Economics Masters Dissertation
University of Cape Town

Measuring Inequality of Opportunity in South Africa

Evelyne. M.Nyokangi

Supervisors: Dr. Patrizio Piraino and Dr. Murray Leibbrandt

Abstract

This paper examines the effect of circumstances on the opportunities available to individuals in South Africa, by quantifying the degree to which inequalities in labour market outcomes are due to circumstances (unequal opportunities). To do so, two distinct Inequality of Opportunity indices are applied to the first wave of the National Income Dynamic Study (NIDS). The *dissimilarity index* estimates the opportunities that need to be reallocated, for all economically active South Africans to have equal access to employment in spite of their circumstances. Whereas the *inequality of economic opportunity index*, estimates the (lower bound) share of total income inequality that can be attributed to differing circumstances. Results from the empirical analyses reveal that circumstances, such as race, gender and parental education, do not contribute significantly to inequalities in accessing employment. This is in contrast to the substantial share of labour market income inequality, found to stem from circumstances. These results suggest that policies aimed at redressing inequities in the labour market, should focus on the channels through which circumstances, especially race and gender impact an individual's opportunities and thus their ability to acquire labour market income.

October 30, 2014

Contents

Introduction

The unequal distribution of outcomes associated with individual well-being across a population, is a topic that draws considerable interest and intense debate. The source of contention lies in whether inequality is "unfair" and on the role of the state in addressing these imbalances. This paper builds on recent literature in the comparatively new Inequality of Opportunities research, in which academics theorise that overall inequality can be partitioned into a share that is attributable to unequal opportunities produced by inherited circumstances and a share due to differentials in individual efforts (Bourguignon, Ferreira and Menéndez, 2007a; Checchi and Peragine, 2010a; Ferreira and Gignoux, 2011). These scholars also propose that only inequalities stemming from unequal opportunities are "unfair" because their existence means that an individual's well-being is constrained by factors out of their control and therefore that outcomes are not purely self-determined.

There has already been considerable research into the causes of South Africa's high rate of unemployment and high levels of income inequality (Leibbrandt, *et al.*, 2008). This paper differs from those studies because it offers an alternative perspective through which inequalities in accessing employment and acquiring labour market income (once the individual enters employment) can be studied. This perspective is important, especially for those concerned about society's ability to provide an even playing field for all its members, because it offers insight into the institutional and economic mechanisms generating inequality traps in South Africa (Ferreira and Walton, 2006).

Therefore, this research is motivated in part by social attitudes of the underlying causes of inequality, and of the role of the state in ensuring the "fair" distribution of outcomes associated with economic well-being. Alesina and Angeletos (2005) hypothesised that perceptions about the causes of inequality determine the extent to which people support redistribution. This was somewhat confirmed by Gaviria (2007), who found that support for policies aimed at redistribution is greater, if inequalities are perceived to arise from unequal opportunities as opposed to differentials in effort. Thus being able to measure the effect of unequal opportunities could prove to be useful in the political arena, where government must often justify public expenditures (Gaviria, 2007).

Studies into the drivers of economic growth have shown that a country's aggregate economic performance is negatively affected by inequalities stemming from unequal opportunities (Bourguignon, Ferreira and Walton, 2007b; Marrero and Rodríguez, 2010). This is because long-standing inequalities in the opportunities available to individuals generate inequality traps that place constraints on the productive capacity of an economy. One source of these constraints is the inefficiencies produced when individuals with "better" inherited circumstances accumulate human capital rather than the most industrious, or in other words the hardest workers or the most highly skilled (Marrero and Rodríguez, 2010). This would not be the case if inequalities were entirely determined by differentials in the effort exerted by the individual. In such a scenario "worse-off" individuals would be motivated to accumulate human capital in order to improve their well-being.

The aim of this paper is therefore to quantify the extent to which existing inequalities in the likelihood of employment and in labour market income are a product of unequal opportunities faced by individuals with different circumstances. To accomplish this objective, a set of Inequality of Opportunity estimates are provided for South Africa using a variety of variables as proxies for circumstances. The results from this analysis suggest that whereas circumstances produce substantial inequalities in the opportunities available for employed individuals to acquire labour market income, their impact on the ability of individuals to access employment is marginal.

Section two of this paper presents the Inequality of Opportunities framework proposed by Roemer (1998) and draws attention to the two distinct approaches through which unequal opportunities can be measured. This part of the paper also contains a concise review of papers that have made a significant contribution to the construction of the methodology implemented in this study. Section three draws from the reviewed empirical literature and provides a unified methodology for estimating the Inequality of Opportunities in accessing employment and acquiring labour market income econometrically. This section also highlights the strengths and weaknesses of the proposed estimation strategies. Section four of this paper then goes on to provide information on the data set on which this analysis was applied and includes information of the variables selected as proxies for circumstances and a discussion of the criteria used to select the sample on which this analysis was based. This section also contains an analysis into whether the sample selection process would result in biased Inequality of Opportunities estimates. Section five reports and interprets the results from the empirical analysis. Finally, the impact of circumstance proxies not included in the initial analysis on Inequality of Opportunity estimates, are presented and discussed in Section six. This expansion partly addresses questions regarding the extent to which circumstances unobserved in the empirical analysis contained in Section five generate unequal opportunities in South Africa. Section seven concludes and suggests areas in which there is scope for further research.

Section 2: Background

2.1 The Foundations of Inequality of Opportunities

The Inequality of Opportunities framework originates from an extensive debate in political philosophy, on the role of the egalitarian planner in ensuring that outcomes are distributed equitably in society (Dworkin, 1981; Arneson, 1989; Fleurbaey, 1995; Roemer, 1998). The general consensus from scholars involved in this debate, was that equal outcomes should not be the goal of an egalitarian planner seeking to maximise social welfare. This is because policies implemented to equalise the outcomes of all individuals, result in "unfair" redistributions, due to the fact that individuals who make decisions advantageous to their outcome end up with the same overall outcome as those who do not (Roemer, 1998). Therefore, by not holding individuals accountable for their choices, such policy interventions tend to **disincentivise,** and resultantly distorts the behaviour of individuals (Fleurbaey, 1995).

The issue of personal responsibility is at the forefront of the egalitarian planner's welfare maximisation problem. This is why scholars assume that the planner has perfect information on all factors affecting the outcome of individuals (Roemer, 1998). This assumption allows the planner to assign responsibility over a subset of factors to the individual, and the individual is held accountable for the impact of these factors on their outcome. Therefore in order to maximise social welfare and bring about equality of opportunity, the egalitarian planner allows inequalities resulting from "responsible characteristics" to persist, but neutralises inequalities resulting from "non-responsible characteristics" through compensation. Fleurbaey (1995) called the first condition for welfare maximisation "the principle of natural reward" and called the second condition "the principle of compensation". Roemer (1998) then made the seminal contribution of formalising these theories through his construction of the conceptual framework through which the literature on Inequality of Opportunities has evolved.

Roemer (2006) firstly proposed that an individual's outcome (whether it be earnings, income or any other indicator of socio-economic status) is entirely determined by their pre-determined circumstances and their efforts. He defined circumstances as traits *exogenous* but intrinsic to the individual ("non-responsible characteristics"). These included race, gender, parental education, parental occupation and region of birth. Efforts on the other hand, were defined as factors that the individual can affect through their choices. These efforts are *endogenous* to the individual ("responsible characteristics"), and include variables such as education level, occupation and region of residence[1]. Roemer (2006) then proposed that a given

[1] It is important to note that in South Africa, these "responsible characteristics" are partly determined by economic, political, social, and other institutions (Gradı´n, 2012).

population can be completely partitioned into *types* and *tranches*, where types are subgroups of individuals with identical circumstances, and tranches are subgroups of individuals exerting the same degree of effort.

In this Roemerian framework, equality of opportunity can be realised in two ways (Roemer and Trannoy, 2013). The first requires between-types inequalities in outcome to be eradicated, by compensating individuals for outcome differentials unambiguously due to the opportunities conferred by their circumstances. This is referred to as the ex-ante compensation principle. The second way is referred to as the ex-post compensation principle, and requires the eradication of within-tranche inequalities, ensuring that individuals exerting the same degree of effort have identical outcomes, irrespective of their circumstances.

Following Roemer's conceptual contribution, the majority of the Inequality of Opportunity literature has been empirical in nature (Bourguignon *et al.*, 2007a; Cogneau and Mesple-Somps, 2009; Pistolesi, 2009; Checchi and Peragine, 2010a; Ferreira and Gignoux, 2011; Singh, 2012; Belhaj-Hassine, 2012; Brunori, Ferreira and Peragine, 2013).

2.2 Review of Empirical Literature

There are a number of studies which have estimated Inequality of Opportunities in developing and developed countries. This section however, contains a brief review of studies that made a significant contribution to the methodology implemented in recent empirical Inequality of Opportunities research. Bourguignon et al. (2007a), for instance, used a parametric approach to estimate Inequality of Opportunities in earnings in Brazil for a sample of urban males. This approach required Bourguignon et al. (2007a) to make some functional form assumptions about earnings. The first was that earnings are linearly related to circumstances, efforts and other unobservable factors. The second was that efforts are a linear function of circumstances and other unobservable factors. These assumptions were necessary because they allowed labour market earnings to be expressed as a function of circumstances and unobservable factors $\ln(y_i) = \psi C_i + \varepsilon_i$.

The parameters estimates derived from an OLS regression of the model defined above were used to generate the distribution of counterfactual earnings $\tilde{\mu_i} = \exp[\bar{C}_i\hat{\psi} + \hat{\varepsilon}_i]$, under the counterfactual that all individuals in the sample have the same set of circumstances[2]. The equalisation of circumstances means that all variations in counterfactual earnings are entirely due to differentials in unobservable

[2] For each circumstance variable, Bourguignon *et al.* (2007a) replaced individual circumstance values with their sample mean. So that $C_i = \bar{C}$ for all individuals in the sample, i =1…N.

factors. Therefore Inequality of Opportunity is the estimated difference between inequality in the actual earnings distribution of earnings and inequality in this counterfactual (hypothesised) distribution.

Checchi and Peragine (2010a) on the other hand, developed and used two distinct non-parametric approaches, to decompose total inequality in earnings in Italy into Inequality of Opportunity and inequality of efforts. These approaches to measuring Inequality of Opportunity did not require them to make functional form assumptions about the relationship between earnings, circumstances and efforts. The first approach partitioned the population into groups of individuals with the same circumstances. Following Roemer (1998), these subgroups were referred to as types, and the Inequality of Opportunity index is used to estimate between-types earning differentials (ex-ante approach). The total inequality in earnings is then decomposed into between-type (Inequality of Opportunity) and within-type (inequality of efforts) components.

In the second approach, the population is partitioned into tranches (ex-post approach). These are groups of individuals that have exercised the same degree of effort. Checchi and Peragine (2010a) did not observe actual effort variables in their estimation of ex-post Inequality of Opportunity. They instead used the individual's percentile in their type's distribution of earnings as a proxy for effort exerted. This means that all individuals in the p^{th} percentile of their type's distribution have exerted the same degree of effort, and consequently belong to the same tranche (Roemer, 1998). Using the defined tranches, total inequality is then decomposed into a within-tranche and a between-tranche component. The within-tranches inequalities are attributed to Inequality of Opportunity, because all individuals in a tranche have exercised the same degree of effort. Therefore any differential in earnings within a tranche are due to differing circumstances.

Ferreira and Gignoux (2011) estimated Inequality of Opportunities in earnings and consumption expenditure, for six Latin American countries. Their study drew from the work of Bourguignon *et al.* (2007a) and Checchi and Peragine (2010a), and estimated ex-ante Inequality of Opportunity using both parametric and non-parametric estimation strategies. The non-parametric method applied by Ferreira and Gignoux (2011) followed that proposed by Checchi and Peragine (2010a) directly. Whereas, the parametric method, utilised the model proposed by Bourguignon *et al.* (2007a), and generated a counterfactual distribution using outcomes predicted by $\widetilde{u_i} = \exp[C_i \, \hat{\psi}]$. In the smoothed counterfactual distribution, the outcome of each individual was replaced with their predicted (conditional mean) outcome level. This smoothing eradicated all within-type differentials because individuals with homogeneous circumstances (same type) had the same predicted outcome. Inequality of Opportunity was estimated as the inequality in the counterfactual distribution.

Ferreira and Gignoux (2011) were also the first to introduce the idea of interpreting Inequality of Opportunity estimates as lower bound estimates of "true" Inequality of Opportunity from all circumstances (observed and unobserved). The rationale being, that as more circumstances are observed in the model, the greater the estimated Inequality of Opportunity will be.

In this paper, Inequality of Opportunity indices are estimated using the non-parametric and parametric strategies proposed by Checchi and Peragine (2010a) and Ferreira and Gignoux (2011). Both strategies were used because they each have certain benefits and limitations.

The non-parametric model for instance, does not make any functional form assumptions, about the relationship between outcome, circumstances, and efforts. This means that there are no concerns about endogeneity due to omitted variable bias when using this strategy (as is the case with the parametric model). A drawback of this strategy however, is that the accuracy of its estimates depends on the number of types observed and the size of the data set. Ferreira and Gignoux (2011) found that Inequality of Opportunity was overestimated when the number of observed circumstances (and types) increases. They attributed this to the increase in the number of types with few (or zero) observations, because sampling variance is relatively higher for those types.

It is therefore crucial to limit the number of circumstances used to define types, when using the non-parametric estimation strategy. Checchi and Peragine (2010a) for example, only observed one circumstance (highest parental education) in their study. Ferreira and Gignoux (2011) on the other hand, used a broad range of circumstances (ethnicity, father's occupation, father's education, mother's education and birth region) but reported both parametric and non-parametric estimations. It is important to present parametric estimates because the parametric model is not as data intensive as the non-parametric model, which means that relatively more circumstances can be observed without adversely affecting the accuracy of the estimates.

A further limitation of the non-parametric estimation strategy is that it cannot be used to estimate the partial effect of each observed circumstance on the outcomes. This is a drawback because the partial effects estimates indicate the comparative contributions of each circumstance to overall Inequality of Opportunity. The decomposition of total Inequality of Opportunity and the identification of the dominant circumstances assist in the prioritisation of policies formulated to address the unequal opportunities caused by specific circumstances. Bourguignon *et al.* (2007a) used the parametric approach to estimate partial effects, and found that parental education is the most important circumstance contributing to overall (observed) Inequality of Opportunity. This finding suggests that there would be significant reductions in unequal opportunities in Brazil, if the mechanisms that allow advantages to be transmitted from one generation to next were disrupted.

A major limitation of the parametric model (and one that has been briefly mentioned) is that the parameter estimates are for the most part biased, due to the correlation between observed circumstances and unobservable factors. Bourguignon *et al.* (2007a) used Monte Carlo simulations to gauge the likely impact of different degrees of bias on the parameters and so the Inequality of Opportunity estimates. The simulations generate a range of parameter estimates and a corresponding ninety percent confidence interval within which the true observed Inequality of Opportunity estimate lies. These results are useful, given that it is unlikely that all factors relevant to the model will be observed in datasets, because the sensitivity of the parameters and consequent Inequality of Opportunity to bias can be assessed.

The parametric and non-parametric strategies offer different estimates of Inequality of Opportunity. Therefore these strategies are taken to be complementary approaches, that can be used to evaluate the robustness of derived Inequality of Opportunity estimates (Ferreira, Gignoux and Aran, 2010; Ferreira and Gignoux, 2011).

