Economics Master's Thesis
University of Cape Town
February 2011

Exploring the Inequality of Opportunities in South Africa

Megan Little

Supervisor: Murray Leibbrandt

Abstract: This paper constructs an Inequality of Opportunities framework for South Africa and seeks to establish whether earnings inequality therein has been driven by differing circumstances or variations in efforts. The analysis has been performed on a 2008 cross-sectional South African dataset, proxying for circumstances using years of parental education and for efforts using an individual's own education in years. The results reveal that parental education affects earnings through two channels – both directly and indirectly (through its effect on efforts which in turn impact on earnings). The direct effect appears more important for females than males, whilst the indirect effect is extremely important for both genders. Further investigation in the form of an intergenerational education mobility analysis reveals a strong link between one's parent's education and one's own education. These results suggest that making a child's education less dependent on their parent' should help to lower earnings inequality in South Africa by making efforts and thus earnings less dependent on circumstances.

Table of Contents

Introduction

Earnings inequality – the unequal distribution of earnings across a group of people – has drawn much debate in the past century at least partly because inequality has widened globally, both within and between countries. Some academics purport that inequality is a natural condition of society, driven by the will of certain individuals to excel and impress - thereby driving their earnings upward relative to the rest. Others argue that inequality forms a trap, restricting the ability of those locked into the lowest earning percentiles to emerge out of poverty, and inducing crime and instability in unequal societies. Thus, the question arises – is inequality good or bad, fair or unfair? And, if we so desire, what is the best way to reduce inequality? Which policies will prove effective and which ineffective?

These and other questions have sparked considerable academic interest and the literature on inequality has burgeoned over the past few decades. Much of the work has been dedicated to accurately estimating the level of inequality and tracking its changes over time. However it has become clear that unveiling the root cause of inequality is crucial to finding the answers to the questions above. The Inequality of Opportunities approach, a relatively new field of study, provides a particularly enlightening framework through which to do so.[1] South Africa's consistently high earnings inequality has prompted extensive study in the area, however the Inequality of Opportunities has hitherto remained largely unexplored in South Africa. This paper therefore augments the current literature by pioneering the exploration of the Inequality of Opportunities in a South African context, thus helping to identify the causes of South Africa's high inequality.

[1] John Roemer's 2000 paper, "Equality of Opportunity", pioneered the literature by contextualising the idea of circumstances and efforts in the earnings inequality framework. Other academics (most notably Bourguignon, Ferreira and Menendez in their 2003 paper, "Inequality of Outcomes and Inequality of Opportunities for Brazil") have augmented his work and applied the analysis in a developing country context.

In the Inequality of Opportunities framework one seeks to identify whether inequality has arisen because of a differential in people's efforts (E) or whether it is due to the varying circumstances (C) that they face (Bourguignon et al, 2003). An individual's characteristics are therefore partitioned into C and E, allowing one's earnings to be the joint distribution of both the efforts exerted by the individual, and the circumstances under which these efforts are exerted. The set C is herein defined as those variables over which the individual has no direct control such as one's race, gender or parents' education. Conversely, the set E contains those characteristics which the individual has the power to influence, namely own education, on-the-job training or decisions to migrate for work purposes. C and E affect one's realised earnings both directly and indirectly (as circumstances can affect efforts and thereby influence earnings).

Now, it would commonly be regarded as 'fair' if earnings inequality had arisen solely because of differences in the efforts of a group, whilst if instead their characteristics were the main cause of inequality then this would largely be considered 'unfair'. The ability to distinguish what proportion of inequality is caused by C and by E therefore furthers our understanding of the roots of inequality and helps to inform the design of well-targeted policies to reduce it.

The majority of work in the Inequality of Opportunities field has thus far been focused in the developed and higher income countries, with fairly limited exploration in the developing world (Bourguignon et al, 2003: 2).[2] As is the case with many new fields of research, academics have encountered significant problems – not only in their attempts to accurately model the Inequality of Opportunities but also in the form of substantial data limitations. The latter problem has been most acute in developing countries, which has hindered progress in the field.

