. Table 1 below presents the distribution of employed workers by metro and race.

Table 1: Distribution (%) of workers by race and metropolitan area

	Wave 1	Wave 2	Wave 3	Wave 4
Durban				
African	63.03	59.48	60.64	64.76
Coloured	2.81	2.52	2.94	3.82
Indian	21.98	22.95	21.51	20.17
White	12.18	15.05	14.92	11.25
Cape Town				
African	23.31	28.69	27.13	25.38
Coloured	45.67	45.94	43.31	44.1
Indian	0.5	1.05	2.16	1.04
White	30.51	24.32	27.41	29.48
Johannesburg				
African	71.48	73.2	76.56	78.11
Coloured	4.97	4.55	3.48	2.52
Indian	2.58	3.05	2.71	3.95
White	20.98	19.2	17.24	15.41

Table 1 above shows the racial distribution of workers across the three main metropolitan areas in SA by wave (year). Given that the African population group is the largest in South Africa, it is not surprising to find that this population group comprises the majority of workers in both Durban (around 62 percent, on average across the waves) and Jhb (approximately 75 percent, on average across the waves). In Cape Town, more than 4 out of every 10 workers are Coloured, while in Jhb and DBN Coloured workers comprise less than 5 percent of workers on average. In DBN, around one-fifth of the population is Indian, while in Jhb and CT Indians are in the minority (less than 4 percent on average). Whites constitute between 11 and 30 percent of workers in each of the provinces, with the largest, but declining, population residing in CT.

Table 2 below provides descriptive statistics on gender employment for workers who reported being employed within the three metropolitan areas. By analysing the table below, it is evident that a greater percentage of male workers are employed in the workplace than female workers across the three metropolitan areas at all waves. At wave two, statistics for the CT metro show that 53.68 percent of employed individual workers are male and the remaining 46.32 percent are female workers. On the other hand, DBN and Jhb statistics suggests about 60.97 percent and 58.8 percent are individuals who reported to be male respectively. By comparing male and female employment, it is clear that male employment is higher than female employment across the three metropolitan areas and this result serves as an indication that gender biased employment is still persistent.

Table 2: Gender Employment across Metros by Wave

	Wave 1	Wave 2	Wave 3	Wave 4
Durban				
Male	57.75	60.97	58.54	56.92
Female	42.25	39.03	41.46	43.08
Cape Town				
Male	52.77	53.68	52.14	53.5
Female	47.23	46.32	47.86	46.5
Johannesburg				
Male	62.56	58.8	60.15	57.31
Female	37.44	41.2	39.85	42.69

Within the context of this study, gender biased employment could be due to nature of occupations existing in the three respective metros or due to more male workers migrating into metros. In SA, male workers are considered to be heads of the household and they often migrate to metropolitan areas for better employment opportunities and higher earnings (Stapleton, 2015). With respect to occupation, empirical evidence in the United States suggest that blue collar jobs (craft and operatives) which are concentrated in metros, are dominated by men whilst women work in clerical and service occupations (Gabriel and Schmitz, 2006).

3.3.2 Labour Market Analysis

Thus far, the study has analysed how individuals are distributed in terms of metro representation, race distribution and gender employment across the three metropolitan

areas. In this section, the study examines how the individuals are distributed across the three respective metros with respect to industry, union membership and whether or not the worker works full-time or part-time. Table 3 below presents the distribution of individual workers with respect to the industry they work in across the three metropolitan areas at all waves. In doing so, the study only provides analysis for industries that are highly represented in metropolitan areas. As a result, the agricultural and mining industries were not part of this analysis as they are not located in metros, but are located in rural areas.

Table 3: Distribution of Workers by Industry across the Three Metros

Durban	**Wave 1**	**Wave 2**	**Wave 3**	**Wave 4**
Manufacturing	22.01	21.36	20.07	19.01
Construction	5.73	6.41	6.11	5.79
Wholesale	22.01	25.23	27.81	26.03
Transport	7.63	7.74	7.11	6.47
Finance	9.8	11.21	12.09	13.64
Community	17.3	18.83	17.83	20.66
Private Household	15.52	9.21	8.98	8.40
Johannesburg				
Manufacturing	14.21	14.08	13.86	16.63
Construction	8.65	9.24	9.27	7.55
Wholesale	23.16	26.73	29.64	27.02
Transport	5.17	4.48	5.35	6.78
Finance	21.57	18.74	18.16	17.83
Community	16.10	14.71	13.58	14.44
Private Household	11.13	12.02	10.13	9.74
Cape Town				
Manufacturing	19.74	17.96	18.95	16.72
Construction	7.76	9.46	7.95	10.41
Wholesale	23.15	26.95	27.13	26.98
Transport	4.94	5.39	5.15	4.4
Finance	16.10	14.37	15.32	14.52
Community	21.27	18.8	16.96	17.16
Private Household	7.05	7.07	8.54	9.82

Amongst the labour industries examined in this study, the wholesale labour industry appears to consist of a high proportion of individual workers supplying their hours to work across the three metropolitan areas at all waves. This sector is followed by the manufacturing industry, community/social service, the finance and the private household industry. However, the

finance industry appears to be dominant in the Jhb metro area and this can be attributed to the fact that Jhb (given its economic landscape) has been regarded as the economic hub of SA. Overall there is variation in the distribution of employment with respect to labour market industries across the three metropolitan areas. However, the majority of workers living and working in the respective metros supply their labour in the wholesale industry.

Table 4 below presents the statistics of union representation across the three metropolitan areas at all waves. Labour unions have historically been present in the labour market and given wage inequalities that have persisted over the years, labour unions have served the primary function of equalising the distribution of earnings across workers (David, 2001). However, looking at Table 4, the results suggest that the majority of workers across the three metropolitan areas at all waves are non-union members. Jhb has the lowest percentage of workers who reported being union members. On the other-hand, DBN has the highest proportion of workers on average who are union members, whilst the CT metropolitan area presented results similar to the DBN metropolitan area.

Table 4: Union Representation across Metros by Wave

	Wave 1	Wave 2	Wave 3	Wave 4
Durban				
Union Member	33.5	45.68	30.0	31.87
Non-Union	66.5	54.32	70.0	68.13
Cape Town				
Union Member	25.55	32.29	23.66	28.26
Non-Union	74.45	67.71	76.32	71.74
Johannesburg				
Union Member	17.65	16.56	20.11	26.73
Non-Union	82.35	83.44	79.89	73.27

Low union representation in the metros is most likely because the types of industries which predominate in these metros are not those which are traditionally unionised. According to Hirsh (2008), unions are predominantly effective in industries such as mining, manufacturing, transportation, communication and utilities. From Table 3, it was clear that the majority of workers living and working in the respective metros work in wholesale, manufacturing, finance and private households.

Turning attention to full-time or part-time work, economic theory has postulated over the years that working full-time implies more hours to work which is in equal proportion to rising hourly earnings (Hirsch, 2004). It has been ideal for workers over the years to demand full-time jobs compared to part-time jobs mainly because part-time jobs are penalised with lower hourly earnings, fewer opportunities for career advancement and less job security compared to working fulltime. However, these factors come as a trade-off mainly because working part-time comes with more friendly working hours, better working conditions and less stress (Aaronson and French, 2004).

Furthermore, empirical evidence suggests that workers working full-time supply more hours to work relative to workers working part-time (Eissa and Hilary, 2004). However, this does not necessarily imply that full-time workers earn higher hourly earnings than their part-time counterparts. Looking at SA, it is legislated by the Basic Conditions of Employment Act of 1997 that in most sectors workers working part-time should be paid more per hour such that there is a premium to part-time employment that persists once differences in observable characteristics between part-time and full-time workers are accounted for (Posel and Muller, 2008). Table 5 below presents some descriptive statistics that show the difference in percentages of workers employed full-time and part-time across the three metropolitan areas by waves. These results justify the economic expectations as more individuals are reported to be full-time workers compared to being part-time across the three metros in all waves.

Table 5: Distribution of Fulltime and Part-time Workers across Metros

	Wave 1	Wave 2	Wave 3	Wave 4
Durban				
Fulltime	94.15	96.12	92.82	92.82
Part-Time	5.85	3.88	7.18	7.18
Cape Town				
Fulltime	90.34	90.52	87.8	86.66
Part-time	9.66	9.48	12.2	13.34
Johannesburg				
Fulltime	92.19	90.16	90.45	91.99
Part-time	7.81	9.84	9.55	8.01

This study has used descriptive statistics to show how employed individuals are represented across the three respective metros with respect to labour outcomes. Results presented

through descriptive statistics suggest there is variation in labour outcomes across the three metropolitan areas over the years. This in turn will affect the distribution of hourly earnings for each individual worker living in any of the three metros. Therefore, the variation presented through descriptive statistics makes it feasible for the study to show whether or not there are differences in hourly earnings across the three respective metros at all waves. Before running estimations, the study presents summary statistics of hourly earnings for the three respective metros as shown in Table 6 below and further depicts the distribution of hourly earnings through the Kernel Density Plot, which gives a visual representation on how earnings differ across the three metros.

3.3.3 Univariate Analysis: Summary Statistics and the Kernel Density Plot

Given the analysis provided in the previous discussion, the expectations are that there are differences in hourly earnings between the three metros. Jhb is expected to show higher hourly earnings compared to the other two metros given that Jhb is considered as the economic hub of SA (Rogerson, 2009). Table 6 below presents summary statistics on the mean hourly earnings for the three metropolitan areas. By analysing the summary statistics of hourly earnings presented in the Table 6 below, it is clear that Jhb has the highest hourly earnings compared to CT and DBN, and DBN has the lowest average hourly earnings compared to Jhb and CT. By comparing CT and Jhb estimates, for instance, hourly earnings are slightly different as Jhb workers earn R22.96 an hour, on average whilst CT workers earn R22.62 an hour on average a difference of R0.34.

Table 6: Summary Statistics of Hourly earnings

	Hourly earnings by Metro (Jhb, DBN and CT)	
	Mean Wage	**Std Dev**
Durban	20.96	30.26
Cape-Town	22.62	27.02
Johannesburg	22.96	35.62

To further elucidate differences in earnings across the metros, a Kernel Density Plot is used. The Kernel Density Plot is able to draw out differences in the earnings distributions of each metro, and so is superior to the analysis of means presented in Table 6.

Figure 1: Kernel Density Plot of Hourly Earnings

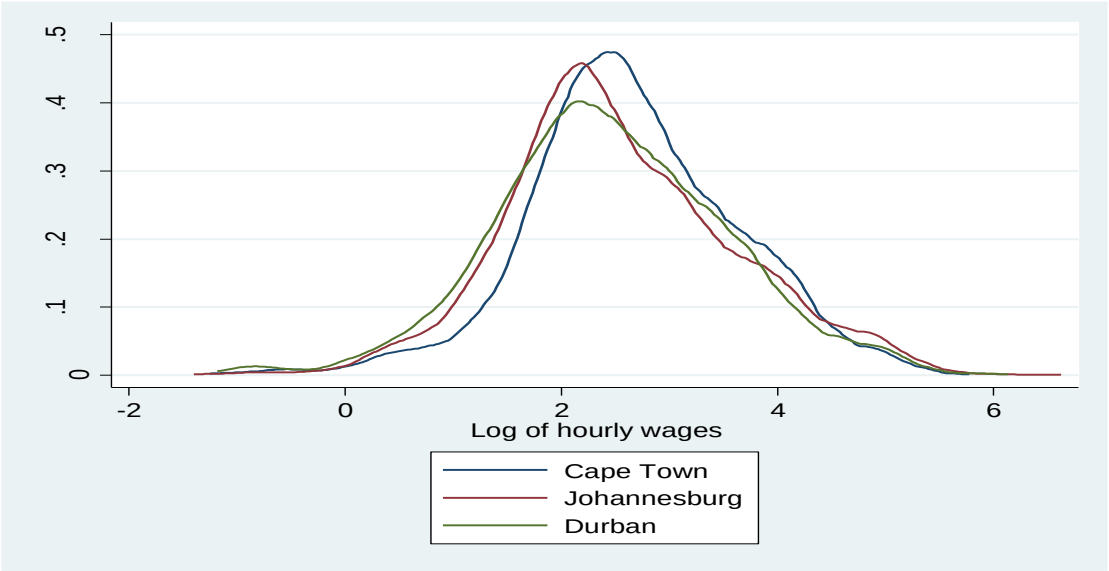


Figure 1 above shows the distribution of the logarithm of hourly earnings in each metropolitan area as estimated using a Kernel Density Plot. All three distributions are negatively skewed. In DBN, the earnings distribution is flatter than for CT and Jhb, suggesting a greater variance in the logarithm of hourly earnings in DBN than in CT and Jhb. The earnings distribution for CT lies mostly to the right of that for Jhb and DBN, indicative of higher earnings, on average, in this metro. However, the upper tail of the earnings distribution for Jhb lies above that for both CT and DBN, which is indicative of there being a larger number of very high income earners in this metro.