It was also found that the ex-ante (types) and ex-post (tranches) approaches offer different estimates of Inequality of Opportunity (Checchi and Peragine, 2010a). The tension between the two approaches stems from the neutrality of the ex-ante approach to effort inequalities (Checchi, Peragine and Serlenga, 2010b). This neutrality means that any within-type redistribution leaves ex-ante estimates unchanged, since the average income of the type does not change. This is not the case for ex-post estimates, where such redistributions could possibly generate different rankings within each type's distribution of income, which would alter the composition of individuals belonging to each tranche. The fact that the two approaches yield different estimates has led to researchers selecting one of the approaches for empirical application[3].

The present study is based on the ex-ante approach to estimating Inequality of Opportunity. The rationale being that our interest lies in the portion of inequality due to circumstances (and subsequent types), and this approach offers the most direct way of estimating Inequality of Opportunity (Brunori *et al.,* 2013). The ex-ante approach is also preferred for pragmatic reasons. The ex-post approach requires each type to be divided into percentiles, in order to identify individuals who have exercised the same degree of effort. This partitioning is more data intensive than that demanded by the ex-ante approach, especially when many circumstance variables are included in the model (Checchi and Peragine, 2010a). It was therefore more appropriate to implement the ex-ante approach given the absence of a sufficient number of observations within each tranche and the impact this has on the precision of ex-post Inequality of Opportunity (Ferreira and Gignoux, 2011).

[3] See Fleurbaey and Peragine (2013) and Checchi *et al.* (2010b) for an in depth examination of the causes of the clash between ex ante and ex post notions and therefore estimates of Inequality of Opportunity.

The final set of papers relate to the preliminary analysis carried out in this study, into inequality of employment opportunities in the South African labour market. The methodology used in this preliminary analysis was drawn from studies that used other approaches to estimate Inequality of Opportunity. Barros *et al.* (2009) contributed to the quantitative analysis of Inequality of Opportunity, through their application of the dissimilarity index (used extensively in segregation literature) to the Inequality of Opportunity framework. This application allowed for the estimation of Inequality of Opportunity in discrete non-money metric measures of well-being, such as access to basic services like electricity, adequate housing and sanitation (Barros *et al.,* 2009), in addition to Inequality of Opportunity in the attainment of educational outcomes (sixth grade) for children in different Caribbean and Latin American countries. This approach required the probability of a child completing sixth grade on time to be estimated using a logistic regression of a binary outcome (=1 if the child completed sixth form on time and =0 if the child did not complete sixth form on time) on a set of circumstances. The circumstances included in the specification of the logistic regression function include gender, parent's education, number of siblings, presence of parents, area of residence (urban or rural) and per capita household income. The dissimilarity index was then estimated as the weighted average of the difference between each individual's estimated conditional probability and the average probability of completing sixth grade on time for the entire population (Barros *et al.,* 2009). Since all individuals belonging to the same type have identical conditional probabilities, the dissimilarity index simply quantifies the extent to which the probability of each type completing sixth grade on time differs from the across type average of completing sixth grade on time. The calculated index is interpreted as the share of opportunities to complete the sixth grade on time, that need to be redistributed from "better-off" to "worse-off" types so that all types are equally likely to complete sixth grade on time.

Finally, the second study related to discrete outcomes was conducted by Yalonetzky (2012), who modified the dissimilarity index proposed by Barros *et al.* (2009) and made it applicable to multidimensional outcomes. This meant that the Inequality of Opportunity framework could also be applied to the growing multidimensional inequality literature.

The next section provides a comprehensive discussion of the methodology applied in this study, to measure the effect of circumstances on labour market outcomes.

Section 3: Methodology

In the empirical application, the number of observable circumstances was limited to race, gender and highest parental education[4]. These circumstances were selected because Barros *et al.* (2009) associated them with the two sources of unequal opportunities. *Differences in social treatment* relate to unequal opportunities due to characteristics such as race and gender. These inequalities occur when certain groups are discriminated against, so that they receive inferior access to available opportunities. This second source relates to the impact of family background characteristics such as highest parental education on the opportunities available to the individual. These *differences in conditions* generate unequal opportunities, because individuals with highly educated parents tend to accumulate higher levels of human capital due to their families having better access to the necessary resources (Barros *et al.,* 2009) [5].

The three vectors corresponding to the observed circumstances are stated below, and each vector contains a finite number of categories.

$$C_{Race} = \{African, Coloured, Indian\ and\ Asian\ (Indian), White\}$$

$$C_{Gender} = \{Male, Female\}$$

$$C_{Highest\ Parental\ Education} = \{No\ Education, Primary\ Incomplete, Primary\ Complete, Secondary\ Incomplete, Matric, Hig$$

Although restricting the number of circumstances to three underestimates Inequality of Opportunity, the restriction increases the odds of there being a sufficient number of observations per type. This is important because the sampling variance is relatively higher for types containing few individuals, and this inflation leads to the overestimation of between-type inequality (Ferreira and Gignoux, 2011). The non-parametric estimation strategy is particularly affected by this overestimation, because its precision depends on the quality of the generated type conditional means, which depends on the number of individuals observed for each type. It was therefore necessary to limit the number of circumstances to race, gender and highest parental education.

---

[4] See Pellicer *et al.* (2011) for an in depth discussion of the race and parental education inequality traps faced by South Africans.

[5] Each individual has a finite set of circumstances, which can be expressed in notational form as $\Omega = \{C_1, C_2, \dots, C_n\}$. Therefore limiting the number of circumstances observed in our analysis to race, gender and highest parental education underestimates Inequality of Opportunity because they are a subset of the full set of circumstances $\Omega = \Omega_{\text{observed}}$.

The circumstances stated above, are used to partition the sample (n) into m distinct types of individuals with identical circumstances. These sub-groups are referred to as types, and each type (T) is defined by a unique set of circumstances, determined through the combination of elements from the three circumstance vectors. The circumstance vectors used in this analysis generate forty-eight distinct types, which can be described in vector form as;

$$T_1 = \{African, Male, No\ Education\}, T_2 = \{African, Female, No\ Education\}, \ldots\ and\ T_{40} = \{White, Female, Higher\ Education\}$$

Finally, a variety of economic status variables (earnings, labour market income, per capita Household Consumption and per capita household income) have been used in previous studies into Inequality of Opportunity when measured using the ex-ante approach. The primary measure of monetary well-being in this paper however is total labour income, measured as real monthly income derived from labour market related activities for the individual[6]. The justification for this is that the unequal distribution of labour market income across employed South Africans, contributes significantly to individual and household level inequality (Leibbrandt *et al.,* 2010). It therefore seems important to base this study on labour market income, as this will shed light on the factors that cause variations in the opportunities that determine how income is distributed amongst employed South Africans.

The employment status of economically active individuals is also used as an outcome of interest in this paper, to estimate Inequality of Opportunity in accessing employment. It was important to quantify the extent to which entrance into employment is determined by circumstances (using the dissimilarity index) before analysing Inequality of Opportunity in labour market income, because labour market income is a significant contributor to an individual's well-being and so having a job is valuable to the individual.

[6] Total labour market income is comprised of earnings from the main and secondary job, casual wages, self-employment income, thirteenth cheque, other bonus, profit share, "helping friends" income and extra piece-rate income (Argent, 2009).

3.1 Ex-ante (Types) Inequality of Opportunity

3.1.1 The Non-Parametric Approach

To estimate ex-ante Inequality of Opportunity non-parametrically two vectors must be defined. The first is the initial income vector of the entire population (1) and the second is a counterfactual distribution in which each individual's income is replaced by the mean income of their type (2). The first vector is expressed as;

(1) $\mathbf{X}^{A} = \{\mathbf{y_1}, \dots, \mathbf{y}_m\}$ where,

$\mathbf{y}_T = \{\gamma_{T1}, \dots, \gamma_{Tn_T}\}$ and $T = 1, 2, \dots, m$.

This vector indicates that the population has been completely partitioned into m types (*T*), defined on a set of observed circumstances. So that the aggregate of individuals belonging to each type (n_T) is equal to the total population $\sum_{T=1}^{T=m} n_T = n$.

Each element in this vector ($\mathbf{y}_T$) denotes the income distribution vector of a specific type *T,* ranked from lowest to highest income. Roemer (2006) refers to this distribution as the opportunity set available to an individual, since it represents the set of possible incomes the individual can earn if they exert different degrees of effort, conditional on their type (*T*). So it follows that a lower rank is equivalent to a lower degree of effort, and a higher rank is equivalent to a higher degree of effort.

The second vector is grounded in the utilitarian version of the ex-ante approach to estimating between type inequalities[7]. The vector is expressed as;

(2) $\mathbf{X}_{\mathbf{B}}^{\mathbf{A}} = \{\mu_{\gamma 1}, \dots, \mu_{\gamma m}\}$ where,

$\mu_{\gamma T} = \left\{\frac{1}{n_T}\sum_{i=1}^{nT}\gamma_{Ti}, \dots, \frac{1}{n_T}\sum_{i=1}^{nT}\gamma_{Ti}\right\}$ for $T = 1, 2, \dots, m$.

[7] The utilitarian ex-ante approach differs from the non-utilitarian approach in one crucial way, which is that whilst the *utilitarian approach* estimates Inequality of Opportunity through the comparison of mean outcome across types. The *non-utilitarian* approach compares the outcome distributions of the various types, through tests of stochastic dominance. Tests of stochastic dominance involve comparing the outcomes of individuals exerting the same degree of effort, but belonging to different types and aggregating these differentials into a single index. The indirect impact of circumstances (through efforts) means that the cumulative distributions of "better-off" types are expected to dominate that of comparatively "worse-off" types at every level of effort (Checchi and Peragine, 2010a).

Each element in the vector $\mathbf{X_B^A}$ denotes the average income vector of a specific type (μ_{yT}), because each individual in the sample has been assigned the average income of their specific type ($\frac{1}{n_T}\sum_{i=1}^{n_T} y_{Ti}$). The smoothing of within-type income differentials eliminates within-type inequalities. These inequalities are acceptable, because they are assumed to be caused by different degrees of effort being exerted by individuals with similar circumstances. A consequence of this smoothing is that there is inequality neutrality within each type, and equality of opportunity is realised when the mean income of each type is the same, $\mu_{y1} = \mu_{y2} = \cdots = \mu_{ym}$. This equality of opportunity equivalency is often not observed, and so the inequality estimated from vector 2 can be completely attributed to the unequal opportunities stemming from the observed circumstances.

For a given scalar inequality index **I,** non-parametric absolute and relative Inequality of economic opportunity index estimates are defined as:

$$\mathbf{IOp}_A = \mathbf{I}\left(\{\mathbf{X_B^A}\}\right) \tag{2.1}$$

$$\mathbf{IOp_R} = \frac{\mathbf{I}(\{\mathbf{X_B^A}\})}{\mathbf{I}(\{\mathbf{X^A}\})} \tag{2.2}$$

The absolute Inequality of economic opportunity index ($\mathbf{IOp}_A$), measures the overall effect of the observed circumstance on the labour market income. This estimate is derived by applying an inequality index (**I**) over the counterfactual distribution $\mathbf{X_B^A}$. The relative Inequality of economic opportunity index estimate ($\mathbf{IOp_R}$) in contrast, is concerned with the share of inequality due to observed unequal opportunities. It therefore compares absolute Inequality of Opportunity, to the inequality observed in the actual income distribution.

3.1.2 The Parametric Approach

The parametric approach to estimating Inequality of Opportunity is derived from the Mincerian earnings regression that estimates labour market earnings (y_i) as a function of the characteristic (x_i) of an individual. These characteristics can be strictly partitioned (according to the Inequality of Opportunity framework) into one of two categories, circumstances ($\mathrm{C_i}$) or efforts ($\mathrm{E_i}$). The Inequality of Opportunity modified Mincerian earnings regression is expressed by equation 1 below.

$$\ln(y_i) = \alpha\mathrm{C_i} + \beta\mathrm{E_i} + \mu_i \tag{3}$$

In the structural model (eq. 3) the full set of circumstance and effort variables are identified and used to estimate individual income. The $\mathbf{C}_i$ vector consists of variables exogenous to the individuals such as gender, race, parental education, parental wealth, father's occupation and place of birth. The $\mathbf{E}_i$ vector on the other hand, consists of variables the individual has control over. These include education level, occupation type and labour market status. The residual term (μ_i) is said to be an independent and identically distributed (i.i.d) random variable with a zero mean and no correlation with factors included in $\mathbf{C}_i$ and $\mathbf{E}_i$, so that $E(\mu|\mathbf{x}) = 0$. This is a fundamental assumption which is necessary if one is to estimate equation 3 using Ordinary Least Squares (OLS). If μ_i is not orthogonal to $\mathbf{C}_i$ and $\mathbf{E}_i$, the parameter estimates $\hat{\alpha}$ and $\hat{\beta}$ will be biased, and it would be incorrect to interpret the estimated coefficients as the marginal effect of $\mathbf{C}_i$ and $\mathbf{E}_i$ on individual income. This is evidently the case with estimates from the structural model.

The orthogonality assumption does not hold in equation 3, partly because the structural model assumes that $\mathbf{C}_i$ is uncorrelated with $\mathbf{E}_i$. This is incorrect according to Bourguignon *et al.* (2007a) because it would mean that the only effect of $\mathbf{C}_i$ on income is the direct effect captured by the term α. This fails to account for the indirect effect of $\mathbf{C}_i$ through $\mathbf{E}_i$ (Roemer, 1998). An example often given to illustrate this dependence is the education level attained by the individual, given that own education is classified as an effort variable in the Inequality of Opportunity framework.

The unequal distribution of educational outcomes, on account of differentials in family background, has been the focus of a substantial body of literature. Breen and Jonsson (2005) for example, found that individuals from wealthy families or educated parents are more likely to obtain higher levels of education than individuals without these familial characteristics. This advantage may occur through a number of possible channels, one of which is the provision of superior home inputs (books and private tuition) into the educational production function (Rønning, 2011). The relationship between own education and family background alludes to the effort variables being at least partially determined by the circumstance variables $corr(C, E) \neq 0$, and as a consequence endogenous to the structural model.

Bourguignon *et al.* (2007a) therefore suggest that the structural model (eq. 3) be supplemented with auxiliary regressions of the various effort variables on the circumstance vector (eq. 3.1).

$$\mathbf{E}_i = \theta \mathbf{C}_i + \nu_i \qquad (3.1)$$

The error term (ν_i) in equation 3.1 captures unobserved circumstances and efforts, and is assumed to be i.i.d and to have a zero mean. Just as with the structural model, using OLS to estimate equation 3.1 implicitly assumes that $\mathbf{C}_i$ and ν_i are orthogonal.

Putting together the structural model (eq. 3) and the auxiliary regressions (eq. 3.1) forms a system of equations from which estimates of (both the direct and indirect impact of) circumstances on individual income can be derived. These estimates will only be unbiased if the aforementioned orthogonality assumptions hold, in addition to the assumption that the errors in (eq. 3) and (eq. 3.1) are uncorrelated $cov(\mu_i, \mathrm{v}_i) = 0$. These assumptions do not hold in empirical application, primarily because relevant but unobserved C and E variables may be correlated with observed C and E variables so that $E(\mu_i|\mathrm{C}_i, \mathrm{E}_i) \neq 0$ and $\mathrm{E}(\mathrm{v}_i|\mathrm{C}_i) \neq 0$ [8]. This failure makes estimating valid parameters from this system of equations challenging.