[2] Recently there have been a few papers that have employed the Inequality of Opportunities framework in developing countries such as Nunez's analysis of Chile (Nunez, 2007), Singh's paper on India (Singh, 2010) and Belhaj Hassine's work in Egypt (Blehaj Hassine, 2010). Further exploration remains limited however.

South Africa however, offers a valuable opportunity for further exploration as the National Income Dynamic Study of 2008 (NIDS) not only garnered a sophisticated level of detail on individual income sources, but also captured important information on parents' education and occupation which are necessary to implement the Inequality of Opportunities' methodology. Thus, by drawing on the pioneering work of previous academics to guide this study, it has been possible to construct an Inequality of Opportunities framework for South Africa and thereby embark on the path to understanding the relative roles of circumstances and efforts in determining South Africa's high earnings inequality. The focus of this paper has been on the determinants of earnings inequality because, as will be shown, earnings have played a pivotal role in South Africa's overall income inequality.

This paper is structured as follows: Section 1 provides a brief overview of inequality in South Africa, revealing the extent of the problem and highlighting the need to address it. Section 2 outlines the methodology and explains the regression based techniques used to decompose the results. A data description is also included to describe the South African data utilized to perform the analysis. Section 3 presents a discussion of the results for the regressions and inequality decompositions thereby exposing the nuances of the South African earnings market and the various channels through which C and E can affect inequality. Section 4 briefly discusses the implications of the results for policy makers and Section 5 concludes.

Section 1: Inequality in South Africa

South Africa has consistently been plagued by high income inequality, and for a long time South Africa's GINI coefficient was amongst the highest in the world (Leibbrandt et al, 2001: 21). This is in part a legacy of apartheid (the political regime that lasted from 1948-1994) which by its very nature forced South Africans of different races to live and work separately. African, Coloured and Asian people were forced to attend inferior schools, earn lower wages than White people and live in crowded substandard regions and this bred a society with high income inequality, the legacy of which continues 14 years post apartheid.

In 1994 apartheid ended and South Africa adopted a free and fair democracy, but since then, income inequality has worsened rather than improved (Leibbrandt et al, 2010b: 4). Inequality within race groups (between members of the same race) has contributed significantly to this problem largely because inequality amongst Africans (which make up about 80% of the South African population) has widened considerably since the end of apartheid (Van der Berg et al, 2006: 7). Whilst African per capita wages have risen since the end of apartheid (thus helping to close the inter-racial income gap), 'between race group inequality' remains high by international standards (Leibbrandt et al, 2010b: 16). When high 'within race group inequality' is combined with high 'between race group inequality' the result is inevitably high total inequality.

Understanding the sources of this high inequality is of utmost importance for the South Africa government if they hope to achieve their stated goal of inequality reduction (South African Government Information, The New Growth Path). South Africa's high rate of unemployment (estimated to be approximately 25% in 2010) has played a part in driving up household level inequality, however the unequal distribution of wage incomes (amongst the employed) appears to have played the largest role (Leibbrandt et al, 2010b: 45). This paper has therefore focused on understanding the determinants of earnings and their distribution in South Africa in an effort to get to the root of what is causing South Africa's high income inequality.

Whilst there is a widespread literature dedicated to income inequality in South Africa, the Inequality of Opportunities remains largely unexplored. This paper's contribution to the existing literature is therefore an exploration of the Inequality of Opportunities in a South African context to develop a better understanding of the reasons for South Africa's high inequality. This has engendered greater understanding as to whether South Africa's inequality has been driven by differentials in people's efforts or by circumstances over which people have no control.

Section 2: Methodology

Whilst the Inequality of Opportunities is a relatively new field of study, a number of seminal works have led the way by deriving the conceptual framework and developing some insightful empirical decompositions. The Inequality of Opportunities was formally conceptualised and introduced by Roemer in 2000 (Roemer, 2000) when he proposed that variations in people's earnings could in part be explained by differentials in the efforts exerted by the individuals and in part be attributed to the varying circumstances they face. Since then, a number of other academics have extended his methodology and proposed alternative inequality decompositions. This paper has employed the techniques described by Fields (Fields, 2003) and Bourguignon et al (Bourguignon et al, 2003) to gain a deeper understanding of the significance of circumstances and efforts in determining inequality in South Africa.