3.4 Conclusion

This chapter has used data from the pooled LFS from 2004 to 2007. The LFS data is a household-sample survey that collects data relating to labour market activities of individuals aged 15 and above who live in South Africa. The study imposed statistical controls such as observing individuals living and working in the respective metropolitan areas, employed and aged between 15 and 64. Results of the descriptive statistics analysis suggest that the majority of individuals are African and the minority are White. African workers are mostly located in Jhb and DBN at all waves, whilst White workers are mostly located in CT. As expected, male employment is higher than female employment across the three metropolitan areas and this result serves as an indication that gender biased employment is still common.

Given the economic nature of the three respective metros, labour market factors relating to industry suggest that majority of the workers work in manufacturing, wholesale, finance and community industries across the three metropolitan areas at all waves. Furthermore, labour market factors relating to union membership and full-time employment suggest that the majority of workers are non-union members and the majority of the workers work full-time. This is in line with economic expectations as these large cities have the capacity to attract high skilled workers (less incentive to join unions) in high-tech jobs that pay high earnings which is deemed as an incentive to work fulltime. The study further made use of summary statistics in an attempt to show hourly earnings differ between the metros. This univariate analysis suggests that differences exist but they appear to be minimal and unclear. The Kernel Density analysis revealed that differences in the earnings distribution exist particularly between Jhb and DBN, and DBN and CT. In contrast, there are minimal differences in the earnings distribution between CT and Jhb.

The next chapter explores methods of estimating the wage equation and uses the Oaxaca Blinder Decomposition to show how differences in earnings occur across the three metropolitan areas. In doing so, the next chapter examines the four different approaches of which are used in examining differences in earnings across the three metropolitan areas.

Chapter 4: Econometric Methodology and Estimation Analysis

4.1 Introduction

The descriptive statistics presented in Chapter 3 reveal that there are minimal differences in hourly earnings across the three metropolitan areas. However, the problem with a purely univariate analysis is that it does not account for other multiple factors that may affect average earnings across the three metropolitan areas. The study must use a multivariate analysis to account for the multiple factors that affect average earnings. These factors may include such as observable characteristics as age, gender, education, labour market factors.

This chapter discusses the methods which the study uses to explain and analyse differences in hourly earnings across the three respective metropolitan areas using multivariate analysis. This enables the study to provide informed analysis on how hourly earnings differ across the three metros and how other factors (personal factors, educational factors and labour market factors) which determine earnings also contribute to differences in hourly earnings across the three metropolitan areas. In doing so, the study uses four approaches to examine this effect.

4.2 Methodology

As discussed in the previous chapter, the study uses the labour force survey (LFS) pooled from a four year period (from 2004 to 2007). The study's conceptual framework is based upon the theory of human capital developed by Mincer (1974) and Becker (1975) on the understanding that investment in personal characteristics and educational factors ultimately increases hourly earnings. This study follows the work by Mallik *et al*. (2014), by modifying the Mincerian Model through the inclusion of explanatory variables that capture personal characteristics (age, gender, race, and marital status) and labour market factors (unionisation, sector of employment, type of industry). Personal and labour market factors are included on the basis that they directly affect the hourly earnings of workers across the three metropolitan areas. The inclusion of such explanatory variables will further improve the overall explanatory power of the model to be estimated.

The study uses four approaches to understand how hourly earnings are determined, how they differ and whether the difference in hourly earnings is due to city endowments or worker characteristics. In doing so, the first approach estimates hourly earnings against the dummy

variables of Jhb, DBN, and CT. Furthermore, given that the study uses pooled cross-sectional data, the first model also takes into consideration how hourly earnings differ with time. In doing so, this implies the study would include wave dummy variables that represent the different times. For the metro dummy variables, the study uses the Jhb metro as the base or references category and for the wave dummy variables the study uses the year 2004 as a reference category. Having both reference categories enables the study to make comparisons on how hourly earnings differ across the metros and across time as shown by equation (1) below.

$$lnY_{it} = \beta_{it} + \sum_{w=4}^{w} \theta_k{}' w_k + \sum_{k=3}^{k} \gamma_{kit}{}' D_k + \boldsymbol{\varepsilon_{it}} \quad (1)$$

Where; $lnY_i$is the dependent variable which is the log of hourly earnings; k; 1,2..:K being the dummy variables for DBN and CT (such that 1 if the outcome is true and 0 otherwise)and the Jhb metro dummy variable is used as the reference category; w: 1,2,..:W being the wave dummy variables for the years 2004, 2005, 2006, and 2007 respectively and the 2004 wave dummy variable is used as a references category (wave 1),and $\varepsilon_i$is the random error term. The purpose of estimating the above equation is to show how hourly earnings differ across the three metropolitan areas before taking into account the effect other explanatory variables (personal, educational and labour market factors) have upon hourly earnings. Equation (1) above is estimated under the assumption that other explanatory variables have no effect on hourly earnings. Given that the dependent variable (hourly earnings) is continuous and presented in log form, this further provides the base for the study to use Ordinary Least Squares (OLS) as a model of estimation.

The second approach modifies equation (1) by including explanatory variables that affect hourly earnings such that the assumption imposed in the first estimation that other explanatory variables (personal, educational, and labour market factors) have no effect on hourly earnings is relaxed. This enables the study to observe how (direction) and by how much (magnitude) the coefficient on the metro dummy variables change after controlling for other explanatory variables. This approach is conducted under the assumption that the returns on observable characteristics are the same across the three respective metropolitan areas.

The study further assumes that workers work and live in the respective metropolitan area of which they report living in. This assumption is made to avoid possible situations where an individual worker reports living in the CT metro but happens to be supplying their hours to work in the Jhb metro. Therefore, in the second approach, the study estimates the conventional model of hourly earnings that is produced by equation (2). This conventional model produced by equation (2) is modified to take into account other explanatory variables that affect hourly earnings as shown by equation (2) below.

$$lnY_{it} = \beta_0 + \sum_{k=4}^{k} \theta_k \, 'w_k + \boldsymbol{\beta'P} + \boldsymbol{\delta'E} + \boldsymbol{\alpha'l} + \sum_{k=3}^{k} \gamma_k \, 'D_k + \boldsymbol{\varepsilon_{it}} \qquad (2)$$

Where **P** Is the vector of personal characteristics, **E** is the vector of educational factors and **L** is the vector of labour market factors. The dependent variable for an individual worker within the labour market remains the logarithm of hourly earnings (shown on the left side of equation (2)) and the independent variables are given by the vector of personal characteristics (age, gender, sex, race, location, etc.), the vector of educational factors (primary to tertiary education), and the vector of labour-market factors (sector and industry of employment, occupation, union membership, etc.). Equation (2) represents a conventional earnings equation except that the study has modified it to take into account wave dummy variables that represent the different data sets generated in different years to obtain one pooled cross-sectional data set. Furthermore, empirical evidence conducted by researchers such as Goldfarb and Yezer (1976), Kim, Liu and Yezer (2009), George (1983), and Motellon *et al*. (2011) all supported the inclusion of such variables.

The third approach lays the foundation for the Oaxaca-Blinder Decomposition technique to be performed. In doing so, the third approach estimates separate regressions for each of the three metropolitan areas thus enabling the study to make objective comparisons on how estimated coefficients differ across the three metropolitan areas. In doing so, the study relaxes the assumption imposed earlier that returns on observable characteristics are identical across the three metropolitan areas. Therefore, this process allows the returns on observable characteristics to differ with respect to each metropolitan area. The study modifies equation (2) to account for this effect and as shown below by equations (3), (4) and (5) respectively, represent the estimation of hourly earnings with respect to each metropolitan area.

$$lnY_{it}^{CT} = \beta_0^{CT} + \sum_{k=3}^{k} \theta_k \, 'w_k + \boldsymbol{\beta}'\boldsymbol{P} + \boldsymbol{\delta}'\boldsymbol{E} + \boldsymbol{\alpha}'\boldsymbol{l} + \varepsilon_{it} \quad (3)$$

$$lnY_{it}^{Jhb} = \beta_0^{Jhb} + \sum_{k=3}^{k} \theta_k \, 'w_k + \boldsymbol{\beta}'\boldsymbol{P} + \boldsymbol{\delta}'\boldsymbol{E} + \boldsymbol{\alpha}'\boldsymbol{l} + \varepsilon_{it} \quad (4)$$

$$lnY_{it}^{DBN} = \beta_0^{DBN} + \sum_{k=3}^{k} \theta_k \, 'w_k + \boldsymbol{\beta}'\boldsymbol{P} + \boldsymbol{\delta}'\boldsymbol{E} + \boldsymbol{\alpha}'\boldsymbol{l} + \varepsilon_{it} \quad (5)$$

Estimations examined above will show that the returns on observable characteristics differ with respect to each metropolitan area. Once the study has examined and analysed the above differences in observable characteristics, the last approach (fourth approach) involves decomposing the mean wage difference between the respective metros using the Oaxaca-Blinder (O-B) Decomposition based on the results obtained in the third approach.

When examining differences in hourly earnings between groups particularly in the labour-market, it is a common practice that researchers examine the mean difference and whether the mean difference between groups is explained by coefficients, shift coefficients or differences between intercepts through the O-B decomposition method (Jones, 1983). The wage differential between groups (the three respective metros in our case) can be divided into explained and unexplained components (Oaxaca, 1973 and Blinder, 1973). The explained component measures productivity characteristics that can be observed such as personal characteristics (age or gender) and labour market factors (experience). The unexplained component is regarded as a residual part that is not accounted for by such differences in earnings determinants and this unexplained component is commonly regarded as a discriminating effect (Jann, 2008; Todd, Goddeeris and Haider, 2010). While much of the research using the O-B technique has analysed gender or race differences, this dissertation focuses on examining the mean differences in hourly earnings between the three metropolitan areas.