The instrumental variables approach is the conventional solution to the endogeneity problem. This solution requires that an instrumental variable (z) that is correlated with observed variables, but uncorrelated with individual income, be introduced into the system (Cameron and Trivedi, 2010). Provided that the instrumental variable (z_i) is correlated with the endogenous regressor (observed C and E) and uncorrelated with the error term, the instrumental variable estimator will be consistent. Although Bourguignon *et al.* (2007a) acknowledge that the use of instrumental variables would theoretically purge the full system of endogeneity. However, they foresee serious difficulties in identifying appropriate instrumental variables for endogenous circumstance and effort variables. This opinion is based on the fact that, when instrumental variables are used in empirical studies, researchers almost always disclose that they are not perfect instruments and that the subsequent estimation should be treated with caution[9].

In light of this Bourguignon *et al.* (2007a) propose that the auxiliary regression (eq. 3.1) be substituted into the structural model (eq. 3), producing the reduced form equation stated below. In estimating the reduced form equation, it is assumed that the observed circumstance variables C_i are uncorrelated with the composite error term ε_i. The composite error term captures unobserved circumstance and the direct effect of unobserved effort variables. It also includes random genetic variation, luck and measurement error (Bourguignon *et al.,* 2007a).

$$ln(\mathrm{y}_i) = \psi \mathrm{C}_i + \varepsilon_i \text{ where,}$$

$$\psi = \alpha + \beta\theta \text{ and } \varepsilon_i = \beta \mathrm{v}_i + \mu_i \tag{4}$$

[8] When estimating (eq. 3) and (eq. 3.1), relevant C and E variables are not included either because of inadequate datasets (which leads to these factors being unobserved) or because they are impossible to measure. These factors then form part of the error term and can introduce endogeneity into the system.

[9] Bourguignon *et al.* (2007a) do not use the instrument variable approach to solve the systems endogeneity problems. They instead estimate the structural model as it is and run Monte Carlo simulations to evaluate the impact of different degrees of bias on the parameters and consequently the Inequality of Opportunity estimates. They then report a range of the reduced model parameter estimate ($\hat{\psi}^*$) and the corresponding interval of Inequality of Opportunity estimates.

From the reduced form model (eq. 4), it is apparent that there will be equality of opportunity when circumstances do not affect income either directly or indirectly through effort or random shocks, so that $\psi = 0$. This however is not the case, and given that the aim of this paper is to estimate the overall portion of total inequality due to differences in the observed circumstances Bourguignon *et al.* (2007a) recommend estimating the reduced form equation because the parameter ψ accounts for the direct (α) and indirect (β θ) effects of observed circumstances. The residual component (ε_i) is then attributed to unobserved circumstance, the direct effect of unobserved effort variables, random genetic variation, luck and measurement error (Bourguignon *et al.,* 2007a).

An advantage of this solution is that one no longer has to contend with the endogeneity issues of the full system. This approach nevertheless has the same drawback as that of the full system*:* the unobserved determinants of earnings contained in the error term (ε) are likely to be correlated with the observed circumstances C_i. This is because although the observed circumstances are economically exogenous they are not econometrically exogenous. Therefore the reduced model may also yields biased parameter estimates (ψ) as a result of the error term not being orthogonal to the circumstance regressors[10].

In spite of the likely bias in the estimated parameter, Ferreira and Gignoux (2011) advocate for the continued use of the reduced model, as long as the estimated opportunity shares are interpreted as lower-bound estimates of overall Inequality of Opportunity[11]. This stipulation allows biased parameter estimates (ψ) to be used in computing Inequality of Opportunity, because the parameter estimate $\hat{\psi}$ captures not only the impact of the observed circumstances but also of other unobservable factors (circumstance or effort) correlated to these observed circumstances. It is important to capture these effects especially if the majority of the variation in the residual component (ε_i) is due to unobserved circumstances. This is because the parameter estimates will account for some of the partial effects of the unobserved circumstances on labour market income, and the resulting Inequality of Opportunities estimate can be better attributed to the full circumstance vector C*[12].

[10] The use of instrument variables when estimating the reduced form model was dismissed on the basis of data availability which made finding an instrument difficult.

[11] According to Ferreira and Gignoux (2011) the vector C* is composed of all circumstances i.e. all factors determining an individual outcome which are exogenous to the individuals. This definition of C* implies that the vector of circumstances observed in this paper is a strict subset of the full circumstance vector, in that $C_i < C_i^*$. Therefore, as more relevant circumstances are observed it is guaranteed that Inequality of Opportunity estimates, both parametric and non-parametric will be higher.

[12] The same principle holds when the bias is caused by relevant effort variables being omitted from equation 4, when the $corr(C_{Observed}, E_{Unobserved}) \neq 0$. This is because the portion of the omitted effort variables correlated with the observed circumstances, allows the parameter estimates ($\hat{\psi}$) to capture some of the partial effects of the unobserved effort variables. The bias can then be attributed to the observed circumstances generating unequal opportunities in acquiring labour market income indirectly, through the unobserved effort variables.

Consider an example where the estimated reduced model parameter vector (ψ) is positively biased. In addition, let this upward bias be as a result of the unobserved relevant variables being positively related to income and the observed circumstances being positively correlated with unobserved variables contained in the error term, such that $\text{corr}\,(C_i, \varepsilon_i) > 0$.. In this scenario it would be appropriate to adhere to Ferreira and Gignoux (2011) suggestion, and interpret the opportunity share estimated using the biased circumstance parameters as a lower-bound estimate of the total Inequality of Opportunity due to all circumstances (C*). This interpretation would be fitting because the positive omitted variable bias was caused by the observed circumstance being partially correlated with unobserved characteristics (of which unobserved circumstances are included). Thus the upward bias is to some extent a result of the observed circumstances capturing the impact of unobserved circumstances on the outcome of the individual. This interpretation is even more appropriate when the variation in ε_i is primarily due to unobserved circumstances.

Despite its endogeneity issues, the reduced form equation has become the standard approach for researchers seeking to estimate the share of total inequality due to inherited characteristics, and it is the model used in this paper.

The first step in quantifying observed Inequality of Opportunity is to estimate the reduced form equation 4 using OLS. The estimated parameters ($\hat{\psi}$) are then used to generate a smoothed counterfactual income distribution (eq. 5).

$$\tilde{u_i} = \exp[C_i\,\hat{\psi}] \tag{5}$$

In the counterfactual distribution ($\tilde{u_i}$) individuals with the same circumstances are assigned the same predicted (conditional mean) income level. This simulated distribution is analogous to a non-parametric smoothed distribution (eq. 2), where individuals were assigned the mean income of their type. This is because in both cases the smoothing of income within each type eliminates all within-type income differentials (due to differences in effort) and we are only interested in the outcome differentials of individuals with different circumstances.

The smoothed distribution of income is the counterfactual distribution in the parametric approach, and it forms the basis for inequality of economic opportunity index estimates in this analysis.

$$\mathbf{IOp}_A = \mathbf{I}(\{\tilde{u_i}\}) \tag{5.1}$$

$$\mathbf{IOp}_R = \frac{\mathbf{I}(\{\tilde{u}_i\})}{\mathbf{I}(y)} \tag{5.2}$$

For a given inequality measure **I**. The counterfactual distribution is used to estimate the parametric absolute Inequality of economic opportunity index (eq. 5.1) and consequently calculate the share of overall inequality attributable to the observed circumstances (eq. 5.2). This share is formally referred to as the relative Inequality of economic opportunity index and the remaining share of overall inequality (1-$\mathbf{IOp}_R$) is attributed to factors contained in the residual term (ε_i).

Ex-ante Inequality of Opportunity can be calculated using the non-parametric or the parametric estimation strategies described in Section 3.1.1 and Section 3.1.2 respectively. In this paper, both estimation strategies are applied in order to evaluate the robustness of the estimated absolute and relative opportunity shares[13]. Results from both strategies are presented in Section 6.

3.1.3 Computing Inequality of Opportunity: The Decomposition of Total Inequality

In order for total inequality of income to be decomposed, an inequality index with certain axiomatic properties has to be selected. Foster and Shneyerov (2000) proposed the General Entropy Class of measures GE (.) and more specifically the Theil-L index, also known as the Mean Logarithmic Deviation GE (0). This is because it is the only member of the general entropy class to use weights based on the type's population shares, in addition to all the properties needed to handle smoothed distributions. The properties are[14]:

(i) Symmetry (anonymity): the index must be invariant to the permutation of any two individuals within a type i.

(ii) Transfer principle: There are two parts to the transfer principle. The first is that the index must be invariant to any transfers within the outcome distribution of a type, if the transfers do not alter the average outcome of the type (within-type transfer insensitivity). The second is that the index is allowed to rise slightly when a transfer is made between individuals belonging to different types (between-type transfer principle).

[13] Given that the parametric approach requiring strong functional form assumptions (especially about the relationship between C_i and ε_i) and the non-parametric approach is imprecise when few observations are observed for each type.
[14] See Foster and Shneyerov (2010) and Ferreira and Gignoux (2011), for a full discussion of the properties of the Generalised Entropy Class of measures and specifically the Mean Logarithmic Deviation GE (0).

(iii) Scale invariance: the index must be invariant to the rescaling of the outcome by any positive scalar.

(iv) Population replication: the index must be invariant to population replication.

(v) Additive decomposability: the index must allow inequality of outcome to be fully decomposed into that due to circumstance differentials and that due to differentials in unobserved but relevant factors.

(vi) Normalisation: the index must take the value of zero if there is equality of opportunity i.e. if the distribution (eq. 2) has the characteristic that the mean outcomes of all of the types are equal.

The parametric and non-parametric Inequality of Opportunity estimates presented for labour market income are to be interpreted as lower bound estimates of "true" Inequality of Opportunity, as is customary in the empirical literature. This interpretation is valid because the inclusion of omitted circumstances into our vector of circumstances is unlikely to reduce the estimated opportunity shares (Checchi and Peragine, 2010a; Ferreira and Gignoux, 2011)[15].

3.1.4 The Strengths and Weaknesses of the Ex-Ante Model

The ex-ante model described in this section has a number of strengths. The first is that computationally it only requires a simple regression of the outcome of interest on a set of observed circumstances. Seeing as the objective is to estimate the share of inequality due to a set of observed circumstances, this model allows for the simple decomposition of overall inequality. This is particularly advantageous for researchers attempting to estimate Inequality of Opportunity for developing countries, which often times have few representative national surveys with limited information on the social origins of individuals. All that is required to estimate the reduced model and to partition the population into types is survey data on circumstance variables such as gender, race, region of birth or some parental status variables (income, occupation or education).

Secondly, because it is highly unlikely that we will ever be able to observe the full set of circumstances (C_i^*) and therefore estimate "true" Inequality of Opportunity, the opportunity share estimated from the observed circumstances (which are a sub set of C_i^*) is a good starting point for any study into "unfair"

[15] See Ferreira and Gignoux (2011) for the full proof of the proposition that the estimated observed Inequality of Opportunity is the lower bound estimate of the "true" Inequality of Opportunity estimated when the full set of circumstances (C*) are observed in our model.

inequalities. These lower-bound estimates of "true" Inequality of Opportunity estimates can also be of use to policy makers seeking to compensate individuals for "unfair" inequalities. A further advantage of this model is that it can be used to estimate the contribution of each observed C_i variable to the opportunity share. This identification is important to policy makers wanting to formulate targeted policies.

The ex-ante model also has a number of weaknesses. The first is that the reduced equation suffers from endogeneity issues stemming from its failure to capture all relevant variables. Even though the qualitative solution of interpreting the opportunity share (calculated from biased parameter estimates) as the lower-bound estimate of opportunity inequality is intuitively appealing, the more econometrically rigorous Niehues and Peichl (2012) solution is beyond the scope of this paper primarily because it requires a large panel data set.

Another weakness is that the ex-ante model presented in this paper can also only be used for continuous outcomes, so researchers interested in estimating Inequality of Opportunity for discrete ordinal outcomes are unable to use this model. This led to the application of the dissimilarity index, which gauges the extent to which existing opportunities are "fairly" allocated by assessing the impact of inherited circumstances on multi-dimensional measures of well-being. In the dissimilarity index framework, equality of opportunity is achieved when all types are equally likely to "achieve" the different levels of outcome. Therefore in the case of a binary variable **x,** where **x**=1 if the individual fall below the poverty line and **x**=0 if the individual fall above the poverty line, there is equality of opportunity if each type is equally likely to be "poor". See Barros *et al.* (2009), Silber and Yalonetzky (2011), Asadullah and Yalonetzky (2012) and Yalonetzky (2012) for applications of the dissimilarity index to the Inequality of Opportunity framework.

The final limitation has to do with the empirical application of the ex-ante approach: in order to generate the counterfactual distributions used to estimate Inequality of Opportunity, only individuals with non-missing entries in their circumstance vector can be included in the selected sample. This restriction can dramatically reduce the sample size and more importantly the number of observations per type. This is problematic given that the preciseness of any estimate intended to summarise between-type differentials in outcome diminishes as the number of observations per type decreases. Therefore, this data driven limitation is only an issue if the dataset being used does not have a sufficient number of people or if it has a lot of missing/invalid responses to relevant variables.

In spite of its weaknesses, the ex-ante model is the model used in this analysis to estimate Inequality of Opportunity in South Africa. The next sub-section contains the methodology used to estimate inequality of employment opportunities for economically active individuals. It was important to carry out this analysis as a precursor to the analysis of Inequality of Opportunity in labour market income, because labour market income is a significant contributor to an individual's well-being. Therefore it is critical to quantify the extent to which entrance into employment (a state that is unquestionably advantageous for economic well-being), is determined by inherited characteristics and the opportunities such characteristics bestow on the individual. This partially addresses the ex-ante models inability to estimate Inequality of Opportunity for all employed individuals, since the access to employment analysis accounts for all employed individuals, even if they reported zero income.

3.2 The Dissimilarity Index (Barros et al., 2009)

The dissimilarity index is implemented when the outcome of well-being is discrete, and it gauges the extent to which existing opportunities are "fairly" allocated.

Therefore, for a given outcome y_i which takes on two values indicating whether an economically active individual is employed (=1) or unemployed (=0), there is equality of opportunity if each type's probability of accessing employment (p^T) is equal to the entire populations average probability of accessing employment (p^*). This definition of equality of opportunity alludes to the importance of access probability gaps (which are the absolute differences between p^T and p^* for each type) in the dissimilarity index approach to estimating Inequality of Opportunity.

The calculated dissimilarity index (%) is interpreted as the share of opportunities to gain employment which need to be reallocated amongst the different types so that all types are equally likely to be employed[16]. This implies that in the dissimilarity index framework, the greater the inequality in accessing employment unambiguously due to circumstances, the greater the calculated dissimilarity index (D). Equation 6 states the dissimilarity index applied to the smoothed distribution of probabilities (p^T).

$$\mathrm{D} = \frac{1}{2p^*}\sum_{T=1}^{m} \mathrm{w}^T |p^T - p^*| \qquad \text{where,} \tag{6}$$

$T=1, 2, ..., m$ and $\mathrm{w}^T = \frac{1}{n_T}$ or sampling weights

In equation 6, p^T denotes type T 's probability of accessing employment and p^* is the average probability of accessing employment for the entire population. The access probability gap of each type is

[16] The estimated opportunity share are also interpreted as lower-bound estimates because as more relevant circumstances are observed it is guaranteed that share of opportunities to gain employment (as estimated by the dissimilarity index will be higher.

then weighted by the proportion of the population belonging to the type ($\frac{1}{n_T}$), so that the index is simply a scalar of the weighted access probability gaps

This stated version of the dissimilarity index was selected over the more recent Yalonetzky (2012) adaptation, because his dissimilarity index is more suitable for multidimensional outcomes than it is for binary outcomes where one state is "preferred" over the other state[17]. Thus the dissimilarity index stated in equation 6 is the simplest method by which to estimate type dissimilarities in opportunities, given that employed is clearly the preferred state (Yalonetzky, 2012).