Gary Fields presents a simple but insightful decomposition that uses a basic Mincerian earnings regression and its variance decomposition to identify the *relative* contributions of the right hand side variables (the determinants of earnings) to the variance of earnings (Fields, 2003). Fields' method has been adjusted for the purposes of this paper to indicate the relative importance of the sets, C and E to earnings inequality. This relative comparison has therefore illuminated whether the role of circumstances has outweighed that of efforts in determining inequality or vice versa and has proved a useful method to analyse the Inequality of Opportunities.

Bourguignon et al present another enlightening study (Bourguignon et al, 2003), that examines inequality in Brazil and through a series of regression based analyses (based on Mincer's earnings regression) and counter-factual simulations helps to ascertain whether inequality has been driven by variations in efforts or circumstances. As will be shown, the manner in which the models are constructed enables the reader to discern the direct effect of one's circumstances on earnings from the indirect effect (where circumstances affect earnings through their impact on efforts).

Drawing on the work of these and other academics it has been possible develop a comprehensive framework to uncover the roles circumstances and efforts have played in determining earnings inequality in South Africa. Below is a full explanation of the Fields and Bourguignon et al methodology accompanied by a description of the NIDS dataset that has been used to perform the analysis.

2.1. Mincerian earnings regressions

Given that labour market dynamics (and specifically variations in the earnings of those employed) are so crucial to inequality in South Africa, the methodology used in this paper revolves solely around earnings and their determinants, excluding non-labour sources of income such as government grants from the analysis. As such, central to the theory is the Mincerian earnings regression which provides the theoretical backing for estimating earnings, allowing the log of earnings (the dependent variable) to be a function of characteristics such as one's race, gender and education amongst others (Willis, 1986: 526). For the purposes of this study, these characteristics have been divided into the sets C, E and X, so that the Mincerian earnings equation in its most simple form is represented by equation 1 below.

$$Ln\,(w_i) = \alpha C_i + \beta E_i + \gamma X_i + u_i \tag{1}$$

Above, w_i represents current individual monthly earnings; α, β and γ are the three coefficient vectors pertaining to C_i, E_i and X_i respectively; C_i and E_i are the sets of circumstance and effort variables and X_i is the set of variables that are traditionally included in the Mincerian wage equation but do not fit conceptually into the sets C_i or E_i. In this paper the C_i variables are defined as individual's race, gender, parental education and parental occupation (all things over which the individual has no control). E_i on the other hand includes an education variable and a dummy indicating whether the individual has ever migrated (interpreted as an E variable because it indicates whether an individual is willing to move for work). X_i variables are defined as age, current location and occupation. Finally, u_i is the residual term, capturing unobserved C_i and E_i (this would include measurement error, random luck and any characteristics that the survey fails to adequately capture and have thus been omitted from the model) (Bourguignon et al, 2003: 5).

By using ordinary least squares to regress equation 1, it is implicitly assumed that, (a) any variables included in the sets, C_i and E_i are independent of each other and that, (b) u_i (unobserved earning determinants) have zero mean and are i.i.d (identically and independently distributed across individuals). It should be clear however that this may not necessarily be the case.

Examining assumption (a) first, it is likely that some form of interdependence exists between the variables with C directly influencing E in some cases.[3] For example, one's own education is likely to be a function of one's parents' education given that a better educated parent will be more likely to provide books, quality homework time and the on-hand assistance a child needs when learning. In addition, many children will follow the example set by their parents, having observed the returns to education that their parents have enjoyed (Bourguignon et al, 2003: 6). This interplay may hold for all C and E variables. Thus C and E may not be orthogonal to each other but rather display some form of interdependence.

[3] Note that the variables contained in the set C are explicitly assumed to be exogenously determined implying that E can not have any impact on C.