Following Ben's (2008) notation, the study decomposes the differences in hourly earnings between CT- Jhb, DBN - Jhb, and CT-DBN to find the mean outcome difference. Looking at the decomposition for CT-Jhb, for instance, this implies we want to find M;

$$M = \bar{Y}_{CT} - \bar{Y}_{Jhb} \quad (6)$$

Where $M$is the difference in the mean outcome for the observations in the CT and Jhb metros. In finding the mean outcome, the study will conduct a twofold decomposition on the mean difference of hourly earnings which divides the wage gap into a part that is explained and a part that is unexplained. The study performs the O-B Decomposition based on a linear regression and is separable in observable and unobservable characteristics of which the mean outcome is the expected value of the outcome variable for CT and Jhb such that:

$$\text{Group M} \in \{CT\ ; Jhb\} \text{ and} \in (\varepsilon_M) = 0 \quad (7)$$

This implies the linear regression model will take the form

$$\bar{Y}_M = \bar{X}'_M \hat{\beta}_M + \varepsilon_m \quad (8)$$

Where;$\bar{Y}_M$is the hourly wage rate ; $\bar{X}_M$is the vector of individual characteristics, $\hat{\beta}_M$ is the estimated coefficients and $\varepsilon_m$is the random error term. The assumption is that error terms are equal to zero. Therefore, by substituting equation (8) into equation (6) we obtain the difference for the mean outcome for both the CT and Jhb metro as shown by equation (9) below.

$$Y_{CT} - Y_{Jhb} = \bar{X}'_{CT}\hat{\beta}_{CT} - \bar{X}'_{Jhb}\hat{\beta}_{Jhb} + \varepsilon_{CT} - \varepsilon_{Jhb} \quad (9)$$

Arranging equation (9) above, the study obtains a threefold decomposition that is produced by equation (10) below which sums the difference in hourly earnings between the two metros into three parts; endowments, coefficients, and interactions.

$$Y_{CT} - Y_{Jhb} = (\bar{X}_{CT} - \bar{X}_{Jhb})'\hat{\beta}_{Jhb} + \bar{X}'_{Jhb}(\hat{\beta}_{CT} - \hat{\beta}_{Jhb}) + (\bar{X}_{CT} - \bar{X}_{Jhb})'(\hat{\beta}_{CT} - \hat{\beta}_{Jhb}) + \varepsilon_{CT} - \varepsilon_{Jhb} \quad (10)$$

Looking at the right of equation (10) the first part represents endowments, the second part represents the coefficient and the last part represents the interaction. The coefficient part

apportions the difference to be due to differences in coefficients, the endowment part apportions the difference to be due to different explanatory variables across the metros and the interaction part accounts for the possibility that explanatory variables may differ and coefficients may also differ across the two metros. However, by decomposing the results, the study aims to divide the wage differential between two groups into a portion that is either explained or unexplained. To do so, the study performs a twofold decomposition under the idea that there is a non-discriminating coefficient vector that should be used to determine the contribution of the differences in the predictors. In other words, the twofold decomposition technique decomposes the mean outcome difference with respect to the non-discriminating coefficient vector (Hlavac, 2014). This non-discriminating coefficient vector would serve to be a set of regression coefficients that would exist in an ideal world without discrimination. Given this notion, let β_k be the non-discriminating coefficient vector. The study assumes that the discrimination is directed towards Jhb, and that there is no positive discrimination for CT such that β_k=$\hat{\beta}_{CT}$. Re-arranging equation (10) and taking into account the non-discriminating coefficient vector, we obtain a twofold decomposition which apportions the difference in explained cross-group differences to explanatory variables and the other portion remains unexplained by the differences as shown by equation (11) below:

$$Y_{CT} - Y_{Jhb} = \left(\overline{X}_{CT} - \overline{X}_{Jhb}\right)' \beta_k + \overline{X}'_{CT}(\widehat{\beta}_{CT} - \beta_k) + \overline{X}'_{Jhb}\left(\beta_k - \widehat{\beta}_{Jhb}\right) + \varepsilon_{CT} - \varepsilon_{Jhb}$$

If discrimination is directed towards Jhb and there is no positive discrimination for CT, then the non-discriminating coefficient vector will be given by β_k=$\hat{\beta}_{CT}$, such that $\overline{X}'_{CT}\left(\hat{\beta}_{CT} - \hat{\beta}_{CT}\right)$ = 0, therefore:

$$Y_{CT} - Y_{Jhb} = \left(\overline{X}_{CT} - \overline{X}_{Jhb}\right)' \widehat{\beta}_{CT} + \overline{X}'_{Jhb}\left(\widehat{\beta}_{CT} - \widehat{\beta}_{Jhb}\right) + \varepsilon_{CT} - \varepsilon_{Jhb} \quad (11)$$

The above equation represents the twofold Oaxaca-Blinder Decomposition between the CT and Jhb metropolitan areas. Looking at the right-hand of equation (11), the first-part represents outcome differential which is explained by group differences and the second part represents the unexplained component of group differences that is attributed to discrimination. The second part shown above also captures the potential effects of differences

in unobserved variables as shown by the inclusion of a random error term (Jann, 2008). Furthermore modifying equation (11), leads to equation (12) and (13) respectively, which represents the twofold decomposition for Jhb-DBN (assuming that the discrimination is directed towards DBN, and that there is no positive discrimination for Jhb such that β_k=$\hat{\beta}_{Jhb}$) and CT-DBN (assuming that the discrimination is directed towards DBN, and that there is no positive discrimination for CT such that β_k=$\hat{\beta}_{CT}$) to be estimated by the study as shown below:

$$Y_{Jhb} - Y_{DBN} = (\overline{X}_{Jhb} - \overline{X}_{DBN})'\beta_k + \overline{X}'_{Jhb}(\widehat{\beta}_{Jhb} - \widehat{\beta}_k) + \overline{X}'_{DBN}(\widehat{\beta}_k - \widehat{\beta}_{DBN}) + \varepsilon_{Jhb} - \varepsilon_{DBN}$$

If discrimination is directed towards DBN and there is no positive discrimination for Jhb, then the non-discriminating coefficient vector will be given by β_k=$\hat{\beta}_{Jhb}$, such that $\overline{X}'_{Jhb}(\hat{\beta}_{Jhb} - \hat{\beta}_{Jhb})$ = 0, therefore:

$$Y_{Jhb} - Y_{DBN} = (\overline{X}_{Jhb} - \overline{X}_{DBN})'\widehat{\beta}_{Jhb} + \overline{X}'_{DBN}(\widehat{\beta}_{Jhb} - \widehat{\beta}_{DBN}) + \varepsilon_{Jhb} - \varepsilon_{DBN} \quad (12)$$

And

$$Y_{CT} - Y_{DBN} = (\overline{X}_{CT} - \overline{X}_{DBN})'\beta_k + \overline{X}'_{CT}(\widehat{\beta}_{CT} - \widehat{\beta}_k) + \overline{X}'_{DBN}(\widehat{\beta}_k - \widehat{\beta}_{DBN}) + \varepsilon_{CT} - \varepsilon_{DBN}$$

If discrimination is directed towards DBN and there is no positive discrimination for CT, then the non-discriminating coefficient vector will be given by β_k=$\hat{\beta}_{CT}$, such that $\overline{X}'_{CT}(\hat{\beta}_{CT} - \hat{\beta}_{CT})$ = 0, therefore:

$$Y_{CT} - Y_{DBN} = (\overline{X}_{CT} - \overline{X}_{DBN})'\widehat{\beta}_{CT} + \overline{X}'_{DBN}(\widehat{\beta}_{CT} - \widehat{\beta}_{DBN}) + \varepsilon_{CT} - \varepsilon_{DBN} \quad (13)$$

After estimating the above twofold decomposition for the above respective metros, the study should enable the reader to see how mean hourly wage differences are different between the three respective metropolitan areas according to the O-B decomposition estimation equations presented above. By doing so, the goal of running the O-B decomposition in this study is to separate the mean difference between the three metropolitan areas (Jhb-CT, CT-DBN, and

Jhb-DBN) to a portion that is explained by observable characteristics (endowments) and into a portion that is unexplained (residual or discriminating components). In other words, decomposing the mean hourly earnings between metros enables the study to single out causes of differences in hourly earnings. This method is useful in the sense that it provides an explanation for the difference between the two groups in a statistical sense which separates the difference to be explained by observable characteristics or discrimination (Fortin *et al*., 2011).

4.3 Possible Estimation Challenges

There are possible estimation problems that may be encountered when estimating the above approaches in the study's analysis of differences in hourly earnings. The first challenge is that of sample selection bias given that the sample used in this study only comprises employed individuals. Sample selection bias has been a concern in research, particularly in the application of econometrics in instances where researchers seek to analyse or estimate consumer expenditures and wage equations (Puhani, 2000). The study has imposed a statistical control that considers only individuals who are actively employed as the sample of the study. This statistical control is selective and restrictive in its approach as it only considers respondents who are actively employed, thus implying that the sample used to examine differences in hourly earnings across the three metros is non-random. Therefore, results presented through the OLS are likely to be unreliable or biased given the fact the dependent variable is observed only for a restricted non-random sample (Heckman, 1979).

Using data obtained from a non-random sample has its own risks for the analysis of the study given the fact that it leads to biased and inconsistent results obtained from OLS estimators. Because this study uses OLS as a method of estimation for the wage equations, then it follows that estimates will be biased upward as the sample used in this study is not representative of the whole population. Furthermore, limiting the study to individuals living in metros and not considering workers living in non-metros is another factor leading to sample selection bias. Heckman (1979) proposed a two-step maximum likelihood model which can be used to avoid the possibility of estimating the model from self-selected samples which treats this selection problem as an omitted variable problem (Vella, 1998 and Puhani, 2000). Models can be estimated via OLS if unobserved heterogeneity influencing the probability of being selected into the sample is not correlated with the unmeasured factors influencing the outcome

variable. Therefore, the study assumes that error terms are not correlated (ignore the selection equation) and the wage equation is estimated via OLS. It is worth noting that the O-B decomposition has a correction for sample-selection bias, which deducts the selection effect from the overall differential and applies it to the standard decomposition formulas to the adjusted differential (Jann, 2008). The O-B Decomposition runs a correction for sample-selection bias if the selection effect (lambda) is estimated in the initial regression. The study observes differences in hourly earnings only for workers who live in the respective metros.

Another challenge is that the study has assumed that workers live and work in any of the three respective metros. This assumption is likely to cause endogeneity bias as each individual of the sample living in one metro can be different in non-observable ways from another individual living in another metro. This unobserved heterogeneity is likely to be correlated with explanatory variables thus resulting in explanatory variables to be correlated with error terms. This effect may lead to bias and inconsistent OLS estimates and this form of endogeneity is caused by an omitted variable.

To remedy inconsistent OLS estimates as a result of omitted variables, the two-stage least squares through the instrumental variable technique can be used as a method of dealing with endogeneity bias (Angrist and Krueger, 2001). The instrumental variable allows for the estimation of simultaneous equations consisting of endogenous regressors (Antonakis *et al*., 2014). This technique requires the study to find a variable that is strongly correlated with the endogenous variable (instrumental relevance) but not correlated with error term (instrument erogeneity). If the instrument satisfies the two conditions than the instrumental variable will be consistent even though it is less efficient than the OLS (Bound, Jaeger and Baker, 1995). It is worth noting that the study acknowledges the likelihood of these estimation concerns. However, the study assumes that there is no omitted variable bias and there is no correlation amongst parameters (omitting one does not result in biased estimates of the other). This assumption can be somewhat be tenuous if the distribution of unobserved characteristics, for each year, differs across the three metropolitan areas. There are a number of dwelling level variables that may be correlated with selection into employment but that are not related to wages such as the presence of children in the dwelling or pensioners in the dwelling or transport costs. Therefore, it is commendable for future research that a selection model be used, as it will ensure that results presented are adjusted for selections bias.

Furthermore, there is not much variation in response to questions relating to earnings across the three metropolitan areas as majority of the respondent's main income is in the form of salaries/wages and they earn above the minimum wage. The study uses population weights in all estimations to avoid possible situations where some groups are over or under-represented because of non-response. This process rebalances the data to accurately reflect the population.

Overall, the study uses four approaches to examine and analyse differences in hourly earnings across Johannesburg, Durban and Cape Town. These four approaches seek to maintain that the study adheres to the objectives that were set in the first chapter. Given that the study has explored methods involved in executing the four approaches, then next step is to provide an analysis of the estimation results. Therefore, the next section provides a detailed analysis of the estimation relation for the four approaches used by this study.