In order to estimate the dissimilarity index (eq. 6), the probability of each individual's ability to access employment conditional on their circumstances (x_k) is estimated parametrically, using a logistic model (eq. 7) (Barros *et al.,* 2009).

$$Ln\left(\frac{P(I=1|x_{1,\dots,}x_m)}{1-P(I=1|x_{1,\dots,}x_m)}\right) = \sum_{i=1}^{m} x_k\beta_k \tag{7}$$

$$\hat{p}_i = \frac{Exp(\ddot{\beta}_0+\sum_{k=1}^{m} x_{ki}\ddot{\beta}_k)}{1+Exp(\hat{\beta}_0+\sum_{k=1}^{m} x_{ki}\hat{\beta}_k)} \tag{7.1}$$

$$\bar{p} = \sum_{i=1}^{m} w_i\hat{p}_i \tag{7.2}$$

This model specification is used because the logistic regression is linear in parameters and it is these estimated parameters that are used to calculate $\hat{p}_i$. This probability is also an estimate of each type's ability to access employment, given that all individuals sharing the same set of circumstances have the same estimated employment access probability. The average probability of accessing employment for the entire population (eq. 7.2) is also calculated as the weighted average of the predicted access probabilities, and it is an estimate of the across type probability of accessing employment (p^*).

3.2.1 Properties of the Dissimilarity Index (D) [18]

Property 1: The dissimilarity index is greater than or equal to zero, D≥0.

The access probability gaps are $|p^T - p^*| \geq 0$ and so $\sum_{T=1}^{m} w^T|p^T - p^*| \geq 0$.

[17] The Yalonetzky (2012) dissimilarity index has its foundations in the literalist definition of equality of opportunity, and is based on a statistic from a test of homogeneity between the multinomial distributions of the different types (Yalonetzky, 2012).

[18] See Barros *et al.* (2010) for full proofs of the stated properties.

Property 2: If there is perfect between-type equality in access opportunities, D=0.

There is perfect equality when the predicted access rate of each type (or individual) is equivalent to the average population access rate i.e. $p^T = p^*$.

Property 3: If there is perfect between-type inequality in access opportunities, D=1.

There is perfect inequality when one type in the population attains one state and the rest of the population (all the other types) attains the other. In the case of employment status there is perfect inequality when one type is employed or unemployed and the rest of the types in the population T-1 are unemployed or employed.

Property 4: The dissimilarity index is insensitive to balanced increase in opportunities.

This is evident when there is a *balanced* increase in opportunities, which occurs when the predicted probability of accessing employment increases for each type in such a way as to preserve the original access rate distribution. The dissimilarity index does not change. The dissimilarity index is insensitive to such an increase given that balanced increases do not alter the proportion of the population in each type or the proportion of the population accessing employment (Yalonetzky, 2012).

Property 5: The dissimilarity index is scale invariant.

Rescaling the outcome by some scalar will not alter the dissimilarity index. This scale invariance is due to the probability of accessing employment conditional on the set of circumstances (p^T) being bounded so that it always lies in the [0, 1] interval.

Property 6: The dissimilarity index exhibits anonymity or symmetry.

The dissimilarity index is invariant to individuals switching between the two dichotomous states of employment and unemployment. This invariance is due to the unchangeable nature of circumstances, which means that this switch can only occur between individuals with homogenous circumstances. As a consequence each switch occurs within each type, so each type's proportion in relation to the entire population, as well as the proportion of each type in the two states and the probability of accessing employment for each type (p^T), will be unaffected. The overall result is that the dissimilarity index will remain unchanged.

Property 7: The dissimilarity index is invariant to population replication.

The dissimilarity index remains unchanged if the population is replicated k times. This is because each type's proportion in relation to the entire population, as well as the proportion of each type in the two states and the probability of accessing employment for each type (p^T), will be unaffected by such a replication.

Property 8: The dissimilarity index is sensitive to transfers between types.

The dissimilarity index is sensitive to transfers of states between types. In the dissimilarity index framework a transfer takes place when an individual in type A becomes employed and as a result an individual in type B takes their position by becoming unemployed. This transfer is referred to as a transfer of an instance of unemployment. The sensitivity of the dissimilarity index to transfers is not necessarily a "bad" characteristic, but it depends on the types between which the transfer occurs. Two transfer scenarios will be presented below.

Scenario 1: The transfer of an incidence of unemployment from a "disadvantaged" type to a more "advantaged" type. This transfer would result in a decrease in the dissimilarity index, due to the predicted access rate of employment of each type converging to the population average access rate and the resulting decline in the weighted average of the access rate gaps.

Scenario 2: The transfer of an incidence of unemployment from a more "advantaged" type to a "disadvantaged" type. This transfer would result in an increase in the dissimilarity index, due to the predicted access rate of employment of each type diverging even further from the population average access rate and the resulting increase in the weighted average of the access rate gaps.

The Barros, *et al.,* (2009) dissimilarity index framework described in this sub-section, is used to gauge the extent to which individuals in South Africa are blocked from a state that would benefit them (employment), due to factors outside of their control. This expansion is necessary for any study into inequality that aims to be comprehensive, because labour market income contributes significantly to individual well-being.

Thus, in summary, the two Inequality of Opportunity frameworks presented in Section 3 are applied to the National Income Dynamics Study (NIDS) Wave 1 cross-sectional data set, and used to estimate the extent to which circumstances affect an individual's ability to acquire labour market income (3.1) and access employment (3.2) in South Africa.

The next section provides information on the NIDS data set and includes a discussion of the criteria used to select the sample on which this analysis of unequal opportunities was based.

Section 4: Data and Sample Selection

This empirical study utilised cross-sectional data from the National Income Dynamics Study (NIDS) panel dataset produced by the South Africa Labour and Development Research Unit (SALDRU) at the University of Cape Town (NIDS, 2008; 2010-2011; 2012). This panel study was commissioned by the Presidency to track a representative sample of South African households over time, with the primary objective of monitoring the well-being (both economic and non-economic) of these households and of the individuals residing within them. The sample was selected in two stages. The first involved the random selection of 400 primary sampling units (PSU's) from the master sample[19]. Then, 8 clusters containing the households chosen to be included in the sample were drawn from the selected PSU's.

NIDS is the ideal dataset with which to analyse Inequality of Opportunity in South Africa, because it contains substantial information on the income, family background, socio-economic factors and demographic characteristics of 28226 individuals. The panel feature also has the added advantage of allowing the pooling of individual information collected at different points in time. This is valuable given that the ex-ante model can only be applied to individuals with non-missing information on the observed circumstances.

Although the NIDS dataset is currently comprised of three waves (collected in 2008, 2010-2011 and 2012), the results section of this paper focuses on Inequality of Opportunity in 2008. The primary reason for why the panel aspect of NIDS was not employed in this analysis is that any study into the dynamics of unequal opportunities requires a long panel data set for there to be some variation in an individual's "responsible" (effort) characteristics. At its current stage, with only two year intervals between the three available waves, the NIDS dataset does not meet the requirements necessary for Inequality of Opportunity dynamics to be studied in South Africa.

In order for the empirical analysis to yield estimates representative of the South African population, the probability of an individual being included in the NIDS sample was calculated using a two-stage cluster survey design[20]. Post stratification weights were then utilised to adjust the NIDS sample so that the age-sex-race marginal totals of the sample correspond to that of the South African population (Wittenberg, 2009).

[19] Statistics South Africa (STATS SA) compiled the master sample of 3000 PSU's from each of the 53 district councils that the South African population had been partitioned into.

[20] The two-stage cluster survey design involved the calculation of, the probability of sampling each of the PSU'S drawn from the master sample and the probability of sampling each of the households in the selected PSU's (Wittenberg, 2009). This final probability corrects for household non-response. Finally, the sample design weights used during the empirical analysis were calculated as the inverse of the calculated probability of inclusion.

4.1 Sample Selection

The Inequality of Opportunities sample is restricted to adults aged between twenty-one and fifty-nine years old, because individuals in this age group tend to be finished with schooling and are more likely to be economically active[21]. The fifty-nine year old upper bound age restriction was set due to the high rates of retirement and exit out of the labour force after the age of fifty-nine, as a result of sixty being the age at which individuals in South Africa become eligible for pensions (Ranchhod, 2009). The labour market status of the individual is important in this study which firstly looks at inequalities in the opportunity to access employment (for economically active individuals) and then the circumstances related inequalities in the labour market income distribution of employed individuals.

The original sample is further restricted to individuals with non-missing entries for their outcome and circumstance variables[22] The practice of excluding individuals with missing entries, for the observed circumstances, is in accordance with previous empirical studies into Inequality of Opportunity (Checchi and Peragine, 2010a; Ferreira and Gignoux, 2011; Singh, 2012). The justification being, that an individual cannot be assigned to a type without a full set of circumstances. This inability to identify distinct types makes it impossible to compute between-types inequality (opportunity inequality).

Individuals with missing outcomes were also excluded from the sample because it is impossible to trace the full outcome distribution of a type, when some individuals within the type have missing outcome entries. This is problematic given that Roemer's ex-ante mean-equalisation definition of equality of opportunity is based on the distribution of outcomes within each type (Ferreira and Gignoux, 2011).

Table 1 contains a list of the circumstances vectors used to estimate Inequality of Opportunity in this study. The number of circumstances and the number of categories within each circumstance was limited, to ensure that an adequate number of observations were observed for each type. This is important because types with low number of observations have relatively high sampling variances, which results in the overestimation of between-type inequalities when the non-parametric estimation strategy is implemented.

---

[21] An individual is classified as economically active in the labour force if they are employed, if they are actively searching for employment or if they are discouraged workers who are not actively searching for employment (Ranchhod, 2009). Therefore in this paper we use the broad definition of unemployment, where the unemployed are those who would like to work, regardless of whether they are actively searching for work or not (Kingdon and Knight, 2004).

[22] Individuals with invalid responses for the observed circumstances were also excluded from the sample. A response is said to be invalid if it does not allow the individual to be assigned to one of the forty-eight distinct types, defined by the observed circumstances. Some of the invalid responses observed in the NIDS dataset are; Don't Know, Refused and Not Applicable.

Table 1. Definition of Circumstance Variables

Gender	
	Male
	Female
Race	
	Black
	Coloured
	Indian
	White
Highest Parental Education	
	No education
	Primary incomplete
	Primary complete
	Secondary incomplete
	Matric
	Higher education

Notes: The Indian category refers to individuals who reported their race as either Indian or Asian.

Table 2 presents the number of types observed for each of our samples and the mean number of observations per type. The results show that approximately 23 to 28 percent of our samples have types with fewer than five observations. This finding is problematic according to Ferreira and Gignoux (2011), who advocate for the use of the parametric estimation strategy, in the assessment of the sensitivity of non-parametric Inequality of Opportunity estimates to these low observations. The parametric strategy is a valid evaluation tool, because it is not subject to the high sampling variance observed for types with low observations.

Table 2. Description of the Sample Partition

	Access to Employment	**Labour Market Income**
Maximum number of types	48	48
Number of types observed	48	47
Mean number of observations per type	150.48	95.60
Proportion of types with fewer than 5 observations	0.23	0.28
Proportion of types with fewer than 10 observations	0.35	0.36
Proportion of types with fewer than 20 observations	0.40	0.43
Sample Size (n)	**7223**	**4493**

Notes:
1. Types are defined by crossing the observed circumstance variables.
2. Own calculations using sample data

The sample selection criteria and resulting sample size for each outcome before and after the restrictions are reported in **Table 3**. The first row of **Table 3** shows that the original sample for the access to employment analysis consists of twenty-one to fifty-nine year olds who are economically active. Whereas the original sample for labour market income, consists of employed twenty-one to fifty-nine year olds. The second row of **Table 3** shows the reduced sample sizes once individuals with missing or invalid responses to the observed circumstance variables are excluded. Further analysis showed that missing or invalid highest parental education entries were the cause of the reduction in sample size (see **Table 1A**in the Appendix)[23]. This finding is expected given that questions regarding parental education is require respondents to recall parental information, which can prove to be challenging. A look at the distribution of the parental education variable for the restricted sample shows that the largest percentage of the

[23] Given that the individual must have non-missing observed circumstances to be included in the sample, Marrero and Rodriguez (2012) expect the average sample size to consist of approximately 2500 individuals. This average sample size figure was calculated from the sample sizes used to estimate Inequality of Opportunity in previous empirical literature.

sample had parents with no education and the smallest percentage had at least one parent with higher education (see **Table 2A** in the Appendix).

Table 3. The Outcome and Respective Sample Sizes 2008-2012

	Access to Employment	**Labour Market Income**
Sample selection criteria		
Original sample size of 21 to 59 year-olds	7610	5248
Of those observations with non-missing and valid circumstances	7223	4973
Of those observations with valid outcomes	7223	4493
Share of original sample	0.9491	0.8561

Notes:

1. The original sample size was 21 to 59 year olds who are economically active (consists of employed and the unemployed using the broad definition of unemployment) for **the Access to Employment** outcome and the employed for the **Labour market income** outcome.
2. Outcomes: - The labour market income outcome is a continuous variable and the sample was restricted to individuals with positive earnings. The **employment status** outcome is binary in nature with the individual being assigned the value 1 if they are classified as employed and 0 if they are classified as unemployed using the broad definition.
3. Circumstances: - To reduce the number of 21 to 59 year-olds with missing or invalid responses, circumstance values were drawn from the three available Waves of the NIDS dataset. This increased the access to employment sample by 3111 (from 4112 to 7223) and the labour market income sample by 2183 (from 2790 to 4973). If there were discrepancies in the responses to the gender and race circumstances across the waves, the most recent response was selected. This however was not the case for parental education, where the highest education level was calculated as the average of the responses given in the three waves.
4. Own calculations using sample data.

Finally, employed individuals who reported zero labour market income were excluded from the final sample (see Row 3 of **Table 3**). [24] These individuals had to be excluded from the analysis of Inequality of Opportunity in acquiring labour market income, because the reduced model (4) and the mean log deviation inequality index censor outcome variables to positive non-zero values.

[24]Although zero labour market income responses are invalid for our analysis, they may be legitimate responses if the individuals were sick or seasonally out of work in the months before they were interviewed.

Although the final selected samples are adequate for our analysis, it is important to ensure that the restricted samples do not differ from the comparable original samples[25]. This would be problematic because the restricted samples will no longer be representative of the population.

4.1.1 The Issue of Inference: Representative Samples

Column 1 and 2 of **Table 4** compare the full and selected samples used to estimate the extent to which access to employment is "fairly" distributed in South Africa. It was found that the restricted samples do not statistically differ from the full original and comparable samples, on any of the specified characteristics. This was not the case for the full and restricted samples used to estimate opportunity inequalities in labour market income for 2008, see Column 3 and 4. The descriptive statistics show that the samples are similar across all of the characteristics except for labour market income.