This suggests that a number of auxiliary regressions need to be run, regressing efforts on circumstances as follows:

$$E_i = \delta C_i + \epsilon_i \qquad (2)$$

Above, ϵ_i represents a vector of unobserved effort determinants, which are assumed to be i.i.d. across individuals and have zero mean. Substituting (2) into (1) yields the Complete Effect equation[4]:

$$Ln\,(w_i) = (\alpha + \beta\delta)C_i + \beta\epsilon_i + \gamma X_i + u_i \qquad (3)$$

Once again u_i is assumed i.i.d. with zero mean, and as equation 3 accounts for some of the endogeneity in the model this improves the plausibility of this assumption holding. It remains likely however, that some correlation exists between u_i and ϵ_i which brings us back to assumption (b). It seems unlikely, in equation 1, that no correlation exists between the observed earnings determinants (E_i) and the unobserved earnings determinants (u_i) as it is almost impossible to completely capture all the C and E variables given the data limitations and extensive array of characteristics that should be contained in the sets. Indeed this argument holds for equation 3 too, as it is likely that u_i (unobserved earnings determinants) and ϵ_i (unobserved effort determinants) may be correlated. The data limitations mean that important information may have been included in the error term which may therefore be correlated with E_i. This may bias the estimation of accurate coefficients in our regressions (Murray, 2006:112).

[4] 'Complete effect' because it captures both the direct and indirect effects (whilst equation 1 captured the direct effect only).

One possible solution to this problem would be to instrument for E_i using an instrument (Z_i) that is correlated with E_i but not with w_i, such that equation 2 would instead be represented by equation 2a below (where Z_i is independent of u_i) (Murray, 2006:113).

$$E_i = \delta C_i + \theta Z_i + \epsilon_i \tag{2a}$$

Researchers have however struggled thus far to identify adequate instruments for E_i for a number of reasons. The first is that it has proved difficult to find an instrument that should not itself be included in the original regression. For instance, assuming years of education is our E_i variable, one proposed instrument could be family background (Bourguignon et al, 2003: 9). Whilst this is strongly correlated with an individual's education, it would fail as a successful instrument in an Inequality of Opportunities context given that family background constitutes an important earnings determinant in its own right. Furthermore, using an aptitude test as an instrument for years of education may overcome the first problem, but in many instances, data on such tests are simply not available. It therefore proved impossible to find a suitable instrument to use in the regression analysis for the purposes of this paper.

As such, the results herein need to be interpreted with some caution given that issues of endogeneity and omitted variable bias may be distorting the results. Improving the quality of available survey data or implementing instrumental analysis would help to increase the reliability of the results. Whilst these measures are beyond the scope of this paper, the results herein have nonetheless been informative and extensions of the paper to account for the problems stated above would help to ensure greater precision.

Returning to equations 1 and 3, a deeper examination of their specification proved insightful, revealing the importance of the direct and indirect effects. Equation 1 captures only the direct effects with C_i affecting w_i through α, and E_i affecting w_i

through β. This as afore mentioned, assumes complete independence between C_i and E_i. Equation 3 however, shows the double effect of C_i on w_i. C_i affects w_i directly through the set of coefficients, α, (for given E_i) and also indirectly through C_i's effect on E_i - which in turn affects w_i (captured by the coefficients $\beta\delta$).

Thus, by including the results from both equations 1 and 3, one is able to distinguish what part of total inequality comes from the direct effect of C on earnings, and what part is due to the indirect effect (of C acting on E and thus affecting earnings and their associated distribution). This is made possible by comparing the distribution of earnings derived using the two different equations and examining the magnitude and significance of the various coefficients in the respective regressions. In addition, an examination of the auxiliary regressions (education as a function of circumstances including parental education) will help to reveal the extent of intergenerational educational mobility in contemporary South Africa.

The exercise is useful because it has quite large consequences for policy. For example, if inequality is due largely to differentials in C which act mainly through the direct channel, then a transfer going to individuals or households with certain characteristics should effectively decrease inequality. If however, the indirect channel is more important (recall the example of parental education affecting own education and thus impacting on earnings) then a conditional transfer (conditional on the child in the household attending school for instance) or improvements in the quality of education at schools that children from poorly educated families attend may prove much more effective. As will be seen, in South Africa, parental education plays a very important role in the determination of one's own education and the ability to discern the direct from the indirect effect of parental education was of paramount importance for the results in this paper.