4.4 Estimation Results (Specification 1 and 2)

Table 7 below presents the results based on the estimation of hourly earnings. As outlined earlier, the first estimation of hourly earnings controls only for the metropolitan area dummy variables and this estimation is regarded as model 1/specification 1. The purpose of this estimation is to show how hourly earnings differ across the three metropolitan areas.

Table 7: Estimation Results for the First and Second Specification

Independent Variables	Model 1	Model 2
Cape Town	0.1503***	-0.2661
Durban	-0.2128***	-0.2813***
Wave 2		-0.1017***
Wave 3		0.0788***
Wave 4		0.1637***
Personal Factors		
Coloured		0.2997***
Indian		0.4142***
White		0.5369***
Male		0.1726***
Age		0.0344***
Age-squared		-0.0003***
Married		0.1223***
Educational Factors		
gr1_gr7		0.0022
gr8_gr11		0.1443*

Matric		0.4014***
Diploma/degree		0.5452***
Labour Market Factors		
Union		0.1068***
Formal		1.2831***
Informal		0.8231***
Fulltime		-0.2661***
Construction		-0.9264**
Wholesale		-0.2384***
Transportation		0.0404
Finance		-0.0205
Community/Social Service		-0.0210
Private Household		-0.0534
Professional		-0.2040***
Technician		-0.3503***
Clerk		-0.5042***
Service/Sales workers		-0.8988***
Craft		-0.7215***
Plant		-0.7265***
Elementary		-0.8930***
Constant	**2.655*** **	**0.9676*** **
R-Squared	0.0170	0.5795
	(0.0000)	(0.0000)
F-statistic	53.07	

*** 1% level of significance; ** 2% level of significance; * 10% level of significance

As discussed earlier, the first specification includes metro controls only, while in the second specification a full set of explanatory variables is included. From the first specification, it is clear that significant differences in earnings exist between the metros. Individuals who work in CT earn on average 15.03 percent more than individuals in Jhb. In contrast, those in DBN earn approximately 21.28 percent on average less than individuals in Jhb. In order to identify whether these earnings differentials persist when observable differences between workers are accounted for, the second specification controls for a range of explanatory variables, including personal, educational and labour market factors.

Once observable differences between workers are taken into consideration, the second specification suggests that there is no significant difference in earnings between individuals in CT and those in Jhb. Controlling for observable differences between workers does, however, reinforce that those working in DBN earn significantly less, on average, than those in Jhb, with the coefficient suggesting a 28.13 percent difference.

In terms of the additional control variables included in the regression, all the estimated coefficients have the expected signs and are consistent with the results of other earnings regressions estimated for South Africa (Haroon et al., 2001; Burger and Yu, 2006; Leibbrandt *et al.,* 2010; Kerr and Teal, 2012). For example, coefficients on education are all positives suggesting that it positively affects hourly earnings, e.g. a worker with matric level of education earns 49.39 percent higher than a worker without schooling. Therefore, education positively affects hourly earnings and this result is consistent with Card's (1999) view that better educated individuals earn higher wages, decrease the probability of being unemployed and occupy prestigious occupations relative to uneducated individuals.

There are significant differences in earnings for each year particularly for the second specification at all levels of significance, where; hourly earnings were high in 2006 and 2007 as opposed to 2004. However, they were not significant differences in hourly earnings between 2004 and 2005. Furthermore, the study takes into account whether estimates reflect robust standard errors as they take into account issues concerning heterogeneity and lack of normality. by looking at the first and second specification, focusing on the metro dummy variables, estimates reflect robust standard errors.

One concern with the results presented in specification 2 is that the estimated model assumes that the returns on observable characteristics are the same for workers regardless of the metro in which they work and reside. It is likely, however, that observable characteristics may be rewarded differently across the metros. For example, the returns on tertiary education may be higher in Jhb than in Dbn, for instance, given that Jhb is the business hub of SA. To address this concern, the estimates presented in Table 8 are for regressions estimated separately for each metro. The dependent and control variables are the same as in specification 2 except for metro dummy variables as each specification is estimated with respect to each metro dummy variable.

4.5 Estimation Results (Third Specifications)

Table 8: Estimation Results for the Third Specification

	Dependent Variable: (ln Wage) (Ordinary Least Squares)

Independent Variables	Durban	Johannesburg	Cape Town
Wave 2	-0.321***	0.026	-.0097
Wave 3	0.051	0.090**	0.102***
Wave 4	0.058	0.241***	0.185***
Personal Factors			
Coloured	0.440***	0.419***	0.329***
Indian	0.316***	0.643***	1.259***
White	0.374***	0.617***	0.653***
Male	0.227***	0.076*	0.201***
Age	0.028***	0.037***	0.044***
Age-squared	-0.0002*	-0.0003**	-0.0004***
Married	0.0996**	0.110***	0.138***
Educational Factors			
gr1_gr7	-0.139	0.009	0.191
gr8_gr11	0.018	0.032	0.431**
Matric	0.277**	0.282**	0.671***
Diploma/degree	0.661***	0.812***	1.056***
Labour Market Factors			
Union	0.098**	0.142***	0.134***
Formal	1.102***	1.493***	1.055***
Informal	0.541***	1.036***	0.737***
Fulltime	-0.253***	-0.316***	-0.182***
Construction	-0.144*	-0.145**	0.012
Wholesale	-0.243***	-0.219***	-0.223***
Transportation	0.062	0.009	0.069
Finance	-0.028	0.047	-0.148**
Community/Social Service	0.139**	-0.089	-0.080
Private Household	-0.265	-0.027	-0.045
Professional	-0.154	-0.255**	-0.208*
Technician	-0.391***	-0.336**	-0.367***
Clerk	-0.541***	-0.573***	-0.398***
Service/Sales workers	-0.839***	-0.989***	-0.815***
Craft	-0.704***	-0.775***	-0.668***
Plant	-0.647***	-0.729***	-0.761***
Elementary	-0.828***	-0.940***	-0.879***
Constant	**1.168*** **	**0.850*** **	**0.5098**
R-Squared	0.5483	0.6188	0.5919
	(0.0000)	(0.0000)	(0.0000)
F-statistic	78.24	89.30	84.39

*** 1% level of significance; ** 5% level of significance; * 10% level of significance

Estimated coefficients on race across the three metropolitan areas vary and as expected all three of the other race groups earn higher hourly earnings than the African race. The coefficient on Coloured race suggests their hourly earnings are higher than the African race across the three metros. White workers living in CT on average earn 65.3 percent higher hourly earnings than African workers and this gap between the two race groups is higher in CT compared to the 62 percent and 37.4 percentage difference for Jhb and DBN respectively.

Looking at the factors of gender and marital status, empirical evidence suggests that the more the worker becomes productive within the workplace, the return to earnings likewise increase in equal proportion as productivity (Eissa and Hilary, 2004). The estimated coefficient on marriage is positive and significant across the three metros at significant levels above 5 percent. Therefore, an individual worker who is married earns higher hourly earnings than an unmarried worker.

Apart from race, gender, and marital status, there is a positive strong relationship between education and hourly earnings received by workers. However, there are no significant differences in hourly earnings between workers with primary/secondary education and those with no schooling across the three metropolitan areas. As expected, results further show that acquiring matric level of education increases hourly earnings compared to workers without matric. This result is true across the three metropolitan areas. By comparison, the coefficient on matric is higher in CT than it is for DBN and Jhb respectively. This implies that the gap in hourly earnings between workers with matric and those without matric is higher in CT than it is in DBN and Jhb respectively. Overall, the estimated coefficients on education appear to be consistent with empirical evidence which suggests that workers who invest in education increase their return to hourly earnings more than the workers without investments in education (Mincer, 1974). Results further suggest that workers who work and live in CT have a greater chance of increasing their hourly earnings by investing in education compared to workers working and living in DBN and Jhb respectively. This is based on the estimated coefficients on education as the CT metro has a larger return to education at all levels of education compared to DBN and Jhb respectively.

Furthermore, by looking at the estimated coefficients for the wave dummy variable, it is clear that there are significant differences in earnings with time for each metropolitan area. To see this, the third specification shows that, there is variation in earnings with time, as for Durban; significant differences in earnings with time were reported between 2004 and 2005 at all levels of significance. However, there were no significant differences between 2004 and 2006; and; 2004 and 2007. However, the reverse is true for CT and Jhb, as significant differences in earnings with time were reported in both metros except for the years 2004 and 2005.

The impact that labour market factors have upon hourly earnings varies according to each metropolitan area and this variation is broken down into five parts. Firstly, hourly earnings differ across industries given that each labour market industry is characterised by factors that distinguishes it from other labour market industries. The coefficient on the transportation industry suggests there are no significant differences in hourly earnings between the transport industry and manufacturing industry and the same can be said with the private household industry in reference to the manufacturing industry across the three metropolitan areas. However, coefficients on industries such as construction, finance and wholesale show that there is variation in hourly earnings by industry across the three metros.

Secondly, estimated coefficients on the different occupations examined in this study appear to be negative across three metropolitan areas. The direction of this effect suggests that all respective occupations earn lower hourly earnings relative to the managerial or legislative occupations and this result is true for all three metropolitan areas. An exception is made for the DBN metro, as the coefficient on professional occupation is insignificant. This suggests there are no significant differences in hourly earnings between workers who are professionals and those occupying managerial positions.

Thirdly, the coefficient on union membership across the three metros appears to be significant thus emphasising the positive effect that unions have upon hourly earnings. This implies that being a union member in any of the three respective metropolitan areas positively increases hourly earnings relative to non-union members and the size of the difference differs across the three metros.

Fourthly, the coefficient on formal across the three metropolitan areas, the direction of the effect implies that workers working in the formal sector have higher earnings relative to domestic workers and this result is significant at all levels of significance. By comparison, the coefficient on formal employment is higher for Jhb than it is for DBN and CT, which implies the gap between formal workers and domestic workers is higher in Jhb than it is in DBN and CT respectively. In contrast, we can see that the coefficient on informal is positive and significant at all levels of significance. The magnitude of the effect is higher in Jhb than it is CT and DBN respectively. The DBN metro has the lowest return which implies that working in an informal sector in DBN on average increases hourly earnings by 54 percent compared to domestic workers at all levels of significance ceteris paribus.

Lastly, the coefficient on fulltime is negative and significant across the three metropolitan areas indicating that workers working fulltime in the SA labour market earn lower hourly earnings compared to workers working part-time. One potential reason for this is that in SA, part-time workers are paid higher hourly rates than fulltime workers because this has been enforced by the Basic Conditions of Employment Act (BCEA) which provides a minimum standard of rights and protection for all workers working in South Africa (Department of Labour, 1997). The BCEA stipulates that in some sectors of employment, minimum hourly wages should be high for workers working lower hours. According to Posel and Muller (2008), workers who work less than 28 hours a week in the domestic service sector earn 10 percent higher hourly earnings relative to workers working longer hours. The gap widens in the wholesale and retail sector, as workers working less than 28 hours a week earn 25 percent higher hourly earnings relative to workers working longer hours in their occupation. These higher hourly earnings are rewarded to part-time workers in an attempt to offset the low benefits that come with occupying a part-time job (Posel and Muller, 2008).

4.6 Oaxaca Blinder Decomposition of Results

The previous section showed the existence of variation in hourly earnings across the three metropolitan areas. This variation in earnings is influenced by factors relating to personal, educational and labour market characteristics. In this section, the Oaxaca-Blinder (O-B) Decomposition method is applied in an attempt to single out factors causing earnings to differ across the three respective metropolitan areas of SA. As explained in the methodology chapter, the O-B decomposition decomposes the mean hourly wage difference into a part that is explained by observable characteristics and into a part that is unexplained (which is commonly considered as the residual or discriminating component). Doing so enables the study to see whether differences across metros are a result of characteristics or are due to a discriminating component. Given that the study examines differences in hourly earnings for the three metropolitan areas, therefore this section is divided into three parts: the first part decomposes mean hourly earnings for the CT-Jhb metro, the second part is for Jhb-DBN and the last part is for CT-DBN respectively.