It is not surprising that the restricted sample statistically differed from the full sample on this particular characteristic, given that employed individuals who reported zero labour market income were excluded from the final sample used in the Inequality of Opportunity analysis. The overall result of these exclusions is that the labour market income of the restricted sample is found to be significantly higher than that of the full sample, for the labour market income outcome.

The similarity of the restricted and the comparable full samples is evidence that the restricted sample is also representative of the population.

The next section contains a full discussion of the results from the empirical examination of unequal opportunities in accessing employment and in labour market income, for South Africa.

[25] The original samples of twenty-one to fifty-nine year olds (stated in **Table 3**) were selected through random sampling (Wittenberg, 2009). This means that the original samples are representative of the South African population.

Table 4. Investigating the Representativeness of the Selected Samples

	Access to Employment		Labour Market Income	
	Full sample	**Restricted sample**	**Full sample**	**Restricted sample**
Age	36.07	36.10	37.43	37.36
	(9.87)	(9.88)	(9.84)	(9.72)
Years of schooling	10.44	10.46	10.53	10.71
	(3.73)	(3.76)	(3.87)	(3.81)
Gender				
Male	0.49	0.49	0.55	0.56
	(0.50)	(0.50)	(0.50)	(0.50)
Female	0.51	0.51	0.45	0.44
	(0.50)	(0.50)	(0.50)	(0.50)
Race				
Black	0.78	0.79	0.74	0.74
	(0.41)	(0.41)	(0.44)	(0.44)
Coloured	0.09	0.08	0.09	0.09
	(0.28)	(0.27)	(0.29)	(0.29)
Indian	0.02	0.03	0.03	0.03
	(0.16)	(0.16)	(0.17)	(0.17)
White	0.10	0.11	0.13	0.14
	(0.30)	(0.31)	(0.34)	(0.35)
Labour market earnings	3547.80	3635.33	4980.03	5439.20
	(7457.46)	(7607.62)	(8422.30)	(8762.86)
Household income (per capita)	3225.36	3278.77	4045.83	4298.16
	(5487.66)	(5592.25)	(6222.38)	(6476.74)
Household expenditure (per capita)	2669.33	2724.93	3260.96	3445.89
	(4661.16)	(4778.12)	(5257.73)	(5486.15)
Labour market status				
Unemployed	0.29	0.28	-	-
	(0.45)	(0.45)		
Employment	0.71	0.72	-	-
	(0.45)	(0.45)		
Sample Size (n)	**7610**	**7223**	**5248**	**4493**

Notes:

1. The Indian category refers to both Indians and Asians.
2. Own calculations using post-stratified weights.
3. The full and restricted samples were compared (on every specified characteristic) using the Two- Sample t-Test statistic. This test statistics tests whether the two samples are likely to be from the same population.

Section 5: Results

5.1 Inequality of Opportunities in Accessing Employment

This analysis uses the dissimilarity index framework proposed by Barros et al. (2009), to quantify Inequality of Opportunities in accessing employment in South Africa. Although the dissimilarity index is calculated using coefficients from a logistic regression (defined by equation 8) of the dummy variable employment (equal to 0 if unemployed and 1 if employed) on a set of observed circumstances, Table 5 reports the odds ratio estimates from the logistic regression. The odds ratios were selected to be reported over the coefficient estimates due to their ease of interpretation as the relative probability of achieving employment for individuals with different characteristics.

$$Employment = \beta_0 + \beta_1 Female + \beta_2 Coloured + \beta_3 Indian + \beta_4 White + \beta_{5-9} HighestParentalEducation + \varepsilon_i \quad (8)$$

The results presented in **Table 5** indicate that the observed circumstance variables (the gender dummy variable, the racial dummy variables and the highest parental education dummy variables) are jointly significant (at the 1 percent significance level) in determining likelihood of employment. Focusing on each circumstance variable individually reveals some interesting results.

A look at the racial dummy variables shows that Africans are less likely to be employed than White, Coloured and Indian individuals all things equal. The race differentials for access to employment are also found to be highly significant (at the 1 percent level), with the differential between Africans and Whites being more severe than that found between Africans and the other designated racial groups. This is both expected and in line with previous findings, since it confirms that unemployment is concentrated amongst Africans (Magruder, 2010).

There are a number of channels through which differences in race can generate access to employment gaps. The first is employer discrimination against Africans, so that non-African workers tend to be preferred and are therefore more likely to be recruited by employers (Kingdon and Knight, 2004). The second is a remnant of Apartheid era segregation policies, which means that Africans tend to be located in areas with relatively high rates of unemployment (Kingdon and Knight, 2004). This spatial segregation hampers their ability to access employment relative to the other racial groups (Kingdon and Knight, 2004). Finally, the segregated educational system promoted under the Apartheid regime prevented Africans from achieving the same level of education as their White counterparts (Mlatsheni and Rospabé, 2002). This creates inequalities in the likelihood of accessing employment, since those with lower levels of education have poorer labour market opportunities than those who have attained higher levels of education.

Table 5: The Reduced-Form Odds Ratios of Observed Circumstances on the Probability of Accessing Employment (Logit Model)

Regressor:	
Gender	
Female	0.4003***
	(0.0403)
Race	
Coloured	1.8676***
	(0.2990)
Indian	3.6635***
	(1.7602)
White	4.3535***
	(1.5029)
Highest Parental Education	
Primary incomplete	0.7859**
	(0.0756)
Primary complete	0.7378*
	(0.1196)
Secondary incomplete	0.7369***
	(0.0723)
Matric	1.1093
	(0.2062)
Higher education	1.0834
	(0.3251)
Constant	4.0846***
	(0.4340)
Prob > F	0.0000
Sample Size (n)	7223

Notes:

1. Although the dissimilarity index is estimated using estimates of coefficient from the logistic regression, it is the odds ratio estimates that have been presented in this table. This is due to their having a more meaningful interpretation.
2. The omitted categories are: African, Female and No education. The Indian category refers to both Indians and Asians.
3. Standard errors are in parentheses. ***Significant at 1%, ** significant at 5% and * significant at 10%.
4. Own calculations using post-stratified weights.

Turning to the gender variable, males are shown to be in a better position to access employment than females all else equal. The odds ratio estimated for the female dummy variable is highly significant (at the

1 percent level), thus signalling the importance of gender in determining the access to employment. This finding concurs with what is already known about the South African labour market, in that the incidence of unemployment is higher for females than males (Mlatsheni and Rospabé, 2002; Ranchhod, 2009). A possible explanation for this overrepresentation of females amongst the unemployed is offered by Bhorat and Oosthuizen (2005), who attribute this finding to a mismatch between the supply and demand of female workers, with females becoming economically active at a faster rate than they are being employed. In addition, as the primary care givers for children residing within the households, females are often not able to dedicate the same amount of time to the job searching process as males (Cichello, Leibbrandt and Woolard, 2012). This can be a major constraint to their search for employment opportunities and can lead to females being less likely to gain employment. Another possible explanation for why females are less likely to access employment compared to males, is that females are discriminated against in the labour market, so that males are more likely to be employed despite having their having the same characteristics (e.g. attained the same level of education or experience) (Bhorat and Oosthuizen, 2005). Kingdon and Knight (2004) also proposed that females face geographical constraints in their search for employment opportunities because they are disproportionally crowded in the rural areas, which are characterised by relatively high rates of unemployment.

The parental education odds ratio estimates show that having a parent with some level of education does not give an outright advantage in likelihood of employment, ceteris paribus. Only individuals that reported matric or higher education as being the highest parental education are more likely to be employed compared to those whose parents had no schooling (although these differentials were not significant at even the 10 percent level). In addition, those who reported that their highest parental education level was no education were predicted to have a higher likelihood of being employed compared to those reporting less than Matric ceteris paribus. These differentials are also significant at the 1 percent level (secondary incomplete), 5 percent level (primary incomplete) and 10 percent level (primary complete).

The parental education results are unexpected given the different channels through which an educated parent could affect their child's likelihood of employment. The two channels chosen to explain this are both rooted in the vast intergenerational transmission literature. Firstly, highly educated parents tend to invest more in the education of their children, and therefore have highly educated children (Nimubona and Vencatachellum, 2007). This coupled with Leibbrandt *et al.'s* (2010) finding that individuals with low levels of education are more vulnerable to unemployment in South Africa, suggests that parental education plays an important role in determining employment. Secondly, well-educated parents tend to possess better network connections because they are more likely to be employed and therefore in contact with other employed individuals (Magruder, 2010). The quality of the connections is important, because these networks can facilitate the exchange of job related information between members, which parents can use to help secure employment for their children[26]. Burns, Godlonton and Keswell (2010) recognise the positive role of networks in an individual's ability to access jobs especially in relatively low skilled jobs, given that the formal recruitment process is expensive.

The results presented in **Table 5**, only support this hypothesis for higher levels of parental education (Matric and Higher education). This may lead one to conclude that once race and gender have been controlled for having a parent with pre-Matric education does not confer an advantage to individuals in the labour market.

By replacing highest parental education with mothers and fathers highest education, it becomes clear that mothers and fathers education impact an individual's likelihood of employment differently. **Table 2A** in the appendix contains estimates from the Logistic regression defined by equation 9, which is stated below:

$$Employment = \beta_0 + \beta_1 Female + \beta_2 Coloured + \beta_3 Indian + \beta_4 White + \beta_{5-9} Mother'sHighestEducation + \beta_{10-14} Father'sHighestEducation + \varepsilon_i \quad (9)$$

Looking first at the mother's highest education circumstance, the results reveal that those with educated mothers are not more likely to be employed compared to those whose mothers have no formal education, ceteris paribus. Some of the estimated coefficients are also highly significant (at the 1 percent or 5 percent level).

Turning to the father's highest education circumstance, those with some level of father's education are consistently more likely to be employed, than those whose fathers are reported to not have had any formal education. More specifically, individuals whose father's highest education was matric had the highest probability of accessing employment in comparison to those whose father's had no education.

Therefore these estimates can be used (to some extent) as evidence for the assertion that the highest parental education variable conceals the importance of having an educated father, even though the estimated differences (with the exception of Matric), are found to be insignificant (at the 10 percent level). The estimates presented in **Table 4A** also suggest, that the concealment may be due to the fact that mother's education dominates the highest parental education composite variable[27].

[26] The quality of the network depends on the number of its members who are employed (Burns et al., 2010)

[27] In 2008, 37 percent of the population reported that both parents had attained the same level of education. For 37 percent of the population, it was the mother with the reported the highest level of education and for the remaining 26 percent it was the father.

5.1.1 Inequality of Opportunity Estimates: The Dissimilarity Index

The results presented in Table 6 correspond to the Barros *et al.* (2009) dissimilarity index framework and can be used to measure Inequality of Opportunities in accessing employment in South Africa.

The first row states the average prevalence of employment in South Africa, and it shows that the average predicted probability of accessing employment was 71.52 percent in 2008. This means that approximately 30 percent of economically active South Africans are predicted to be unemployed. These findings are consistent with previous estimates of the unemployment rate in South Africa, and are troubling given the significant contribution of labour market income to the financial well-being of an individual (Cichello *et al.*, 2012).

Table 6: The Average Prevalence of Employment and the Inequality of Employment Opportunities (Dissimilarity Index of Access to Employment)

Prevalence of Employment ($\bar{P}$) (%)	71.52
Dissimilarity Index of Employment (%)	7.70
Sample Size (n)	7223

Notes: Own calculations using post-stratified weights.

The dissimilarity index estimate of Inequality of Opportunities in accessing employment is reported in the second row of **Table 6**. The reported dissimilarity index estimate was 7.7 percent in 2008. Although there are no previous studies applying the dissimilarity index framework to equality of access to employment, these results do indicate that between-types dissimilarities in accessing employment are low in South Africa. This is due to the fact that the predicted probability of accessing employment for each type does not substantially differ from the average predicted probability of accessing employment (for the entire population). Thus in 2008, only 7.7 percent of the opportunities to access employment would have had to be reallocated from "better-off" types to "worse-off" types in order to eliminate between-type differences in employment access rates.

These estimates therefore indicate, that the contribution of circumstances to the employment access differentials, are low. This is surprising given the characteristics of South Africa's unemployed and, more specifically, given the impact gender, race and own education have on the ability of an individual to access formal employment (Kingdon and Knight, 2004). In their study into inequality and employment in South Africa, Leibbrandt *et al.* (2010) found that females experience particularly high rates of unemployment, as do Africans and Coloureds. Recalling the strong correlation between parental education and own education, their finding that unemployment is prevalent amongst those with low levels of education might in addition to the others lead one to expect double digit estimates of the dissimilarity index. This however is not what is observed in the results presented in **Table 6**.

A possible explanation for the low estimated Inequality of Opportunities in employment is that the homogeneous treatment of the different categories of employment may be concealing important type differentials in outcome. There are currently three defined types of formal employment in South Africa; regular employment, casual employment and self-employment. Occupations are also grouped into three broad categories; managerial or professional, semi-skilled and elementary. Therefore any study interested in quantifying the impact of circumstances on likelihood of employment accurately would benefit from accounting for the different types of employment and occupations. Especially given that those in formal employment and in managerial or professional occupations tend to be "better-off".

5.2. Inequality of Opportunities in Acquiring Labour Market Income

This section contains a discussion of the results from the reduced-form regression used to parametrically estimate Inequality of Opportunity, when labour market income is being used as a proxy for the wellbeing of an individual. This model was proposed by Bourguignon *et al.* (2007a) as an alternative to an estimation strategy based on the full structural model. The regression in its reduced form is appropriate for this analysis because its specification does not distinguish between the direct and the indirect effect of the observed circumstances on outcome inequality and we are interested in estimating the overall share of inequality in well-being due to circumstances. Thus the parameter estimates on C_i can be used to decompose the inequality index and estimate the overall observed opportunity share of inequality.

The first and second columns of **Table 7** show results from the estimation of equation 10 (stated below), where the log of labour market income is regressed on a set of observed circumstances for 2008.

$$Log(LabourMarketIncome) = \beta_0 + \beta_1 Female + \beta_2 Coloured + \beta_3 Indian + \beta_4 White + \beta_{5-9} HighestParental\ Education + \varepsilon_i \tag{10}$$

The reduced form model defined by equation 10 most probably suffers from endogeneity and therefore bias, due to unobserved variables contained in the residual term (ε_i) being correlated with the observed circumstance variables. Refer back to the methodology section for the full discussion of the endogeneity issues plaguing this model. Therefore, although the parameters estimated are most likely biased the sign, relative magnitude and significance of the coefficients can still be used to gauge the general effect (direct and indirect) of the observed circumstances on labour market income (Bourguignon *et al.,* 2007a).

The results presented in **Table 7** indicate that the observed circumstance variables (race, gender and highest parental education) are jointly highly significant at the 1 percent level and explain a significant portion of the variation in labour market income, 27 percent. Moving on to the individual circumstance parameter estimates.

Looking first at the race circumstance, the racial dummy variables are found for the most part to be highly significant (at the 1 percent level) in determining the labour market income of an individual. The estimated coefficients also have the expected signs, with Africans earning less than the other races all things held equal. In addition, while the income differential between Africans and Coloureds is small, the same cannot be said for the differential between Africans and Indians or the differential with Whites.