2.2. Inequality of opportunities breakdown

Having examined the earnings regressions and the relationship between earnings and C and E, the next step is the inequality decomposition. The Fields' decomposition decomposes the earnings regressions discussed above to reveal the percentage contribution of C and E to earnings inequality, thus allowing a comparison of their relative importance to inequality. Bourguignon et al present an alternative decomposition that manipulates the earnings regressions to isolate the effect of C, thus unveiling the channels through which C affects earnings (both direct and indirect). Below is a full explanation of both Fields' and Bourguignon et al's methods.

2.2.1. The Fields Decomposition

Fields (Fields, 2003) presents a neat decomposition that manipulates an ordinary Mincerian earnings equation to reveal the percentage contribution of the independent variables to the variance of the dependent variable (in our case, log of earnings). One is therefore able to see the relative contributions of the various independent regressors (and the unexplained error term) to earnings inequality.

Fields decomposes the equation by taking the variance of both sides of the regression. The log variance of earnings is therefore attained on the left hand side (a simple inequality measure) whilst the variance of the right hand side is manipulated such that one can estimate the contribution of each right hand side variable to the total log variance of earnings. This is presented as a proportion of the total log variance of earnings, and as such can be viewed as relative percentage contributions to inequality (relative to the other included regressors and the error term).[5] The results were found to be robust across inequality measures and as such Fields' method provided a simple yet insightful way to compare the relative contributions of the sets C and E to earnings inequality in this paper.

[5] See the Appendix for a full explanation of Fields' 2003 methodology.

2.2.2. Two-Step Inequality of Opportunity decomposition

Bourguignon et al present a second decomposition that augments that of Fields' by providing greater clarity on the channels through which C and E act when they impact on earnings inequality. Discerning what proportion of total inequality is due to C and what to E is done in a two-step procedure. First, the actual distribution of earnings is determined (from the regression in equation 3) and an inequality measure is calculated (in our case the GINI Coefficient). This is named GINI 1.

Next, to isolate the collective effect of efforts and the unobserved earnings determinants (u_i) the mean value for the set of C variables is calculated and assigned to every individual (i.e. the mean years of parents' education is assigned to everybody instead of their own personal level of parents' education). Then, using the same regression form defined by equation 3 and the same beta coefficients from the previous regression (which have been stored), a second regression is run having equalised circumstances across individuals. See equation 4 below.

$$Ln\ (w_i) = \left(\hat{\alpha} + \hat{\beta}\hat{\delta}\right)\overline{C_i} + \ \hat{\beta}\hat{\epsilon}_i + \hat{\gamma}X_i + \ \hat{u}_i \tag{4}$$

$\hat{\alpha}$, $\hat{\beta}$, $\hat{\delta}$ and $\hat{\gamma}$ are the estimated coefficients from equations 2 and 3. $\hat{\epsilon}_i$ and $\hat{u}_i$ are the predicted residuals from equation's 2 and 3 respectively. X_i are the standard X_i variables and $\overline{C_i}$ are the mean circumstance variables (the same value assigned to all individuals, i).

From this counterfactual distribution a second inequality measure is derived (GINI 2). GINI 2 captures the effect that efforts and the unobserved residual have on inequality after eliminating the effect of circumstances. So, to isolate the part that circumstances have played in determining inequality, one simply needs to compare the GINI 1 and GINI 2 as the difference in the inequality measure can only be due to differentials in the circumstances. Of course, to ensure optimal accuracy it is worthwhile minimising the importance of the residual term as far as possible given the data available.

2.2.3. Strengths and weaknesses of the models

The Bourguignon et al method has a number of strengths and limitations. One of its major strengths is that it uses a relatively simple decomposition that yields easily interpretable yet insightful results. This means that the results can be explained fairly easily to policy makers with limited economic knowledge and also provide a simple but reliable platform for further researchers to build upon. In addition, whilst expanding the set of C and E variables included will improve the accuracy of the results, it is possible to perform the analysis with fairly simple variables (provided that parental information has been captured)[6]. This means that the model can be used in developing countries that lack extensive datasets.