Table 9 below presents the O-B decomposition results for CT-Jhb, Jhb-DBN and lastly CT-DBN. The dependent variable is the logarithm of hourly earnings and the results thus far have been expressed in logarithm scale. When analysing the decomposed mean difference across the three metros, the study retransforms the results into the original scale such that decompositions are presented in percentage terms as shown by Table 9 below:

Table 9: Oaxaca Blinder Decomposition Results

	Oaxaca Blinder Decomposition for the Metros	
	Amount Attributable	**P-Value**
Cape Town - Johannesburg		
Difference	13.8	0.000
Explained	22.2	0.000
Unexplained	93.1	0.018
Johannesburg - Durban		
Difference	26.9	0.000
Explained	96.9	0.257
Unexplained	30.9	0.000
Cape Town - Durban		
Difference	44.4	0.000
Explained	13.4	0.000
Unexplained	27.4	0.000

Firstly, the mean hourly earnings difference between the CT and the Jhb metro is 13.8 percent and this result is significant at all levels of significance. Looking at the effect that observable characteristics have upon the difference, it is clear that if workers living in Jhb had the same observable characteristics as workers living in CT then their hourly earnings would increase by 22.2 percent which is the explained component of the difference. However, 93.1 percent of the gap or difference in hourly earnings between the two metros remains unexplained and this result is significant at significant levels above 1.8 percent. The explained and unexplained components cancel each other out which is why on average one is not seeing a significance difference at a 1 percent significance level.

Moving on to the mean hourly earnings difference between Jhb and DBN, results presented in Table 9 suggest the mean hourly earnings difference between the two metros to be 26.9

percent. The coefficient on the explained component is insignificant, thus implying that there would be no difference in hourly earnings if workers living in DBN had the same observable characteristics as workers living in the Jhb metropolitan area. Looking at the discriminating component, 30.9 percent of the gap in hourly earnings between the two metros remains unexplained and this result is significant at all levels of significance. This is a clear indication that Jhb earnings are higher compared to DBN because of discrimination.

Lastly, the difference in hourly earnings between the CT and DBN metropolitan area is 44.4 percent as shown by Table 10 above. Given that the coefficient on the explained difference is significant, and if workers living in the DBN metro had the same observable characteristics as workers in the CT metro, then hourly earnings for DBN workers would increase by 13.4 percent. Furthermore, 27.4 percent of the difference in hourly earnings between the two metros remains unexplained and attributed to discrimination. By comparison, the gap in hourly earnings between these two metros appears to be large compared to other two gaps that were examined above.

There are three economic implications that can be drawn from the O-B decomposition of wage differentials across the three metropolitan areas presented in this study. The first implication is that of rigidities across the three metropolitan areas. Results presented by the O-B decomposition suggest the existence of wage differentials across the three metropolitan areas. As stated out by Azzoni and Servo (2002), the variation in hourly earnings is a sign that rigidities are strong within the labour market across the three metropolitan areas. This implies that wages do no equalise across the three metropolitan areas thus income inequalities will remain prominent across the country as these metros are the largest within the country.

The second economic implication of the three is that of the human capital and other observable characteristics. As argued by Motellon *et al,* (2011), differences in the returns on observable characteristics across metropolitan areas is a clear indication that some metros provide a higher return on investments in human capital and other observable characteristics

compared to other metros and this can be a policy implication for government. O-B Decomposition results for CT-Jhb and for CT-DBN suggests that the differences in hourly earnings is a result of observable characteristics. This is to say, the return on human capital and other observable characteristics is higher in CT than it is in Jhb and DBN and this result is significant at all levels of significance. Therefore, identical investments in human capital and other observable characteristics yield varying returns and the return is higher in CT than they are in Jhb and DBN. The third economic implication that can be drawn from these results is that of variation across the three labour markets. As stated by Mallik *et al,* (2014), variation across labour market is an indication that policy decisions by government are more likely to be relevant when targeted wherever labour market problems emerge. This implication can serve as a redistributive element implemented by government, particularly when labour markets are more diverse internally. The variation in the third specifications and in the O-B decomposition is a clear indication that labour markets across the three respective metros are diverse internally as a result of variation in hourly earnings and in the returns on observable characteristics. As suggested by empirical evidence, government policies will be effective when implemented correctly in dealing with labour market challenges.

Overall, it is evident that imbalances within the labour market persist across the three metropolitan areas. The first imbalance is that returns on observable characteristics are higher in CT compared to Jhb and DBN. This implies that if workers had the same observable characteristics, CT workers would earn higher than Jhb and DBN workers. This effect is attributed to both the returns on observable characteristics and discrimination. Labour market discrimination has proven to be persistent, particularly between Jhb and the DBN metropolitan areas. For the Jhb and DBN metros, the difference in average hourly earnings is a result of labour market discrimination. This implies that the Jhb metro discriminates higher than the DBN metro, as a result of these differences between the two metropolitan areas. The same can be said between CT and DBN, as differences in earnings are a result of both returns on observable characteristics and discrimination. Looking at the unexplained part, the difference is partly due to the fact that CT discriminates higher than the DBN metro. The reliability of the results is questionable given possible estimation challenges that are likely to affect models used to present these results. The reliability of the results is based upon the

assumptions made earlier in the possible estimation challenges such that if relation of the assumptions questions the reliability of these results.

4.7 Conclusion

The study used a multivariate analysis in examining and analysing differences in hourly earnings across the three metropolitan areas. In doing so, the study estimated the conventional wage equation where the dependent variable is given by the logarithm of hourly earnings. At first, the study estimates two models: the first model (specification 1) estimates the logarithm of hourly earnings upon the metro dummy variables such that there is no consideration for persona, education and labour market factors. The second model (specification 2) estimates the logarithm of hourly earnings upon metro dummy variables and the vector of personal, educational and labour market factors. Conducting both estimations was done upon the basis of verifying how wages vary across the three metros before and after the inclusion of control variables. If the included controls (specification 2) reduce the coefficient on the metro dummies then part of the wage difference detected in the first model (specification 1) is due to different metro distributions of the control variables. Results in the first specification suggested there are differences in hourly earnings across the three metropolitan areas. However, after controlling for observable characteristics in the second specification, results suggest there are no significant differences in hourly earnings between CT and Jhb but differences in earnings between DBN and Jhb exist. The inclusion of observable characteristics reduced the coefficient on DBN thus implying differences in earnings are due to the distribution of control variables. Both these models (particularly the second model) were conducted upon the assumption that returns on observable characteristics are identical across the three metros. However, relaxing this assumption leads to the third estimation where the logarithm of hourly earnings was estimated with respect to the logarithm of hourly earnings. The results showed there's variation in observable characteristics across the three metropolitan areas thus laying foundation for the O-B Decomposition.

The study decomposed differences in hourly earnings into a portion that is explained by observable characteristics and into a portion that is unobserved which is commonly considered as a discriminating component. Running the O-B Decomposition enabled the study

to capture economic implications as a result of differences in earnings across the three metropolitan areas. As presented in Table 9, the results suggest that differences in hourly earnings between CT and Jhb is a result of observable characteristics as the return to observable characteristics is higher in CT than it is in Jhb. The second decomposition between the Jhb and DBN metro suggests that the mean difference in hourly earnings is due to a discriminating component that is unobserved (E.g. ability or motivation). Lastly, decomposition between CT and DBN suggests that the mean difference in hourly earnings is due to both observable characteristics and discrimination. Overall, three economic implications can be drawn from the results presented by the O-B Decomposition. The first is that, variation within the labour market is an indication that labour markets are rigid, thus emphasising the prevailing differences in hourly earnings. The second economic implication is that the returns on observable characteristics differ according to each metropolitan area as some metropolitan areas reward highly compared to other metropolitan areas. And the last implication is that variation and diversity in labour markets makes it easier for government to implement policies. These implications are made on the basis that there is variation and on the basis that the return to observable characteristics is higher in some metropolitan areas compared to others.

CHAPTER FIVE: CONCLUSIONS AND RECOMMENDATIONS

5.1 Summary

Within a perfectly competitive labour market, firms and workers are said to be homogenous in that workers are identical in skills and firms are identical in the products produced within the market. The neoclassical paradigm indicates that wages amongst workers, across labour markets, regions and metropolitan areas will equalise. Multiple theories and methods have been developed by economists that deviate from the view that wages are equal amongst worker within the labour market and instead suggest that earnings differ amongst workers across labour markets, regions and metropolitan areas. Amongst these wage differential theories are: human capital, wage discrimination, compensating wage differentials, Efficiency Wage Theory, and institutional rigidities in the form of trade unions. For example, human capital affects earnings and workers vary in their stock of human capital thus varying investments in human capital by workers. This leads to varying returns which is then associated with varying distributions in earnings. Further to this, polarised development and segmented labour markets (primary and secondary labour markets) suggest that earnings differ at an aggregate level across regions and metropolitan areas.

There are also multiple factors that may lead to differences in earnings across metropolitan areas. Some metropolitan areas have the capacity to attract workers with high value human capital relative to other metropolitan areas, therefore resulting in varying returns on investments in human capital that affect the distribution of earnings. The size of a metropolitan area can also affect the distribution of earnings with labour markets such as those of large metropolitan areas being associated with higher earnings relative to small metropolitan areas which are associated with lower earnings. Apart from this, some metropolitan areas may have industrial structures that pay higher earnings relative to other metropolitan areas and this may create a diverge in earnings across metropolitan areas.

Given the above factors, this study aimed to examine differences in earnings across Johannesburg, Durban and Cape Town in an attempt to show whether differences in earnings exist and to find plausible explanations for these differences. Examining and analysing differences in earnings amongst workers across labour markets, regions, and metropolitan

areas is essential as it enables the study to provide policy implications, tests wage differential theories and derive quantitative estimates of important economic relationships.

The study used both univariate and multivariate analysis in examining differences in hourly earnings across Jhb, DBN and CT. In doing so, the study used the LFS data which is a household-based sample survey collected by Statistics South Arica (StatsSA). Results presented by the univariate analysis suggest that differences in hourly earnings exist for the three respective metropolitan areas. However, the magnitude of the difference appears to be minimal, unclear, and inconclusive. As a result, the study used the multivariate analysis in which the study estimates the wage equation. The dependent variable is given by the logarithm of hourly earnings and regressed upon the vector of personal characteristics, educational factors, and labour market factors. The inclusion of these parameters was based upon the idea that they affect the logarithm of hourly earnings as suggested by empirical evidence (Azzoni and Servo, 2002; Motellon *et al.,* 2011 and Mallik *et al.,* 2014).

The study used four approaches to examine the differences in hourly earnings for the three respective metros. The first approach (specification 1) included estimating the logarithm of hourly earnings upon the metro dummy variables to examine the effect of the differences before the inclusion of other controls. This approach is taken as a point of reference. The second approach (specification 2) included estimating the logarithm of hourly earnings upon the metro dummies and observable characteristics. The purpose of both estimations was to show how metro earnings vary as controls are included. Overall, results suggest that before the inclusion of control variables, there are significant differences in hourly earnings across the three metros. However, after controlling for other observable characteristics, there are no significant differences in hourly earnings between Jhb and CT. This implies the difference in hourly earnings between the two metros is a result of varying distribution of control variables.