These findings are not surprising given South Africa's legacy of systematic racism, but it is interesting that almost fifteen years into democracy large racial disparities in income continue to persist. A possible explanation for the inequality is that labour market opportunities for Africans are inferior to those of the other races. This view is supported by Leibbrandt *et al.* (2010) who discovered that although African wages grew rapidly between 1993 and 2008, they continue to be lower than those of the other racial groups. The between race disparities are even more pronounced between Whites and Africans, with White workers earning 4.4 times more than African workers on average (Leibbrandt *et al.,* 2010).

There are a number of channels through which race leads to disparities in income in South Africa. The first is that racial discrimination creates barriers to entry into high paying jobs and that this leads to the over representation of Africans in low paying jobs and whites in high paying jobs. The second explanation arises from the indirect impact of race (a circumstance variable) on labour market income through own education (an effort variable), as expressed in equation 3.1. Previous studies have found that racial segregation in the schooling system and the under-funding of schools comprised predominantly of Africans has resulted in low levels of own educational attainment amongst Africans (Gradı´n, 2012). Race has also had an intergenerational impact on own educational attainment through parental education with low levels of parental education amongst Africans resulting in low levels of educational attainment for their children (Nimubona and Vencatachellum, 2007). This has led to Africans working in less-skilled occupations, where individuals earn lower wages. Another factor that may be causing the racial gap is that the ability of Africans to earn high incomes in the labour market is undermined by their being overrepresented in areas associated with low labour market opportunities to acquire higher levels of income (Gradı´n, 2012).

Table 7: Reduced- Form OLS Regressions of Observed Circumstances on Labour market income

Regressor:	
Gender	
Female	-0.4573***
	(0.0506)
Race	
Coloured	0.1813*
	(0.1014)
Asian & Indian	0.8136**
	(0.3854)
White	1.1200***
	(0.1204)
Highest Parental Education	
Primary incomplete	0.2658***
	(0.0753)
Primary complete	0.0680
	(0.1475)
Secondary incomplete	0.5027***
	(0.0721)
Matric	0.7711***
	(0.0998)
Higher education	1.1558***
	(0.1435)
Constant	7.4893***
	(0.0726)
Prob > F	0.0000
R-squared	0.2734
Sample Size (n)	4493

Notes:

1. Omitted categories are: African, Male and No education.
2. Standard errors are in parentheses. ***Significant at 1 %, ** significant at 5% and * significant at 10%.
3. Own calculations using post-stratified weights.

The gender coefficient indicates that the labour market income of males is significantly (at the 1 percent level) greater than that of females, all things equal. This is in line with what is known about the two genders disparate labour market experiences, characterised by low female participation rates and discrimination against female workers once they enter the labour market. This leads to females not being equally considered for jobs even though they have the same qualifications as their male counterparts and not earning the same amount despite being equally productive (Barros *et al.,* 2009). The overall outcome is the persistence of a large and significant male-female wage differential. This result is also likely to reflect the fact that the number of hours worked by females and their ability to work productively while on the job, is negatively affected by women being primarily responsible for child bearing and child rearing. It is argued that these two roles lead to females earning less than males because they significantly reduce the number of years that women (of working age) spend on the job, which has an adverse effect on "the "development of human capital and work ethic" (Leibbrandt *et al.*, 2013: 7).

Turning finally to parental education, it is clear that having educated parents confers an advantage to individuals. This is because there is a positive and predominantly highly significant (at the 1 percent level) income differential between individuals with some parental education and those with uneducated parents, all things being equal. It is observed that the higher the degree of education attained by a parent, the larger the income differential is compared to a situation of no education. This links to the vast intergenerational mobility literature where it has been found that economic status is often passed from parents to their children, and since highly educated parents are likely to be high wage earners it follows that their children will also be high wage earners (Solon, 1999).

5.2.1 Inequality of Opportunity Estimates: The Inequality of Economic Opportunity Index

Table 8 shows the results from the ex-ante decomposition of labour market income inequality in South Africa. This decomposition is carried out because it allows the overall labour market income inequality to be decomposed in a way that derives the portion of overall inequality that can be attributed to a set of observed circumstances. This reveals the extent to which unequal opportunities stemming from factors outside an individual's control, impact the individual's ability to acquire labour market income in adulthood. The decomposition was performed on counterfactual distributions generated using both parametric and non-parametric estimation strategies, in which the counterfactual was that all individuals with the same observed circumstances (of the same type) acquire the same income. This smoothing eliminates within-type differentials in income, so that the remaining inequality estimated from the counterfactual is due to differences in circumstances.

The Inequality of Opportunity estimates are presented as levels $(\mathbf{IOp}_A)$ and as shares $(\mathbf{IOp}_R)$ of total income inequality. The $\mathbf{IOp}_A$ is simply the overall (observed) Inequality of Economic Opportunity index, whereas $\mathbf{IOp}_R$ is the ratio of overall (observed) Inequality of Economic Opportunity and the Total Inequality estimated from the actual distribution of labour market income.

Table 8: Scalar Indices of Inequality of Opportunity: Labour Market Income

Total inequality (E_0)	0.7462 (0.0455)
Non-parametric estimates	
$\mathbf{IOp}_A$	0.2225 (0.0198)
$\mathbf{IOp}_R$	0.2982
Incidence % opportunity inequality	**29.82**
Parametric estimates	
$\mathbf{IOp}_A$	0.2603 (0.0204)
$\mathbf{IOp}_R$	0.3488
Incidence % opportunity inequality	**34.88**
Sample Size (n)	4493

Notes: Own calculations using post-stratified weights.

The first row of **Table 8** presents the mean log deviation estimates of total inequality in labour market income and total inequality is estimated as 0.7462 in 2008. This estimate shows that South Africa exhibits high levels of inequality, compared to other middle-income countries. Ferreira and Gignoux (2011) for example found that overall inequality ranges from 0.557 (in Peru) to 0.692 (in Brazil). The comparison above should be used as a motivation to accurately estimate "unfair" inequalities, especially in countries characterised by high levels of inequality such as South Africa, so as to better address societal concerns over unequal opportunities.

As in previous empirical studies into unequal opportunities, the discussion contained in this section will focus on the relative measure of Inequality of Opportunity. Firstly, it is clear from the results that non-parametric Inequality of Opportunity estimates are lower compared to their parametric counterparts. This is unexpected given that Bourguignon *et al.* (2007a) proposes that the parametric estimation strategy allows for a finer treatment of the circumstance variables thus enabling a more precise decomposition of inequality. This is in supposed contrast to the non-parametric approach, which is plagued by large sample variations when there are few observations made for each type. Although this does not appear to be the case in South Africa, it does not detract from the fact that the closeness of the parametric and non-parametric estimations instil confidence in the presented results.

In 2008, the share of overall inequality attributable to unequal opportunities stemming from the observed circumstances was calculated as 29.82 percent and 34.88 percent for the non-parametric and parametric estimation respectively. So that on average, over a third of total inequality is completely explained by three circumstances (race, gender and highest parental education).

It therefore appears that collectively three circumstances play a major role in determining income inequality in South Africa. This is confirmed by Piraino (2012), who estimated that 25.04 percent of the inequalities in the distribution of gross income in South Africa is determined by circumstances. This Inequality of Opportunity share is lower than those estimated in this study, which is understandable given that Piraino (2012) only observed two circumstances (race and father's education) in that particular study.

The discrepancy between the Inequality of Opportunity share estimated by Piraino (2012) and those estimated in this paper is evidence that Inequality of Opportunity shares calculated from a set of observed circumstances are lower bound estimates of the "true" Inequality of Opportunity share. This is because the inclusion of additional circumstance (gender in the case of this analysis) increases the estimated Inequality of Opportunity shares. It is therefore expected that if relevant circumstances currently not observed in the income regression defined by equation 10 were included, the already high estimated Inequality of Opportunity share would be even greater.

The next section explores Inequality of Opportunities in accessing employment and in the acquisition of labour market income, when mother's education, father's education and father died before the individual was fifteen years old are included in the model specifications observed as circumstances. These alterations to the initial specification were necessary because although these circumstances are relevant to the study of unequal opportunities they are plagued by missing and invalid responses. This leads to data insufficiency problems that adversely affect the accuracy of the Inequality of Opportunity estimates, since only individuals with non-missing entries on the set of observed circumstances can be included in the sample used to explore Inequality of Opportunity in South Africa.

Section 6 therefore addresses three questions. The first regards whether the highest parental education (which does not indicate whether it is the mother or father with the highest education level) fully captures the impact of mother's and father's education on an individual's opportunities and therefore their probability of gaining employment or earning labour market income. The second regards whether the observed circumstances contributes the most to the formation of unequal opportunities in South Africa, and the third regards whether there are circumstances unobserved in the initial measurement of Inequality of Opportunities that generate unequal opportunities. Thus even though there are other circumstances which may create unequal opportunities their exploration goes beyond the limits of this paper.

Section 6: Inequality of Opportunity and the Observed Circumstances

There are six distinct specifications of circumstances in the analysis contained in this section (see **Table 9**). All of the specifications include gender and race as observed circumstances. These two variables were selected as regressors in the baseline specification (Column 1), for two reasons. The first is that they can cause outcomes to be distributed unequally if *differences in social treatment* lead to certain groups being discriminated against, in ways which generate differentials in the opportunities that determine an individual's ability to firstly access employment, and then acquire certain levels of labour market income (Barros *et al.,* 2009). The second, is that the race and gender variables were the only circumstances with negligible (or no) missing responses in the NIDS dataset.

The fifth specification (Column 5) introduces father died before the individual was fifteen years old as an explanatory variable. This circumstance was included because parental loss can create differences in the socio-economic conditions an individual is raised in, which go on to generate unequal opportunities through various channels (Ardington and Leibbrandt, 2009). One channel is that the presence of a father in the formative years is an indicator of the socio-economic status of the household the individual was raised in, because the death of a father more often than not implies poorer socio-economic status, which lends their children relatively inferior opportunities (Ardington, 2008).

Table 9: Reduced-Form Circumstances Specifications

1	2	3	4	5	6
Gender	Gender	Gender	Gender	Gender	Gender
Race	Race	Race	Race	Race	Race
	Father's Highest Education		Father's Highest Education		Father's Highest Education
		Mother's Highest Education	Mother's Highest Education		Mother's Highest Education
				Father Died before Fifteen years-old	Father Died before Fifteen years-old

Thus, in this section, the main results from the specification and estimation of Inequality of Opportunities are presented. Special attention is paid to the contribution of each set of circumstances in explaining the unequal opportunities estimated using the dissimilarity (6.1) and inequality of economic opportunity (6.2) indices.

6.1 Inequality of Opportunities in Accessing Employment

The odds ratio estimates from the logistic regressions of the employment dummy variable (=1 if employed and =0 if unemployed) on the circumstances included in each of the specifications, are reported in **Table 5A** in the appendix**.** **Table 10** then presents the dissimilarity index estimates of inequality of access to employment opportunities, derived from the logistic regressions.

The results reported in **Table 10** show that the estimated prevalence of employment and of inequalities in accessing employment is stable across the various specifications, with the 69.6 to 71.6 percent of the population estimated to be employed and dissimilarity index (%) estimates that range between [7.47, 8.53]. The small interval give credibility to the estimates of Inequality of Opportunity in accessing employment reported in the results section. The dissimilarity index estimates stated in Row 2 also corroborate the findings discussed in Section 5.1.1, because they show that circumstances do not impact the likelihood of accessing employment to a large extent. These estimates also indicate that a significant portion of the estimated inequality is due to gender and race.

Table 10: The Average Prevalence of Employment and the Inequality of Employment Opportunities (Dissimilarity Index of Access to Employment)

	1	**2**	**3**	**4**	**5**	**6**
Prevalence of Employment ($\bar{P}$) (%)	71.24	71.75	71.60	71.85	69.61	69.75
Dissimilarity Index of Employment (%)	7.47	7.74	7.87	8.07	8.05	8.53
Sample Size (n)	7610	6474	7063	6314	5110	4224

Notes: Own calculations using post-stratified weights.

Looking at column 1, the dissimilarity index estimate is interpreted to mean that 7.47 percent of the opportunities to access employment need to be reallocated from "better-off" types to "worse-off" types if between-type opportunity inequalities in employment access rates are to be eliminated. This estimated percentage of opportunities to be distributed only increases by approximately 2 percentage points after three additional circumstances have been added, thereby confirming the significant role of gender and race in generating unequal opportunities. The apparent large contribution of gender and race to unequal employment opportunities is conceivable given that the odds ratio estimates from each logistic regressions defined by specification 1 to 6 (see **Table 5A** in appendix**)** show that the magnitude and significance of the gender and race variables do not differ very much between the specifications (remain significant at the 1 percent level). This is not surprising given the direct and indirect channels through which gender and race can generate differentials in opportunities across a group of people which then leads to unequal distributions in outcomes.

In order to assess whether father's and mother's education impact the opportunities of their children and therefore their ability to access employment to different degrees, the baseline specification is compared to the specifications which control for father's highest education (Column 2) and mother's highest education (Column 3) separately.

The results reported in Column 1 of **Table 5A** are as expected, males are more likely to access employment than females all else equal and Africans are less likely to access employment compared to the other racial groups all else equal. Gender and race are also significant determinants of employment (at the 1 percent level). Results presented in Column 2 show that having a father that has completed secondary school makes an individual more likely to access employment, compared to having a father with no schooling ceteris paribus. This advantage however is only significant (at the 5 percent level) when matric is the highest level of education attained The estimates also reveal that all else equal, individuals who reported that their father attained pre-matric schooling are slightly less likely to access employment compared to those whose fathers had no schooling, although not significantly so (at even the 10 percent level). Additionally, the gender and race estimates do not differ very much between the two specifications except for the Indian dummy variable which loses significance (5 percent level) and declines by 1 percentage point. It is therefore not surprising that the estimated dissimilarity index increased by a miniscule 0.27 percentage points (see **Table 10**).

The odds ratio estimates presented in Column 3 of **Table 5A** are derived from specification 3, where father's highest education is replaced with mother's highest education. It is clear from the estimates that the only level of mother's education to confer an advantage to individuals seeking to access employment is higher education. This is because individuals reporting that their mothers did not receive any formal schooling are more likely to be employed than those whose mothers are reported to not have post-matric qualifications, ceteris paribus. More specifically, individuals whose mothers are reported to have attained primary incomplete, primary complete and secondary incomplete were significantly (at the 1 percent level) less likely to gain employment than those whose mothers did not go to school, all things equal. The dissimilarity index estimates increased by 0.5 percentage points between Column 1 and 3 (see **Table 10**). This is a larger increase than that calculated between Column 1 and 2. Thus, it appears that differentials in mother's highest education have a greater impact on the individual's opportunities and therefore their probability of accessing employment, than differentials in father's highest education.

Finally, the last specification (Column 6) is compared to the baseline (Column 1) and the specification controlling for both parents highest education (Column 4), in order to comment on the effect of increasing the number of circumstances observed.

Column 4 of **Table 5A** in the appendix shows that individuals whose father's did not die before the age of fifteen were less likely to access employment than those whose father did die before this age, ceteris paribus. The results stated in Column 5 of **Table 10** show that the dissimilarity index increased by 0.6 percentage points between Columns 1 and 5. It is therefore clear that although the presence of a father in the formative years is a channel through which unequal opportunities can emerge, this circumstances impact is minimal compared to that of the other observed circumstances. It is nonetheless clear that the inclusion of circumstances found to explain access to employment into the reduced model, should increase the Inequality of Opportunity estimated by the dissimilarity index.