The weaknesses mainly revolve around difficulties in fully capturing all relevant variables as well as issues of endogeneity and the failure to accurately account for interdependence between variables. Ideologically the concepts of characteristics and efforts are quite vague and not only is there contention over which variables rightly belong in the respective sets C and E, but the survey questions themselves often fail to fully capture the information needed (for instance people may forget what level of education their parents had). This means that important information may be included in the error term, making an accurate diagnosis of the roles that C and E have had on inequality, difficult to achieve.

Furthermore, the model used by Bourguignon et al does not adequately account for the interdependence between variables. Whilst the Complete Effect model improves the results by accounting for the role that the C variables have had on the E

[6] Notably, one's circumstances and efforts are likely to include a broad range of variables and this paper has relied on a very limited number due to data limitations. Extending the model to include extra information on circumstances (such as origin of birth or quality of medical care when young) and extra information on efforts (such as on-the-job training or quality of education attained - rather than just quantity in years of education) would improve the model substantially.

variables, the interplay between the C variables themselves is largely ignored. For instance, in the inequality decomposition each individual's personal C variables are changed to the mean C variables in order to pinpoint the role of efforts. It is however likely that this change will have an effect on the other C variables. If for example every individual is assigned the mean parental education, the impact that this would have on parental occupation would not be fully captured. Developing a General Equilibrium model with a number of different representative agents may help to address this problem.

Finally, whilst this paper has focused solely on individual earnings inequality, a worthy extension would be the analysis of household level income inequality (in line with the model discussed by Bourguignon et al 2003). Individuals typically live together in households, cohabiting with friends and family, meaning that their own income and standard of living are not just a function of their personal earnings but also depend on their household's collective income and their household's structure. Thus, an employed individual living in a household with no other earners (for instance with children, non labour force participants or unemployed individuals) is likely to share his earnings with the others in the household, thereby lowering his own proportion of the earnings (Bourguignon et al 2003: 21). Broadening the inequality analysis to the household level would therefore incorporate the unemployed and non-labour force sectors of the population (a significant proportion of South Africa's population) in the inequality analysis.

Through this extension one could therefore gain deeper insight into the effect C and E have on non-labour earnings (grants, remittances etc) and their effect on the determination of household structure (including household size, number of adults and number of employed adults in the household). This would not only aid policy makers in the design of inequality reducing policies but would also test whether existing grants are reaching individuals living under adverse circumstances.

The model presented by Barros et al (Barros et al, 2010) could also be utilized to compare the relative significance of non-labour and labour income sources in determining household level inequality, thus revealing how significant non-labour income sources (such as government grants) have been in reducing inequality. Because NIDS Wave 1 gathered substantial detail on the sources of non-labour income and also on household characteristics, the requisite data to perform this extension are readily available. However, the econometric issues that confront one are fairly intractable and a thorough understanding of earnings inequality was felt to be an important first step in understanding the Inequality of Opportunities in South Africa.

Thus whilst there are a number of limitations to the models employed in this paper, the Inequality of Opportunities framework has not been applied in South Africa before and it has the potential to launch an important discussion on the role of circumstances and efforts in determining inequality in South Africa.

2.3. Data Discussion

This paper uses Wave 1 of the National Income Dynamic Study (NIDS) 2008 as a national cross-sectional data set. This wave is the first of a number of waves of a national panel study that has been commissioned by the South African Presidency to track the changes in income, expenditure, access to services and other measures of well-being for roughly 30 000 individuals in South Africa. The 2008 base wave is a representative sample for South Africa as a whole (NIDS Technical Paper No. 1, 2009: 1). The degree of detail on sources of both labour and non-labour income, household structure and parental characteristics make the NIDS dataset particularly useful in the context of Inequality of Opportunities. Whilst the second wave of the NIDS study has not yet been completed (thus the panel aspect of the NIDS datasets cannot be exploited) it has nonetheless been useful in its cross-sectional form.

NIDS employed a stratified 2-stage cluster sample design to ensure that the results were representative for South Africa as a whole (NIDS Weights Paper, 2009: 2). Firstly, the population was divided into 53 strata - by Province, Transitional Metropolitan Councils and District Councils. Within each of these strata, STATS SA had systematically selected 3000 Primary Sample Units (PSUs) using probability proportional to size for their use in prior studies such as the Labour Force Surveys. For the purposes of NIDS, 400 PSUs were randomly chosen from the 3000 available. Then, from within each PSU, 8 clusters were systematically drawn (each cluster is a non-overlapping sample of dwelling units or households) and these households were contacted and surveyed. NIDS' target population was private households from all over South Africa (including those resident in workers' hostels, convents and monasteries[7]) (NIDS Weights Paper, 2009: 2).