The third approach estimates the logarithm of hourly earnings for each metropolitan area separately. Results presented in this third approach suggest there's variation in the returns on observable characteristics and this lays foundation for the O-B Decomposition. As stated in the methodology chapter, the O-B Decomposition apportions differences in hourly earnings between metros into a portion that is explained by observable characteristics and into a portion that is unexplained which is considered as a discriminating component. Results

presented by the O-B Decomposition are in line with wage differential findings reached in the first and second approach. According to this reasoning, there were significant differences between CT and Jhb (at a 10 percent significance level) in the first specification. However, there were no differences in earnings after controlling for other characteristics and this change is an effect of the distribution in observable characteristics. The O-B Decomposition reaffirms this result between the two metros, as O-B Decomposition results suggest that the mean difference in hourly earnings is 13.8 percent and 22.2 percent of this difference and is explained by observable characteristics. The same can be said between CT and DBN, as the O-B decomposition results suggest the mean difference in hourly earnings is 44.4 percent and 13.4 percent of the difference is explained by observable characteristics, whilst 27.5 percent of the mean difference in hourly earnings is unexplained.

There are three economic implications that can be drawn from these results. The first economic implication is that of prevailing rigidities across the three metros. These prevailing rigidities imply that earnings do not equalise across the three respective metropolitan areas meaning that the inequalities will remain prominent across the country. Prevailing inequalities are problematic to a government that advocates the elimination of inequalities, particularly if inequalities prevail amongst workers with identical skills. This would be a clear indication of inefficiencies.

The second economic implication is that different returns on human capital in metros like CT comprise higher returns on human capital and other observable characteristics compared to Jhb and DBN. This economic implication implies returns on human capital and other control variables are higher in CT regardless of investment in human capital or other control variables made by workers in Jhb and DBN. Differences in earnings across the three metros will prevail if government does not establish policies that seek to rectify these imbalances and doing so is a matter of importance as it implies inefficiencies across metropolitan areas. These inefficiencies are based upon the view that homogenous workers in skills vary in the distribution of earnings across metropolitan areas as some metros have higher returns on observable characteristics compared to others. This inefficiency is a challenge to government that advocates for redistribution and equality amongst workers across labour markets, regions and metropolitan areas. Furthermore, these imbalances imply that wage inequalities amongst

workers with identical skills across metropolitan areas will remain in place if redistributive measures are not implemented by government.

The third economic implication is that the more diverse labour markets are internally, the easier it becomes for government to implement policies in the presence of labour market challenges (Mallik *et al.,* 2014). Results presented in the third specification suggest there is variation in the returns on observable characteristics even though there are marginal differences in some coefficients. This is a clear indication that the three metros are diversified internally, thus government can easily implement policies in the presence of labour market challenges.

Given the above conclusion, there is potential for the study to be broadened for future research purposes. In this sense, the study only observed variations in earnings for three metropolitan areas, whilst there are eight distinguished metropolitan areas in South Africa. Broadening this study will give an overall picture of the differences in earning across eight distinct metropolitan areas of South Africa. In addition to future research, it must be noted that examining wage differences across metropolitan areas can be associated with sample selection bias and possible endogeneity in the form of omitted variable bias. This study has remedied these threats by assuming unobserved heterogeneity influencing the probability of being selected into the sample is not correlated with the unmeasured factors influencing the outcome variable and that there is no omitted variable bias. However, relaxing these assumptions for future research will be useful as it allows for the estimation of both the Heckman Selection Model and the Two-Stage Least Squares, thus more consistent estimates of wage differences across Johannesburg, Durban and Cape Town compared to just using OLS.

References

Aaronson, D. and French, E. (2004), "The Effect of Part-Time Work on Wages: Evidence from the Social Security Rules." Journal of Labour Economics, Vol. 22 (April 2004): 329 52.

Angel, D. P. and Mitchell, J. (1991), "Intermetropolitan wage disparities and industrial Change". Economic Geography 67 (2): 124– 35.

Angrist, J., & Krueger, A. B. (2001). "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments", National Bureau of Economic Research (No. w8456).

Antonakis, J., Bendahan, S., Jacquart, P. and Lalive, R. (2014), "Causality and Endogeneity: Problems and Solutions", The Oxford Handbook of Leadership and Organisation: 93-117.

Antonczyk, D., DeLeire, T. and Fitzenberger, B. (2010), "Polarization and Rising Wage Inequality: Comparing the U.S and Germany", IZA Discussion Paper No. 4842.

Appleton, S., Hoddinott, J. and Krishnan, P. (1999). "The Gender Wage Gap in Three African Countries". Economic Development and Cultural Change, 47(2): 289-312.

AL-Samarrail, S. and Reilly, B. (2008), "Education, Employment and Earnings of Secondary School and University Leavers in Tanzania: Evidence from a Tracer Study." Journal of Development Studies, Vol. 44(2): 258-288.

Azzoni, C. R. and Servo, L. M. (2002). "Education, Cost of Living and Regional Wage Inequality in Brazil". Papers in Regional Science, 81(2002): 157-175.

Baum-Snow, N., & Pavan, R. (2012). "Understanding the city size wage gap", The Review of economic studies, 79(1): 88-127.

Becker, Gary S. (1957). "The Economics of Discrimination", 2nd ed. Chicago: Univ. Chicago Press.

Becker, G.S. (1962). "Investment in Human Capital: A Theoretical Analysis". The Journal of Political Economy, 70(5): 9-49.

Becker, G.S. (1964). "Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education", New York: Columbia University Press for the NBER.

Becker, G. S. (1975). "Investment in human capital: effects on earnings. In Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education", Second Edition (pp. 13-44). NBER.

Becker, G.S. and Chiswick, B.R. (1966), “The Economics of Education and the Distribution of Earnings”, American Economic Review, 56 (2): 358-369.

Becker, G.S. and Murphy, K.M. (1992). “The Division of Labour Coordination Costs and Knowledge.” Quarterly Journal of Economics 107: 1137–60.

Beeson, P. E. and Randall W. E. (1989), "Identifying Productivity and Amenity Effects in Interurban Wage Differentials." Review of Economics and Statistics: 443-52.

Beeson, P.E. (1991), "Amenities and Regional Differences in Returns on Worker Characteristics." Journal of Urban Economics: 41-224.

Beeson, P. E and Groshen E. (1991), "Components of City-Size Wage Differentials, 1973 1988." Economic Review, Federal Reserve Bank of Cleveland, Vol.4: 10-24.

Bell, B.D, and Pitt, M.K, (1998), “Trade Unions Decline and the Distribution of Wages in the UK: Evidence from Kernel Density Estimation”, Oxford Bulletin of Economics and Statistics, Vol.60(4): 509-528

Bellante, D. (1979), "The North-South Differential and the Migration of Heterogeneous Labor." American Economic Review: 166-75.

Bettencourt, L. M., Lobo,.J., Helbing, D., Kühnert, C., and West, G. B. (2007). “Growth, Innovation, Scaling, and the Pace of Life in Cities.” Proceedings of the National Academy of Sciences, Vol. 104(17): 7301-7306.

Bhorat, H. (2000), “Wage Premia and Wage Differentials in the South African Labour Market”, DPRU Working Paper, No. 00/43.

Black, B. and Henderson, V. (1999), “A Theory of Urban Growth”, Journal of Political Economy, Vol. 107(2): 252-284.

Blanchflower, D. and Bryson, A. (2002), “Changes over time in union relative wage effects in the UK and the US Revisited”, NBER Working Paper No. 9395

Blinder, A.S. (1973), “Wage Discrimination: Reduced Form and Structural Estimates”, Journal of Human Resources, Vol. 8 (4): 436–455.

Blumenthal, P., Wolman, H.L. and Hill, E. (2009), “Understanding the Economic Performance of Metropolitan Areas in the United States.” Urban Studies, Vol. 46(3): 605-627.

Booth, A. L. (1995), “The economics of the trade union”. Cambridge University Press.

Booth, A. (2014), “Wage Determination and Imperfect Competition”. Institute for the Study of Labour, Discussion paper no.8034.

Bound, J, Jaeger, D.A. and Baker, R.M. (1995), “Problem with Instrumental Variable Estimation When the Correlation between the Instruments and the Endogenous Explanatory Variable is Weak.” Journal of American Statistical Association, Vol. 90: 430.

Bowels, S. Gintis, H. and Osborne, M. (2001), “The Determinants of Earnings: A Behavioural Approach.” Journal of Economic Literature, Vol. 39(4): 1137-1176.

Brock, J.F. (1949), “The Cape Coloured People: their Pattern of Health and Disease.” South African Medical Journal, Vol. 23(50): 1000-1010.

Burger, R.P. and Yu, D., 2006. “Wage Trends in Post-Apartheid South Africa: Constructing an Earnings Series for South Africa from Household Survey Data”. Labour Market Frontiers, Vol.8: 1-8.

Button, K.S, Loannidis, J.P.A, Morkrysz, C., Nosek, B.A, Flint, J., Robinson, E.S.J, and Munafo, M. R, (2013), “Power Failure: Why Small Sample Size Undermines the Reliability of Neuroscience”. Nature Reviews Neuroscience, 14(5): 365-376.

Card, D. (1999), “The Casual Effect of Education on Earnings.” Handbook of Labour Economics, Vol. 3: 1801-1863

Card, D. (2001), “The Effect of Unions on Wage Inequality in the U.S. Labour Market.” Industrial and Labour Review, Vol. 54, No. 2: 296-315

Card D., Lemieux T., and Riddell, W.C. (2004), “Unions and Wage Inequality”, Journal of Labour Research, Vol.25(4): 519-559

Carlino, G.A., (1986), “Do Regional Wages Differ?”, Business Review: 17-25

Chiswick, B.R (1974), “Income Inequality: Regional Analysis within a Human Capital Framework.” National Bureau of Economic Research, Vol. 74(1): 0-16

Chiswick, B.R. and Mincer J. (1972), “Time Series Changes in Personal Income Inequality in the United States from 1939, with projections to 1985.” Journal of Political Economy, VOLUME: 34-66.

Coelho, P. R., & Ghali, M. A. (1971),” The end of the North-South wage differential”, The American Economic Review, 932-937.

Darity, W.A. and Mason, P.L. (1998), “Evidence on Discrimination in Employment: Codes of Color, Codes of Gender.” The Journal of Economic Perspectives, Vol. 12(2): 63-90.

David, .H. (2001), “Wiring the Labour Market.” The Journal of Economic Perspectives, Vol. 15(1): 25-40.

Department of Labour (1997), “Basic Conditions of Employment Act”, Government Gazette, 390: 18491,

Dickie, M., and Gerking, S., (1987), “Interregional Wage Differentials: An Equilibrium Perspective”, Journal of Regional Science, Vol.27(4): 571-585

Dielman, T.E. (1983) “Pooled Cross-Sectional and Time Series Data: A Survey of Current Statistical Methodology.” American Statistical Association, Vol. 37, No.2: 111-122.

Dumond, J.M., Hirsch, B.T. and Macpherson, D.A. (1999), “Wage Differentials across Labour Markets and Workers: Does Cost of Living Matter?” Economic Inquiry, Vol 37 (4): 577–598.

Dunlop, J. T. (1957). “The Task of Contemporary Wage Theory”. In The Theory of Wage Determination Vol.11(3): 3-27

Eberts, R. W. and Schweitzer, M.E. (1994), "Regional Wage Convergence and Divergence." Economic Review, Federal Reserve Bank of Cleveland, Vol.2: 26-37.

Eissa, N and Hoynes, H.W. (2004), “Taxes and the Labour Market Participation of Married Couples: The Earned Income Tax Credit.” Journal of Public Economics, Vol.88: 1931 1958.

Elder, T. E., Goddeeris, J. H., and Haider, S. J. (2010), “Unexplained gaps and Oaxaca–Blinder decompositions”, Labour Economics, 17(1), 284-290.

Farber, S.C., and Newman, R.J., (1987), “Accounting for South/Non-South Real Wage Differentials and for Changes in Those Differentials Over Time”, The Review of Economics and Statistics, Vol.69(2): 215-223

Felsenstein, D. (2010), “Marshallian Theory of Regional Agglomeration.” Papers in Regional Science, Vol. 89 (1): 155-172.

Fortin, N., Lemieux, T. and Firpo, S. (2011), “Decomposition Methods in Economics.” Handbook of Labour Economics, 4: 1-102.