The results presented in Column 4 of **Table 10** for example, show that the estimated unequal opportunities in the model with both highest parental education variables is greater than the same estimation with only race and gender (Column 2), but smaller than the same estimation including all the circumstances (Column 6). It makes sense that as the number of circumstances observed increases so do the unequal opportunities estimates. This highlights the importance of treating all Inequality of Opportunity estimates as lower-bound estimates of "true" Inequality of Opportunity, which can only be estimated when the full set of circumstances (C*) are observed. The estimate increases as more circumstances (and therefore types) are observed, regardless of whether the included variables are significant (at even the 10 percent level). The increases in the dissimilarity index estimates in response to additional circumstances however, are significantly smaller than the increases observed for the Inequality of Opportunity in labour market income estimates.

6.2. Inequality of Opportunities in Acquiring Labour Market Income

The estimates from the OLS regressions of labour market income on the circumstances included in each of the specifications are reported in **Table 6A** in the appendix. **Table 11** then presents the Inequality of Economic Opportunity estimates, derived from these OLS regressions.

Table 11: Scalar Indices of Inequality of Opportunity: Labour Market Income

	1	**2**	**3**	**4**	**5**	**6**
Total inequality (E_0)	0.7364	0.7477	0.7413	0.7421	0.7418	0.7427
	(0.0448)	(0.0468)	(0.0457)	(0.0470)	(0.0512)	(0.0546)
Non-parametric estimates						
$\mathbf{IOp}_A$	0.1676	0.2364	0.2138	0.2766	0.1574	0.3150
	(0.0146)	(0.0199)	(0.0154)	(0.0221)	(0.0136)	(0.0223)
$\mathbf{IOp}_R$	0.2276	0.3162	0.2883	0.3727	0.2122	0.4241
Incidence % opportunity inequality	**22.76**	**31.62**	**28.83**	**37.27**	**21.22**	**42.41**
Parametric estimates						
$\mathbf{IOp}_A$	0.2181	0.2845	0.2447	0.2853	0.2211	0.2996
	(0.0185)	(0.0215)	(0.0190)	(0.0213)	(0.0204)	(0.0212)
$\mathbf{IOp}_R$	0.2961	0.3805	0.3301	0.3845	0.2980	0.4034
Incidence % opportunity inequality	**29.61**	**38.05**	**33.01**	**38.45**	**29.80**	**40.34**
Sample Size (n)	4757	4025	4385	3917	3108	2538

Notes: Own calculations using post-stratified weights.

The results reported in Row 1 of **Table 11** show that the estimated total inequality in labour market income is stable across the various specifications, with it lying in the following interval [0.7364, 0.7477]. This is in contrast to the estimated non-parametric and parametric opportunity shares (%), which have broad intervals [21.22, 42.41] and [29.61, 40.34] respectively. These estimates indicate that a significant portion of the estimated income inequality is due to the circumstances observed in the analysis.

The Inequality of Opportunity shares intervals are greater for the non-parametric estimates than they are for the parametric estimates (with the exception of Column 6). The fact that the non-parametric estimates fair worse compared to parametric estimates when the model specification is altered is not surprising, given that the non-parametric approach suffers from large sample variations when more circumstances are included, because fewer observations will be observed for each type (Bourguignon *et al.,* 2007a). This is not assisted by the decline in sample size, due to the exclusion of individuals with invalid responses to the additional circumstances observed. The overall result was a reduction in the number of observations per type. This made the estimation of income differentials between the types more imprecise, especially if one is implementing the non-parametric estimation strategy.

Looking at column 1, it is clear that the opportunity shares presented in **Table 11** also indicate that a significant portion of the estimated labour market income inequality is due to gender and race, since it shows that approximately 26 percent of the total inequality in labour market income can be attributed to gender and race. This estimated opportunity share increases to 34.84, 30.92 and 25.51 percent when father's education (Column 2), mother's education (Column 3) and father died before 15 years old (Column 4) are included separately.

By comparing across Column 1, 2 and 3 of **Table 11**, it is clear that father's education contributes more to the creation of unequal opportunities and therefore the unequal distribution of income than mother's education. This result is explained by the parameter estimates in Column 4 of **Table 6A** where both highest mother's and father's education variables are controlled for. Although these variables are on the whole highly significant in explaining labour market income, the magnitude of the highest father's education parameter estimates are consistently larger than that of highest mother's education. This indicates that all else equal, a given level of father's education has a greater impact on income (compared to the father having no formal education) than an identical level of mother's education (compared to the mother having no formal education).

Just as with quantifying unequal opportunities in accessing employment comparing across Columns 1, 4 and 6 of **Table 11** shows that as the number of observed circumstances increases, so does the estimated share of total inequality due to unequal opportunities. The results show that the inclusion of both mother's and father's education increases the opportunity shares presented in Column 1 by approximately

11.85 percentage points, and then only by approximately 3.53 percentage points once the father died before the age of fifteen years-old is controlled for. This finding supports Ferreira and Gignoux (2011) suggestion of treating the estimates opportunity shares as lower-bound estimates of the "true" opportunity share.

Conclusion

South Africa is characterised by a high rate of unemployment and high level of inequality in income. This has led to numerous studies being undertaken to identify the causes and to offer solutions to these problems. This study adds to the existing literature by exploring the extent to which unequal opportunities derived from factors beyond the individual's control generate inequalities in accessing employment and in the acquisition of labour market incomes. Therefore this paper not only adds to the understanding of income inequality but also of access to employment inequalities in South Africa. This is useful given that South Africa's high unemployment rate is often cited as the cause of the unequal distribution of income across the population (Magruder, 2010).

In order to quantify the extent to which unequal opportunities lead to the unequal distributions of outcomes, this study utilised the inequality of economic opportunity index and the dissimilarity index and constructed a framework through which Inequality of Opportunities can be measured in South Africa. A broad set of circumstances were observed in this paper. These include gender, race, the individual's mother's highest education, the individual's father's highest education and whether the individual's father died before the age of fifteen. The results show that although circumstances do create unequal opportunities in accessing employment and in acquiring labour market incomes across individuals in South Africa, they do have a more significant impact on the acquisition of labour market income once an individual gains employment than they do on the individual's likelihood of accessing employment.

It is also clear from the results that the majority of these unequal opportunities stem from gender and race circumstances. This was expected given that people of African descent and females in South Africa are significantly less likely to be employed and to earn high levels of labour market incomes, as a result of their "accumulation of past and present disadvantaged characteristics" (Gradı´n, 2012: 219). Africans for example, tend to have low levels of education, access an inferior quality education, live in poor households, reside in areas characterised by high unemployment rates and have jobs in low paying occupations (Gradı´n, 2012). Gradı´n (2012) proposes that these characteristics (particularly those regarding education) have led to Africans being relatively disadvantaged both in their probability of accessing employment and earning higher levels of income compared to the other designated races.

The results contained in this study are also in sharp contrast to those calculated in similar Inequality of Opportunities studies. This is because, despite observing more circumstance variables, the estimated Inequality of Opportunity shares in these studies are significantly lower than those estimated for South Africa. Checchi *et al.* (2010b) observed five circumstances (gender, nationality, geographical location, parental education and parental occupation) and estimated Inequality of Opportunity shares ranging from 17.24 percent (in Belgium) to 0.3 (in Cyprus). Ferreira and Gignoux (2011) also estimated comparatively lower Inequality of Opportunity shares despite observing five circumstances (gender, ethnicity, parental education, father's occupation and region of birth) in their ex-ante analyses of unequal opportunities in five Latin American countries. Although these opportunity shares were estimated on a range of outcomes of well-being (earnings, per capita household expenditure and per capita household income), it is still possible to draw from this cross country comparison that relative to other countries a significant portion of South Africa's inequality is "unfair" and should be compensated for[28].

This paper was therefore partly motivated by the need for state intervention in ensuring the equitable distribution of outcomes in society, especially through policies that compensate individuals for inequalities due to unequal opportunities. Thus, by identifying the circumstances through which unequal opportunities produce inequalities in the specified labour market outcomes, the state can design redistributive policies that target inequitable inequalities more effectively. These policies will also not distort behaviour because individuals are only compensated for inequalities caused by (observed) unequal opportunities, which means that they still have the incentive to exert higher degrees of effort in order to improve their relative well-being. Without such policies, unequal opportunities create inequality traps that are unfair and that create inefficiencies in the economy (Gaviria, 2007; Marrero and Rodríguez, 2010).

Finally, although this paper focused on the inequalities in accessing employment and in acquiring labour market income which can be explained by circumstances, there are a number of extensions that would be beneficial to researchers investigating the extent to which inequalities are self-determined in South Africa. The first would estimate Inequality of Opportunities at the household level using per capita household expenditure or per capita household income as proxies for the "well-being" of the household. This is a worthy extension for a number of reasons. The first is that it allows for an analysis into the wellbeing of households which is advantageous because individuals cohabiting in the same household often share resources. Therefore any analyses into the well-being of individuals and inequality in a population would be incomplete if labour market income was the sole focus and the only proxy for wellbeing, since it does not account for expenditures made possible by income not derived from the labour market. In South Africa the main source of non-labour earnings for a majority of the population are government grants and other government income[29]. Thus by using per capita household income or per capita household expenditure a significant portion of the South African population (those who are not employed or economically active) can be accounted for in analyses of Inequality of Opportunity (Leibbrandt, Woolard, McEwen and Koep, 2008).

[28] See Brunori *et al.* (2013) for more cross country comparisons of ex-ante Inequality of Opportunity analyses.

[29] In South Africa government grants include the child grant, disability grant, old age pension, foster care grants and care dependency. Other government incomes include UIF income for unemployed individuals and workmen's compensation (Leibbrandt *et al.,* 2008).

The second extension relates to the selection of access to employment as an outcome in a study of Inequality of Opportunity in the South African labour market. This is because the Inequality of Opportunity in accessing employment estimates reported in this paper are low, which could lead one to conclude that circumstances play a limited role in determining whether an individual gains employment. This estimate however, does not account for the fact that there are different types of formal employment and occupations in South Africa. *Leibbrandt et al.* (2010) found that individuals who are in regular employment and those employed in managerial or professional positions fair better in their potential earnings, than the alternative forms of employment and occupations[30]. Therefore, a more useful analysis would be to estimate the impact of circumstances on the individual's ability to access different types of employment and occupations. Although this analysis is beyond the scope of this paper it is possible to estimate Inequality of Opportunity in multinomial outcomes using Yalonetzky's (2013) dissimilarity index of multidimensional Inequality of Opportunity.

This study would have also benefited from the inclusion of more circumstances because this leads to the population being partitioned finely, which increases the number of types observed in the sample thereby making the analysis of unequal opportunities more realistic. This extension however requires a large sample and a richer dataset and although NIDS contains a range of circumstance variables, missing and invalid responses reduce the sample considerably which adversely affects the accuracy of the Inequality of Opportunity estimates. Nevertheless, some of the relevant variables currently not observed in our regression are father's occupation, parental income and place of birth (rural or urban). The variables relating to father's occupation and parental income are strongly related to family background, and are significant determinants of labour market opportunities (Magruder, 2010). The importance of these family background circumstance variables are also related to the larger empirical literature on intergenerational mobility, especially in the transmission of economic status across generations. Place of birth is also an important circumstance, currently not observed in this model. Kingdon and Knight (2004) found that being born in rural areas can significantly restrict the income opportunities open to an individual. It is therefore expected that the inclusion of these variables into our analysis on unequal opportunities and the impact they have on labour market income (as circumstances) would result in higher Inequality of Opportunity estimates.

Finally, the South African economy is skill intensive which means that education is the predominant channel through which circumstances indirectly impact labour market outcomes. Therefore future work ought to examine Inequality of Opportunities in the attainment of education. This particular analysis would utilise the dissimilarity index of multidimensional Inequality of Opportunities proposed by Yalonetzky (2012), to estimate the effect of circumstances on an individual's probability of achieving

[30] There are three types of employment (regular employment, casual employment and self-employment) and occupation categories (managerial or professional, semi-skilled and elementary) in South Africa.

specified educational levels. These specified education levels include categories such as: primary incomplete, primary complete, secondary incomplete, matric and higher education. The results of this empirical analysis would add to our understanding of the direct impact of circumstances on a variable acknowledged to cause large inequalities both in the probability of getting jobs and the level of income one can earn, between individuals who have attained different levels of education.

Bibliography

Alesina, A. and Angeletos, G. (2005). Fairness and Redistribution. *American Economic Review*, 95(4): 960–980.

Ardington, C. (2008) Orphanhood and Schooling in South Africa: Trends in the vulnerability of orphans between 1993 and 2005. A Southern Africa Labour and Development Research Unit Working Paper Number 16. Cape Town: SALDRU, University of Cape Town

Ardington C. and Leibbrandt M. (2009) Parental Loss and Schooling: Evidence from Metropolitan Cape Town. A Southern Africa Labour and Development Research Unit Working Paper Number 42. Cape Town: SALDRU, University of Cape Town

Argent, J. (2009). Household Income: Report on NIDS Wave 1, *National Income Dynamics Technical Paper*, No. 3. *University of Cape Town: Southern Africa Labour and Development Research Unit.*
Available at: http://www.nids.uct.ac.za/documents/technical-papers/110-nids-technical-paper-no3/file

Arneson, R. (1989). Equality and Equal Opportunity for Welfare. *Philosophical Studies*, 56: 77-93.

Asadullah, M. N. and Yalonetzky, G. (2012). Inequality of Educational Opportunity in India: Changes over Time and Across States. *World Development, Elsevier*, 40(6): 1151-1163.

Barros, R.P., Ferreira, F. H. G., Vega, J.R.M. and Chanduvi, J. S. (2009). Measuring Inequality of Opportunities in Latin America and the Caribbean. Washington, D.C.: The World Bank.

Belhaj-Hassine, N. (2012). Inequality of Opportunity in Egypt. *World Bank Economic Review*, World Bank Group, 26(2): 265-295.

Bhorat, H. and Oosthuizen, M. (2005). The Post-Apartheid South African Labour Market, *DPRU Working Paper* 05/93, Cape Town: Development Policy Research Unit, University of Cape Town. Available at: http://www.dpru.uct.ac.za/sites/default/files/sites/default/files/DPRU%20WP05-093.pdf

Bourguignon, F., Ferreira, F. H. G. and Menéndez, M. (2007a). Inequality of Opportunity in Brazil. *Review of Income and Wealth*, 53: 585–618.

Bourguignon, F., Ferreira, F. H. G. and Walton, M. (2007b). Equity, Efficiency and Inequality Traps: A research agenda. *Journal of Economic Inequality,* 5: 235-256.

Breen, R. and Jonsson, J. O. (2005). Inequality of Opportunity in Comparative Perspective: Recent Research on Educational and Social Mobility. *Annual Review of Sociology*, 31:223–243.

Brunori, P., Ferreira, F. H. G and Peragine, V. (2013). Inequality of Opportunity, Income Inequality and Economic Mobility: Some International Comparisons. Policy Research Working Paper 6304, World Bank, Development Research Group.

Burns, J., Godlonton, S. and Keswell, M. (2010). Social networks, employment and worker discouragement: Evidence from South Africa. *Labour Economics*, *17*(2), 336-344.

Cameron, C.A. and Trivedi, P. K. (2010). *Microeconometrics using Stata: Revised Edition*. Texas 77845 USA: Stata Press.