As this study focuses only on labour income (earnings), the sample was reduced to include only employed adults of working age that had finished schooling (ages 25-60). This meant that the total number of observations in the sub-sample under consideration was just over 5000 – a relatively small sample size. For this reason, when the results in this paper were weighted to yield nationally representative results, the weights were calculated to adjust for the probability of sampling each PSU and for the probability of surveying each specific household within the PSUs (correcting for household non-response) (NIDS Weights Paper, 2009: 2).

Notably, the results in this paper were not adjusted to take the stratification into account (i.e. to adjust the weights so that the age-sex-race marginal totals in the NIDS sample match those for South Africa as a whole). This was because the limited sample size made the calculation of the necessary standard errors impossible, as in some instances there was only a single sampling unit in a stratum. Failing to adjust for stratification will not affect the sign or magnitude of the coefficients in the regressions, but will only affect the associated standard errors. In fact, failing to adjust for stratification is likely to inflate the size of the standard errors slightly,

[7] Note that those residing in students' hostels, old age homes, hospitals, prisons and military barracks were excluded.

thereby biasing our results towards over-rejecting the null of significance (Deaton, 1997: 12). As such, our results will not be jeopardised by the failure to adjust for stratification.

Due to the wide variations in the labour experiences of men and women, the analysis was performed separately by gender (Bourguignon et al, 2003:13 & Oaxaca, 1973: 694). The dependent variable, log of earnings, has been derived as the log of the sum of all labour income per individual (recorded as Rands earned per month)[8]. Then, as afore mentioned the set C includes dummies for the individual's race and gender, parent's education (the mean years of schooling attained by both parents), the difference between father and mother's education in years of schooling (to test whether mother and father's education play a different role) and father's occupation (a dummy indicating which occupational category the father worked in[9]). The set E includes own education (in years), education squared (to allow for nonlinearities in returns to education) and a migration dummy (indicating whether the individual has moved from the area in which they were born). Finally the set X (included for completeness of the Mincerian earnings regression) includes age, age squared, a dummy for whether the individual lives in a rural or urban area as well as a dummy indicating in which occupational category the individual works (Bourguignon et al, 2003: 17).

Bourguignon et al employ an insightful technique by dividing the sample into age cohorts (grouping those born between 1940-1950, 1950-1960 etc into cohorts) before performing the analysis. This meant that Bourguignon et al's results capture the change in inequality sources over time and over generations without relying on a second dataset. The relatively small sample of employed adults in NIDS Wave 1 (roughly 5000) meant that it was impossible to examine the sample in age cohorts without drastically lowering the reliability of the results (as there would be too few

[8] Log of earnings was used to ensure a roughly normal distribution of earnings.
[9] The 9 occupation categories are: 1.Legislators, senior officials and managers, 2. Professionals, 3. Technicians and associate professionals, 4. Clerks, 5. Service workers and shop and market sales workers, 6. Skilled agricultural and fishery workers, 7. Craft and related trades and workers, 8. Plant and machinery operators and assemblers, 9. Elementary occupations. These are in line with the categories delineated International Standard Classification of Occupations (ISCO-08).

observations per cohort to allow for meaningful examination). In addition the widely used national datasets in South Africa from the early 1990s made it difficult to form a comparison over time given the dearth of information on parental characteristics. It is however likely that the cause of inequality will vary considerably over generations and by grouping everybody into one large group important detail may be lost. The nature of the panel study that NIDS Wave 1 (2008) forms a part of, will hopefully open up great possibilities in the future though.

Section 3: Results

3.1. Regression Results

Below is a full discussion of the results from the following regressions: (i) the Partial Effects model (equation 1) for males and females; (ii) the Auxiliary model (equation 2) of efforts regressed on circumstances, for males and females (iii) the Complete Effects model (equation 3) for males and females.