Freeman, R. B. (1980). “Unionism and the Dispersion of Wages”. Industrial & Labour Relations Review, 34(1), 3-23.

Freeman, R. and J. Medoff (1984), “What Do Unions Do?” Industrial and Labour Relations Review, Vol.38(2): 244

Fuchs, V. R. (1959),” Differentials in Hourly Earnings by Region and City Size.” National Bureau of Economic Research Occasional Paper No. 101.

Fuchs, V.R. (1967), “Differentials in Hourly Earnings by Region and City Size, 1959”, National Bureau of Economic Research, Vol. 7(1): 1-5

Gabriel, P. E. and Schmitz, S. (2006), “The Impact of Gender Differences in Occupational Attainment on the Relative Earnings of Young Workers.” Applied Economics Letters, Vol. 13(10): 615-619

Gallaway, L., (1963), “The North-South Wage Differential”, Review of Economics and Statistics, Vol.45: 264-272

Garen, J. (1988). “Compensating wage differentials and the endogeneity of job riskiness”. The Review of Economics and Statistics, Vol.70(1) 9-16.

Gerking, S. D., & Weirick, W. N. (1983), “Compensating differences and interregional wage differentials”, The Review of Economics and Statistics, 483-487.

George, J. (1983). “Inter-metropolitan Wage Differentials in the United States”. The Measure of Labour Cost: 309-332.

Girijasankar M, Parikshit K.B, John H, and Richard S, (2014), “Do the Determinants of employability and earnings returns produce similar outcomes in metropolitan and regional labour markets? The case of New South Wales, AUDTRALIA”, Regional Studies, Vol. 48:10, pp. 1706-1718

Glaeser, E. L., Kallal, H. D., Scheinkman, J. A., and Shleifer, A. (1992)."Growth in cities", Journal of Political Economy, 100(6): 1126-1152.

Glaeser, E.L. and Maré D.C, (1994), "Cities and Skills", National Bureau of Economic Research Working Paper No. 4728

Glaeser E.L, Mare D.C, (2001), “Cities and Skills”, Journal of Labour Economics, Vol. 19(2), pp.316-342

Glaeser E.L, and Kahn M.E, (2004), “Sprawl and Urban Growth”, Handbook of Regional and Urban Economics, Vol.4: 241-2527

Glaeser E.L, and Resseger M.G, (2009), “The complementarity between Cities and Skills” The Journal of Regional Science, vol. 50(1): 221-244

Glaeser, E. L., Resseger, M., and Tobio, K. (2009),” Inequality in cities”, Journal of Regional Science, 49(4), 617-646.

Glaeser, E.L, and Gottlieb, J.D, (2009), “The Wealth of Cities: Agglomeration Economies and Spatial Equilibrium in the United States”, Journal of Economic Literature, Vol.47(4): 983 – 1028

Goldfarb, R. S., and Yezer, M. J. (1976). "Evaluating alternative theories of intercity and inter regional wage differentials", Journal of Regional Science, 16(3), 345-363.

Green, T. K. (2003), “Discrimination in workplace dynamics: Toward a structural account of disparate treatment theory”, Harv. CR-CLL Rev., 38, 91.

Hamermesh, D. S., and Wolfe, J. R., (1990), “Compensating Wage Differentials and the Duration of Wage Loss. Journal of Labor Economics, Vol.8(1): 175-197

Hanushek, E., (1973), "Regional Differences in the Structure of Earnings", Review of Economics and Statistics: 204-13.

Haroon, B., Leibbrandt, M., Maziya, M., Van der Berg, S., and Woolard, I., (2001). “Fighting Poverty: Labour markets and inequality in South Africa”. UCT Press

Heckman, J.J, Loncher L.J, and Todd P.E, (2003), “Fifty Years of Mincer Earnings Regressions”, National Bureau of Economic Research, Vol. 1

Heckman, J. (2013), “Sample selection bias as a specification error”, Applied Econometrics, Vol.31(3):129-137.

Hirsch B.T, (2004), “Why do Part-Time Workers Earn Less? The role of Worker and job Skills”, IZA Discussion Paper No. 1261

Hirsch B.T, (1978), “Earnings Inequality across Labour Markets: A Test of the Human Capital Model”, Southern Economic Journal, Vol.54(1): 32-45

Hirsch, B.T, and Schumacher, E.J, (1998), “Unions, Wages, and Skills”, The Journal of Human Resources, Vol.33(1): 201-219

Hirsch, B. T. (2008), “Sluggish institutions in a dynamic world: Can unions and industrial competition coexist?”, The Journal of Economic Perspectives, 22(1), 153-176.

Hlavac, M. (2014), “Oaxaca: Blinder-Oaxaca Decomposition in R”. Available at SSRN 2528391.

Holzer, H.J., and Ihlanfeldt, K.R., (1998), “Customer Discrimination and Employment Outcomes for Minority Workers”, Quarterly Journal of Economics: 835-867

Hwang, H.S., W. R. Reed and C. Hubbard (1992), “Compensating Wage Differentials and Unobserved Productivity", Journal of Political Economy, Vol.100(4):835 - 858.

Industrial Development Corporation, (2013), “South African Economy: An Overview of Key Trends Since 1994”, Department of Research and Information.

Jacobs, J., (1969), “The Economy of Cities”, New York: Vintage.

Jann, B., (2008), “The Blinder-Oaxaca Decomposition for linear regression models”, The Stata Journal, Vol. 8(4), 453-479.

John W. Budd, (2005), "The Effect of Unions on Employee Benefits: Updated Employer Expenditure Results", Journal of Labour Research, Transaction Publishers, vol. 26(4): 669-676

Johnson, G.E., (1983), "Inter-metropolitan Wage Differentials in the United States." In The Measurement of Labour Cost, Chicago and London: University of Chicago Press, 309 30.

Jones, F. L. (1983), “On decomposing the wage gap: a critical comment on Blinder’s method”, Journal of Human Resources, Vol. 18: 30-126.

Kahn, L. M. (1980), “Union spill-over effects on organized labour markets”. The Journal of Human Resources, Vol.15(1): 87-98.

Katz, L. F. (1986), “Efficiency wage theories: A partial evaluation. In NBER Macroeconomics Annual 1986, Vol.1: 235-290.

Kerr. A, and Francis. T, (2015), “The Determinants of Earnings Inequalities: Panel Data Evidence from KwaZulu-Natal, South Africa “Journal of African Economics Vol. 24 (4): pp 530 558

Kiefer, N. M., and Smith, S. P. (1977), “Union impact and wage discrimination by region”. Journal of Human Resources, 521-534.

Kim, D., Liu, F., and Yezer, A. (2009). "Do inter-city differences in intra-city wage differentials have any interesting implications". Journal of Urban Economics, 66(2009), 203–209.

Kim, J., Scott, C., (2012), “Robust Kernel Density Estimation”, Journal of Machine Learning Research: 2529-2565

Krumm, R., (1984), “Regional Wage Differentials, Labour Supply Responses and Race”, Journal of Regional Science, Vol.24: 431-442

Kuper, H., (1960), “Indian People in Natal, Pietermaritzburg: Natal University Press.

Lazear, E.P, (1996), “Performance Pay and Productivity”, The American Economic Review, Vol.90(5): 1346-1361

Leibbrandt, M., Woolard, I., McEwen, H., and Koep, C., (2010) “Employment and inequality outcomes in South Africa: What role for labour market and social policies?” Report to the OECD

Leibbrandt, M., I. Woolard, A. Finn, and J. Argent (2010), “Trends in South African Income Distribution and Poverty since the Fall of Apartheid”, OECD Social, Employment and Migration Working Papers No. 101

Lentz R, and Mortensen D.T, (2010),” Labour Market Models of Worker and Firm Heterogeneity”, Annual Review of Economics, Vol. 2(1): 577-602

Leontaridi, M., (1998), “Segmented Labour Markets: Theory and Evidance”, Journal of Economic Surveys, Vol.12(1): 103-109

Lewis, H.G, (1986), “Union Relative Wage Effects: A Survey of Macro Estimates”, Journal of Labour Economics, 1-27

Lucas, R.E., (1988), ‘‘On the Mechanics of Economic Development’’, Journal of Monetary Economics, Vol. 22: 3–42.

Malcomson, J. M. (1981), “Unemployment and the efficiency wage hypothesis”. The Economic Journal, 91(364): 848-866.

Mallik, G., Basu, P.k., Hicks, J., and Sappey, R., (2014), “Do the Determinants of Employability and Earnings Returns Produce Similar Outcomes in Metropolitan and Regional Labour Markets? The Case of New South Wales, Australia”, Regional Studies, Vol.48(10): 1706-1718.

Manning, A. (2010), “Imperfect Competition in the Labour Market”. Centre for Economic Performance, Discussion paper No.981

Marshall, A., (1890), “Principles of Economics” an Introductory Volume, 8th edition. London: Macmillan

Mincer, J., (1958), “Investment in Human Capital and Personal Income Distribution”, Journal of Political Economy, Vol.66(4): 281-302

Mincer, J., (1974), “Schooling, Experience, and Earnings”, National Bureau of Economic Research, Distributed by Columbia University Press, New York.

Mincer, J. (1974),” Progress in Human Capital Analysis of the distribution of earnings”.

Montgomery, E., (1992), "Evidence on Metropolitan Wage Differences across Industries and Over Time." Journal of Urban Economics: 69-83.

Motellon, E., Lopex-Bazo, E., and El-Attar, M. (2011),"Regional heterogeneity in wage distributions: Evidence from Spain". Journal of Regional Science, 51(3), 558-584.

Oaxaca, R., (1973),” Male–female wage differentials in urban labour markets”, International Economic Review, Vol.14 (3): 693–709.

Pager, D., Western, B., and Bonikowski, B., (2009), “Discrimination in a low-wage labor market a field experiment”, American Sociological Review, 74(5), 777-799.

Polachek S.W, (2007), “Earnings Over the Lifecycle: The Mincer Earnings Function and Its Applications”, State University of New York at Binghamton and IZA, Discussion Paper No.3181.

Porter, M.E., (1990), “The Competitive Advantage of Nations”, New York: Free Press.

Posel, D, and Muller C, (2008), “Is there Evidence of Wage Penalty to Female Part-Time Employment in South Africa?”, South African Journal of Economics, Vol. 76(3): 466 479

Puhan, P, (2000), “The Heckman Correction for Sample Selection and its Critique”, Journal of Economic Surveys, Vol. 14(1): 53-68

Rauch, James E., “Productivity Gains from Geographic Concentration of Human Capital: Evidence from the Cities,” Journal of Urban Economics 34 (1993), 380–400.

Reis, J.A, Barros R.P (1991), “Wage inequality and the distribution of education: A study of the evolution of regional differences in inequality in metropolitan Brazil”, Journal of Development Economics 36(1): pp.117–143

Roback, J., (1988), "Wages, Rents, and Amenities: Differences Among Workers and Regions." Economic Inquiry: 23-41.

Rogerson, C., (2009), “Johannesburg: Pro-Poor Growth through Support of the SMME Economy”

Romer, P.E., (1986), “Increasing Returns and Long-Run Growth”, Journal of Political Economy, Vol. 94: 37-1002

Samuel, B., Gintis, H., and Osborne, M., (2001), “The determinants of Earnings: A Behavioural Approach”, Journal of Economic Literature, Vol.39(4): 1137-1176

Sandrine R, (2002), “How did Labour Market Racial Discrimination Evolve after the End of Apartheid”, South African Journal of Economics, vol. 70(1) pp185-217.

Sahling, L., and Smith, S., (1983), “Regional Wage Differentials: Has the South Risen Again?”, Review of Economics and Statistics, Vol.65: 131-135

Schultz, T, W. (1974), “Marriage, Family, Human Capital, and Fertility”. Journal of Political Economy, Vol 82(2), 76-110

Seekings, J. (2011), “Race, class, and inequality in the South African city”, The New Blackwell Companion to the City, 532-546.