Checchi, D. and Peragine, V. (2010).Inequality of Opportunity in Italy. *Journal of Economic Inequality*, 8(4): 429-450.

Checchi, D, Peragine, V and Serlenga, L. (2010). Fair and unfair income inequalities in Europe. *IZA Discussion Paper 5025. Institute for the Study of Labor, Bonn, Germany.*

Cichello, P., Leibbrandt, M., and Woolard, I. (2012). Labour Market: Analysis of the NIDS Wave 1 and 2 Datasets. No. 78. Southern Africa Labour and Development Research Unit, University of Cape Town.

Cogneau, D. and Mesple-Somps, S. (2009). Inequality of Opportunity for Income in Five Countries of Africa. *Research on Economic Inequality*, 16: 99-128.

Dworkin, R. (1981). What is Equality? Part 2: Equality of Resources. *Philosophy & Public Affairs*, 10: 283-345.

Ferreira, F. H. G. and Walton, M. (2006).Inequality of Opportunity and Economic Development. Policy Research Working Paper 3816. Washington: The World Bank.

Ferreira, F. H. G., Gignoux, J. and *Aran*, M. (2010). Measuring Inequality of Opportunity with Imperfect Data: the Case of Turkey. *Journal of Economic Inequality*, 9(4): 651-680.

Ferreira, F. H. G. and Gignoux, J. (2011). The Measurement of Inequality of Opportunity: Theory and Application to Latin America. *Review of Income and Wealth*, 57(4):622-657.

Fleurbaey, M. (1995). Three solutions for the compensation problem. *Journal of Economic Theory*, 65: 505-521.

Fleurbaey, M and Peragine, V. (2013). Ex ante versus ex post equality of opportunity. *Economica*, 80(317): 118-130.

Foster, J.E. and Shneyerov, A.A. (2000). Path Independent Inequality Measures. Journal of Economic Theory, 91: 199-222.

Gaviria, A. (2007). Social Mobility and Preferences for Redistribution in Latin America. *Economía*, 8(1): 55-96.

Gradı´n, C. (2012). Race, Poverty and Deprivation in South Africa. *Journal of African Economies*, 22(2): 187–238.

Kingdon, G. G. and Knight, J. (2004). Race and the Incidence of Unemployment in South Africa. *Review of Development Economics*, 8: 198–222.

Leibbrandt, M., Woolard, I., McEwen, H., & Koep, C. (2010). Employment and Inequality Outcomes in South Africa. *University of Cape Town: Southern Africa Labour and Development Research Unit*.

Leibbrandt, M., Woolard, I. and de Villiers, L. (2009). Methodology: Report on NIDS Wave 1, *National Income Dynamics* Technical Paper, No. 1. *University of Cape Town: Southern Africa Labour and Development Research Unit*. Available at: http://www.nids.uct.ac.za/documents/technical-papers/108-nids-technical-paper-no1/file

Leibbrandt, M., Lilenstein, K., Shenker, C. And Woolard, I. (2013). The influence of social transfers on labour supply: A South African and international review. No. 112. *University of Cape Town: Southern Africa Labour and Development Research Unit*.

Magruder, J.R. (2010). Intergenerational Networks, Unemployment, and Persistent Inequality in South Africa. *American Economic Journal: Applied Economics*, 2(1): 62-85.

Marrero, G. A. and Rodríguez, J. G. (2010). Inequality of Opportunity and Growth. Working Paper No.

154, ECINEQ.

Marrero, G. A. and Rodríguez, J. G. (2012). Inequality of Opportunity in Europe. *Review of Income and Wealth*, 58: 597–621.

Mlatsheni, C. and S. Rospabé. (2002). *Why is Youth Unemployment so High and Unequally Spread in South Africa? School of Economics University of Cape Town*. Working Paper 02/65.

Niehues, J and Peichl, A. (2012). *Bounds of unfair inequality of opportunity: Theory and evidence for Germany and the US*. No. 3815. CESifo working paper: Public Finance.

Nimubona, A.D. and Vencatachellum, D. (2007). Intergenerational Education Mobility of Black and White South Africans. *Journal of Population Economics*, 20: 149–82.

Pellicer, M., Ranchhod, V., Sarr, M., & Wegner, E. (2011). Inequality Traps in South Africa: An overview and research agenda. No. 57. Southern Africa Labour and Development Research Unit, University of Cape Town.

Peragine, V. (2004). Ranking Income Distributions According to Equality of Opportunity. *The Journal of Economic Inequality*, 2(1): 11-30.

Piraino, P. (2012). Inequality of opportunity and intergenerational mobility in South Africa. *2nd World Bank Conference on Equity*.

Pistolesi, N. (2009). Inequality of Opportunity in the Land of Opportunities, 1968–2001. *Journal of Economic Inequality*, 7: 411–433.

Ranchhod, V. (2009). Labour Market: Analysis of the NIDS Wave 1 Dataset. NIDS Discussion Paper No. 12. Southern Africa Labour and Development Research Unit, University of Cape Town.

Roemer, J.E. (1998). *Equality of Opportunity*. Cambridge: Harvard University Press.

Roemer, J.E. (2002). Equality of opportunity: A progress report. *Social Choice and Welfare*, *19*(2):455-471.

Roemer, J. E. (2006). Economic Development as Opportunity Equalization. Yale University, Cowles Foundation for Research in Economics.

Roemer, J. and Trannoy, A. (2013). Equality of opportunity. Cowles Foundation Discussion Paper No. 1921, forthcoming in Handbook of Income Distribution.

Rønning, M. (2011). Who benefits from homework assignments? *Economics of Education Review*, 30(1): 55–64.

Singh, A. (2012). Inequality of Opportunity in Earnings and Consumption Expenditure: The Case of Indian Men. *Review of Income and Wealth*, 58: 79–106.

Silber, J. and Yalonetzky, G. (2011). Measuring Inequality in Life Chances with Ordinal Variables. *Research on Economic Inequality*, 19: 77-98.

Solon, G. (1999). Intergenerational Mobility in the Labor market. *Handbook of Labor Economics*, 3: 1761-1800.

Southern Africa Labour and Development Research Unit. National Income Dynamics Study 2012, Wave 3 [dataset]. Version 1.1. Cape Town: Southern Africa Labour and Development Research Unit [producer], 2013. Cape Town: DataFirst [distributor], 2013.

Southern Africa Labour and Development Research Unit. National Income Dynamics Study 2010-2011, Wave 2 [dataset]. Version 2.1. Cape Town: Southern Africa Labour and Development Research Unit [producer], 2012. Cape Town: DataFirst [distributor], 2013.

Southern Africa Labour and Development Research Unit. National Income Dynamics Study 2008, Wave 1 [dataset]. Version 5.1. Cape Town: Southern Africa Labour and Development Research Unit [producer], 2012. Cape Town: DataFirst [distributor], 2013.

Wittenberg, M. (2009). Weights: Report on NIDS Wave 1, *National Income Dynamics Technical Report*, No. 2. *University of Cape Town: Southern Africa Labour and Development Research Unit*.
Available at: http://www.nids.uct.ac.za/home/index.php?/Nids-Documentation/technical-papers.html

Yalonetzky, G. (2012). A Dissimilarity Index of Multidimensional Inequality of Opportunity. *Journal of Economic Inequality*, 10(3):343-373.

Appendix

Table 1A. Full Sample: - Percentage of missing or invalid data circumstance entries (%)

	Access to Employment	Labour Market Income
Race	0%	0%
Gender	0%	0%
Highest parental education	5.09%	5.24%
Sample Size (n)	7610	5248

Notes:

1. The sample sizes under consideration are 21 to 59 year-olds who are economically active (access to employment) and employed (labour market income).
2. Invalid entries fall into three categories: Don't Know, Refused and Not Applicable. Whilst valid entries fall into the six categories specified in **Table 1**.
3. Own calculations using sample data.

Table 2A. The distribution of the Highest Parental Education circumstance (%) for the restricted sample

	Access to Employment	**Labour Market Income**
No education	28.11	28.39
Primary incomplete	19.67	18.36
Primary complete	6.09	5.56
Secondary incomplete	30.23	29.23
Matric	12.77	14.78
Higher education	3.13	3.68
Sample Size (n)	7223	4973

Notes:

1. The sample sizes under consideration are 21 to 59 year-olds with non-missing circumstances who are economically active (access to employment) and employed (labour market income).
2. Own calculations using post-stratified weights..

Table 4A: The reduced-form odds ratios of observed circumstances on the probability of accessing employment (Logit models)

Regressor:		
Gender		
	Female	0.3835 ***
		(0 .0420)
Race		
	Coloured	2.0766 ***
		(0.4103)
	Asian & Indian	2.3340*
		(1.1747)
	White	3.0036***
		(1.1602)
Mother's highest education		
	Primary incomplete	0.6764***
		(0.0810)
	Primary complete	0.6117**
		(0.1226)
	Secondary incomplete	0.6320***
		(0.0890)
	Matric	0.8940
		(0.2353)
	Higher education	0.9359
		(0.5314)
Father's highest education		
	Primary incomplete	1.1482
		(0.1367)
	Primary complete	1.0320
		(0.2160)
	Secondary incomplete	1.1796
		(0.1763)
	Matric	2.6401***
		(0.7054)
	Higher education	1.8530
		(1.0187)
Constant		4.5142***
		(0.5024)
Prob > F		0.0000
Sample size		6097

Notes:

1. Although the dissimilarity index is estimated using estimates of coefficient from the logistic regression, it is the odds ratio estimates that have been presented in this table. This is due to their having a more meaningful interpretation.
2. The omitted categories are: African, Female and No education.
3. Standard errors are in parentheses. ***Significant at 1 %, ** significant at 5% and * significant at 10%.
4. Own calculations using post-stratified weights.

Table 5A: The reduced-form odds ratios of observed circumstances on the probability of accessing employment (Logit models)

	1	2	3	4	5	6
Regressor:						
Gender						
Female	0.4018***	0.3923***	0.3868***	0.3793***	0.3825***	0.3756***
	(0.0384)	(0.0418)	(0.0393)	(0.0411)	(0.0411)	(0.0434)
Race						
Coloured	1.6987***	1.9101***	2.0459***	2.1186***	1.8308***	1.9086***
	(0.3014)	(0.3217)	(0.3606)	(0.3918)	(0.3303)	(0.3596)
Indian	3.8417***	2.8946**	3.9001***	2.9258**	3.3676**	2.4903*
	(1.7144)	(1.4602)	(1.9613)	(1.5146)	(1.7769)	(1.3419)
White	4.2886***	4.3159***	4.6364***	4.1392***	3.9124***	3.2654**
	(1.4074)	(1.6931)	(1.6126)	(1.6785)	(1.4900)	(1.5268)
Fathers Highest Education						
Primary incomplete		0.9025		1.0984		1.0029
		(0.0842)		(0.1270)		(0.1358)
Primary complete		0.7896		0.9825		1.0311
		(0.1474)		(0.2052)		(0.2591)
Secondary incomplete		0.9114		1.1261		1.0658
		(0.1103)		(0.1626)		(0.1707)
Matric		1.6147**		2.0490***		1.9035**
		(0.3291)		(0.5358)		(0.5454)
Higher education		1.1348		1.3287		1.2269
		(0.4540)		(0.6483)		(0.7231)
Mothers Highest Education						
Primary incomplete			0.7277***	0.6790***		0.7017**
			(0.0779)	(0.0811)		(0.1040)
Primary complete			0.6870**	0.5870***		0.5514**
			(0.1227)	(0.1186)		(0.1383)
Secondary incomplete			0.6987***	0.6408***		0.6590***
			(0.0735)	(0.0890)		(0.1014)
Matric			0.9576	0.7798		0.9214
			(0.2189)	(0.1925)		(0.2597)
Higher education			1.4096	0.9344		1.2463
			(0.5823)	(0.4597)		(0.7675)
Father Died before Fifteen years-old						
No					0.9046	0.9193
					(0.1116)	(0.1203)
Constant	3.4583***	3.6413***	4.3208***	4.2910***	3.6178***	4.3108***
	(0.3033)	(0.3875)	(0.4653)	(0.4733)	(0.5077)	(0.6730)
Prob>F	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Sample Size (n)	7610	6474	7063	6314	5110	4224

Notes:

1. The omitted categories are: African, Female, No education and Yes. The Indian category refers to both Indians and Asians.
2. Standard errors are in parentheses. ***Significant at 1 %, ** significant at 5% and * significant at 10%.
3. Own calculations using post-stratified weights.

Table 6A: The Reduced- Form OLS Regressions of Observed Circumstances on Labour market income

	1	2	3	4	5	6
Regressor:						
Gender						
Female	-0.4444***	-0.4427***	-0.4547***	-0.4449***	-0.4443***	-0.4521***
	(0.0479)	(0.0513)	(0.0517)	(0.0515)	(0.0622)	(0.0659)
Race						
Coloured	0.2984***	0.1384	0.1945*	0.1218	0.1476	-0.0923
	(0.1157)	(0.1039)	(0.1099)	(0.1064)	(0.1236)	(0.1081)
Indian	1.0393***	1.0526***	0.8700**	1.0073***	1.1026**	1.0127***
	(0.3998)	(0.2617)	(0.4129)	(0.2641)	(0.5122)	(0.3003)
White	1.5608***	1.0984***	1.1770***	1.0157***	1.5496***	0.8882***
	(0.1303)	(0.1242)	(0.1272)	(0.1191)	(0.1240)	(0.1224)
Fathers Highest Education						
Primary incomplete		0.3101***		0.2000**		0.3172**
		(0.0757)		(0.0822)		(0.1413)
Primary complete		0.2269*		0.1218		0.2837
		(0.1329)		(0.1358)		(0.1882)
Secondary incomplete		0.5774***		0.4349***		0.5975***
		(0.0745)		(0.0779)		(0.1076)
Matric		0.7852***		0.5913***		0.9090***
		(0.1040)		(0.1169)		(0.1442)
Higher education		1.2403***		1.0396***		0.9916***
		(0.1779)		(0.2127)		(0.2214)
Mothers Highest Education						
Primary incomplete			0.3032***	0.1836***		0.1756**
			(0.0586)	(0.0636)		(0.0851)
Primary complete			0.1104	0.0778		-0.0381
			(0.1492)	(0.1137)		(0.1474)
Secondary incomplete			0.4933***	0.2350***		0.1369
			(0.0783)	(0.0830)		(0.1111)
Matric			0.6939***	0.3465***		0.2858*
			(0.1183)	(0.1111)		(0.1480)
Higher education			0.9694***	0.3268*		0.3706*
			(0.1601)	(0.1944)		(0.2232)
Father Died before Fifteen years-old						
No					0.1686*	0.0147
					0.0951	(0.1042)
Constant	7.7651***	7.5279***	7.5492***	7.4925***	7.5910***	7.3986
	(0.0551)	(0.0667)	(0.0675)	(0.0703)	(0.1014)	(0.1224)
Prob>F	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
R-squared	0.2217	0.3025	0.2592	0.3077	0.2214	0.3339
Sample Size (n)	4757	4025	4385	3917	3108	2538

Notes:

1. Omitted categories are: African, Male, No education and Yes. The Indian category refers to both Indians and Asians.
2. Standard errors are in parentheses. ***Significant at 1 %, ** significant at 5% and * significant at 10%.
3. Own calculations using post-stratified weights.