Shapiro, C. and Stiglitz, J. E. (1984), “Equilibrium unemployment as a worker discipline device”, The American Economic Review, 74(3), 433-444.

Shapiro, J. (2006), “Smart Cities – Quality of Life, Productivity, and the Growth Effects of Human Capital, forthcoming: Review of Economics and Statistics”

Smith, R. S. (1979). “Compensating wage differentials and public policy: a review”. Industrial and Labor Relations Review, 339-352.

Snow N, and Pavan R, (2012), “Understanding the City Size Wage Gap”, Review of Economic Studies, Vol. 79(1): 88-127

Statistics South Africa Census, (2011), Statistical Release (Revised): P0301.4

Stigler,G.J. (1962). “Information in the Labour Market”. Journal of Political Economy, 70(5), 94-105

Stiglitz, J.E, (1976), “The Efficiency Wage Hypothesis, Surplus Labour, and the Distribution of income in L.D.C.s”, Oxford Economic Papers, New Series, Vol.2: 185-207

Stiglitz, J., (1976), “Prices and Queues as Screening Devices in Competitive Markets”, IMSSS Technical Report No.212, Stanford University.

Stapleton, C. (2015). The Migrant Network Effect: An empirical analysis of rural-to-urban migration in South Africa (No. 504).

Storper, M and Scott, J. (2009), “Rethinking Human Capital, Creativity and Urban Growth”. Journal of Economic Geography, pp147-167

Vella F., (1998), “Estimating Models with Sample Selection Bias: A Survey”, The Journal of HumanResources, Vol. 33(1): 127-169

Yellen, J. (1984).” Efficiency Wage Models of Unemployment.” American Economic Review, Vol.74(2): 280-289

Verma. V, Gagliardi. F, Ferretti .C, (2009), “On pooling of data and measures” Working paper no.84

Viscusi, W.K, (1978), “Wealth Effects and Earnings Premiums for Job Hazards”, Harvard University Press, Employment Hazards Review 60: 408-416

Weiss A. (1980), “Job Queues and Layoffs in Labour Markets with Flexible Wages”, Journal of Political Economy, Vol.88 (3): 526-538

Williams R., (2015), “Specification Error: Omitted and Extraneous Variables”, University of Notre Dame.

Willis R.J, (1986), “Wage determinants: A survey and reinterpretation of human capital earnings function”, Handbook of Labour Economics: 525-602

Appendices

Appendix 1: Oaxaca - Blinder Decomposition for the Jhb and CT metros

```
Blinder-Oaxaca Decomposition

Number of strata   =         1                  Number of obs      =     10365
Number of PSUs     =     10365                  Population size    =  13994319
                                                Design df          =     10364
                                                Model              =    linear
Group 1: ctjhb = 0                              N of obs 1         =      2264
Group 2: ctjhb = 1                              N of obs 2         =      3014

------------------------------------------------------------------------------
             |             Linearized
lwage |     exp(b)   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
overall      |
     group_1 |   16.81404   .4263534   111.30   0.000     15.99873    17.67089
     group_2 |    14.7771   .3926295   101.36   0.000     14.02717    15.56712
difference   |   1.137844   .0417908     3.52   0.000     1.058806    1.222783
explained    |   1.222582   .0403883     6.08   0.000     1.145921     1.30437
unexplained  |     .93069   .0282266    -2.37   0.018     .8769729    .9876973
-------------+----------------------------------------------------------------
explained    |
expten       |   1.222582   .0403883     6.08   0.000     1.145921     1.30437
-------------+----------------------------------------------------------------
unexplained  |
expten       |   1.577129   .6394609     1.12   0.261     .7123604    3.491682
      __cons |   .5901165   .2423737    -1.28   0.199     .2638092    1.320035
------------------------------------------------------------------------------
```

Appendix 2: Oaxaca - Blinder Decomposition for the Jhb and DBN metros

```
Blinder-Oaxaca Decomposition

Number of strata   =         1                  Number of obs      =     10365
Number of PSUs     =     10365                  Population size    =  13994319
                                                Design df          =     10364
                                                Model              =    linear
Group 1: jhbdbn = 0                             N of obs 1         =      3014
Group 2: jhbdbn = 1                             N of obs 2         =      2456

------------------------------------------------------------------------------
             |             Linearized
lwage        |     exp(b)   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
overall      |
     group_1 |    14.7771   .3926295   101.36   0.000     14.02717    15.56712
     group_2 |   11.64421   .2809021   101.76   0.000     11.10641    12.20806
difference   |   1.269051   .0455433     6.64   0.000     1.182845     1.36154
explained    |   .9692141   .0267355    -1.13   0.257      .918199    1.023064
unexplained  |   1.309361   .0325764    10.83   0.000     1.247037      1.3748
-------------+----------------------------------------------------------------
explained    |
expten       |   .9692141   .0267355    -1.13   0.257      .918199    1.023064
-------------+----------------------------------------------------------------
unexplained  |
expten       |   1.291138    .512773     0.64   0.520     .5927614    2.812323
      __cons |   1.014114   .4052191     0.04   0.972     .4633662    2.219469
------------------------------------------------------------------------------
```

Appendix 8: Oaxaca - Blinder Decomposition for the CT and DBN metros

```
Blinder-Oaxaca Decomposition

Number of strata   =         1                  Number of obs      =     10365
Number of PSUs     =     10365                  Population size    =  13994319
                                                Design df          =     10364
                                                Model              =    linear
Group 1: ctdbn = 0                              N of obs 1         =      2264
Group 2: ctdbn = 1                              N of obs 2         =      2456

------------------------------------------------------------------------------
             |             Linearized
lwage |     exp(b)   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
overall      |
     group_1 |   16.81404   .4263534   111.30   0.000     15.99873    17.67089
     group_2 |   11.64421   .2809021   101.76   0.000     11.10641    12.20806
difference   |   1.443982    .050538    10.50   0.000      1.34824    1.546524
explained    |   1.133696    .036354     3.91   0.000     1.064629    1.207245
unexplained  |   1.273694   .0402693     7.65   0.000     1.197155    1.355127
-------------+----------------------------------------------------------------
explained    |
expten       |   1.133696    .036354     3.91   0.000     1.064629    1.207245
-------------+----------------------------------------------------------------
unexplained  |
expten       |   2.128338   .7737009     2.08   0.038     1.043693    4.340188
      __cons |   .5984454   .2180425    -1.41   0.159     .2929912    1.222347
------------------------------------------------------------------------------
```

UNIVERSITY OF KWAZULU-NATAL
INYUVESI YAKWAZULU-NATALI

02 November 2015

Mr Sakhile Keith Mpungose (209516715)
School of Accounting, Economics & Finance
Westville Campus

Dear Mr Mpungose,

Protocol reference number: HSS/1617/015M
Project title: Differences in earnings across Cape Town, Johannesburg & Durban

Full Approval – No Risk / Exempt Application

In response to your application received on 28 October 2015, the Humanities & Social Sciences Research Ethics Committee has considered the abovementioned application and the protocol have been granted **FULL APPROVAL.**

Any alteration/s to the approved research protocol i.e. Questionnaire/Interview Schedule, Informed Consent Form, Title of the Project, Location of the Study, Research Approach and Methods must be reviewed and approved through the amendment/modification prior to its implementation. In case you have further queries, please quote the above reference number.

PLEASE NOTE: Research data should be securely stored in the discipline/department for a period of 5 years.

The ethical clearance certificate is only valid for a period of 3 years from the date of issue. Thereafter Recertification must be applied for on an annual basis.

I take this opportunity of wishing you everything of the best with your study.

Yours faithfully

.....
Dr Shenuka Singh (Chair)

/ms

Cc Supervisor: Dr Colette Muller
Cc Academic Leader Research: Dr Harold Ngalawa
Cc School Administrator: Ms Seshni Naidoo

Humanities & Social Sciences Research Ethics Committee
Dr Shenuka Singh (Chair)
Westville Campus, Govan Mbeki Building
Postal Address: Private Bag X54001, Durban 4000
Telephone: +27 (0) 31 260 3587/8350/4557 **Facsimile:** +27 (0) 31 260 4609 **Email:** ximbap@ukzn.ac.za / snymanm@ukzn.ac.za / mohunp@ukzn.ac.za
Website: www.ukzn.ac.za

UNIVERSITY OF KWAZULU-NATAL
INYUVESI YAKWAZULU-NATALI

02 November 2015

Mr Sakhile Keith Mpungose (209516715)
School of Accounting, Economics & Finance
Westville Campus

Dear Mr Mpungose,

Protocol reference number: HSS/1617/015M
Project title: Differences in earnings across Cape Town, Johannesburg & Durban

Full Approval – No Risk / Exempt Application

In response to your application received on 28 October 2015, the Humanities & Social Sciences Research Ethics Committee has considered the abovementioned application and the protocol have been granted **FULL APPROVAL.**

Any alteration/s to the approved research protocol i.e. Questionnaire/Interview Schedule, Informed Consent Form, Title of the Project, Location of the Study, Research Approach and Methods must be reviewed and approved through the amendment/modification prior to its implementation. In case you have further queries, please quote the above reference number.

PLEASE NOTE: Research data should be securely stored in the discipline/department for a period of 5 years.

The ethical clearance certificate is only valid for a period of 3 years from the date of issue. Thereafter Recertification must be applied for on an annual basis.

I take this opportunity of wishing you everything of the best with your study.

Yours faithfully

.....
Dr Shenuka Singh (Chair)

/ms

Cc Supervisor: Dr Colette Muller
Cc Academic Leader Research: Dr Harold Ngalawa
Cc School Administrator: Ms Seshni Naidoo

Humanities & Social Sciences Research Ethics Committee
Dr Shenuka Singh (Chair)
Westville Campus, Govan Mbeki Building
Postal Address: Private Bag X54001, Durban 4000
Telephone: +27 (0) 31 260 3587/8350/4557 **Facsimile:** +27 (0) 31 260 4609 **Email:** ximbap@ukzn.ac.za / snymanm@ukzn.ac.za / mohunp@ukzn.ac.za
Website: www.ukzn.ac.za

Proofreading and Correcting

TO WHOM IT MAY CONCERN:

This letter serves to confirm that I have proofread a copy of the following Dissertation/ Thesis/ Journal Article.

TITLE: **DIFFERENCES IN EARNINGS ACROSS CAPE TOWN JOHANNESBURG AND DURBAN**

RESEARCHER: **SAKHILE KEITH MPUNGOSE**

The general areas covered in this proofreading include:

- Spelling - with special reference to English UK spellings of specific words.
- Correction of grammatical errors: syntax, concord etc.
- General editing to improve the language and vocabulary used and to, where necessary, adjust to make the work more academic in tone and style.
- Comments on general layout in terms of consistency in style: bullet lists, Figure and Table headings, Chapter headings and sub-headings.
- Checking that in-text authors appear in the Reference List
- Comments on and corrections of the Reference List entries

PLEASE NOTE:
I have made suggestions to the researcher in terms of corrections which s/he may choose/choose not to put into effect in the final copy. I take no personal responsibility for the student's decisions in implementing or rejecting my suggestions and corrections in the Final Draft

Date: 02 MARCH 2016

LINDA ROBINSON (MRS) (BA HDE)

lindarobinson@hotmail.co.za
083 356 3591

turnitin Turnitin Originality Report

Final Draft by Sakhile Mpungose

From Final proposal - Part 1 (Moodle 33240767) (2015 ECON8MRW1 Research Methodology for Advanced Economics (Moodle 9616244))

- Processed on 10-Dec-2015 11:35 AM CAT
- ID: 546057771
- Word Count: 22417

Similarity Index
6%
Similarity by Source
Internet Sources:
4%
Publications:
3%
Student Papers:
1%