(i) Partial Effects model

Following the methodology of Bourguignon et al the log of earnings should be regressed on the full set of C, E and X variables as defined by equation 5 below (separately for men and women which is why gender is not included as an independent variable):

$$\begin{aligned}\ln(w_i) = {} & \gamma_1 Age_i + \gamma_2 {Age_i}^2 + \gamma_3 Urban_i + \gamma_{4-13} Occupation_i \\ & + \alpha_1 African_i + \alpha_2 Coloured_i + \alpha_3 Parent\_education_i \\ & + \alpha_4 Education\_difference_i \\ & + \alpha_{5-14} Father\_occupation_i + \beta_1 Education_i + \beta_2 {Education_i}^2 \\ & + \beta_3 Migrate_i + u_i \end{aligned} \tag{5}$$

However, before doing so, some preliminary investigation into the nature of the South Africa labour market will prove insightful given that South Africa's history of apartheid has had important consequences for the way in which the labour market operates. More specifically, a strong relationship has developed between education and occupation as will be shown through an examination of the VIF, correlation tables and regression output.

First the variance inflating factors were examined (see Table 1 in the Appendix). The variance inflating factor (VIF) is defined as $1/(1-R^2_{1,2,..n})$ where $R^2_{1,2,...n}$ is the coefficient of collinearity between variables 1 to n (n is the last explanatory variable). It inflates the variance of the estimated parameters in response to evidence of collinearity. The results suggest that high collinearity exists between age and age squared, and education and education squared (which is to be expected) but certain of the occupation categories also had high VIFs indicating possible collinearity with the other independent variables. The correlation tables confirm these findings (see Table 2 in the Appendix), indicating a strong correlation between one's occupation and one's education, both for the overall population and within each race group separately.

Finally, a multinomial logit model was run to test whether one's education (captured as education and education squared) is a significant determinant of one's occupational category (controlling for race, age and location). The results indicate that for both men and women, education and education squared are consistently jointly significant in determining one's occupational category as the results in Table 3 of the Appendix indicate,[10] suggesting that some collinearity may be present between the variables. Including occupation and education as independent variables in our main regression could therefore bias our results and so occupation was dropped, as education would capture the effect of occupation.[11]

[10] For both men and women, the joint hypothesis that the coefficients of education and education squared were equal to zero for every occupational category was tested. In both cases, education and education squared were jointly highly significant.

[11] The same was concluded for father's occupation, as parental education was included.

Therefore the model was run as defined by equation 6 (below) for males and females separately:

$$\ln(w_i) = \gamma_1 Age_i + \gamma_2 {Age_i}^2 + \gamma_3 Urban_i + \alpha_1 African_i + \alpha_2 Coloured_i + \alpha_3 Parent\_education_i + \alpha_4 Education\_difference_i + \beta_1 Education_i + \beta_2 {Education_i}^2 + \beta_3 Migrate_i + u_i \quad (6)$$

Table 4 in the Appendix reveals the regression results from equation 6. As mentioned above, this model is called the Partial Effects model as it captures only the direct effects of C_i and E_i. Below is a discussion of the results for females and males.

<u>Females</u>

The results indicate that the C variables (the racial dummy variables, mean parental education and the difference between the parents' education) are jointly highly significant in determining earnings. The E variables (own education, education squared and the migration dummy) are similarly jointly highly significant. Examining each variable coefficient individually reveals some somewhat surprising results however.

Looking first at the X variables, although the coefficients of age and age squared are of the correct sign (positive for age and negative for age squared thus capturing the concave relationship between age and earnings) they are not significant at even the 10% level in determining earnings. This is a surprising result and one that is at odds with intuition as well as previous findings in South Africa (Leibbrandt et al, 2010). Possible reasons for this could be due to omitted variable bias and the inclusion of an adequate instrument for occupation (and parental occupation) may help to overcome this.[12] Furthermore, the collinearity between parental education and own education is likely to be having distortionary effects on the regression. Whilst multicollinearity will not decrease the power of the model as a whole, it can have quite marked effects on the significance and magnitude of individual regressors and this may explain the surprising results. As will be shown in the discussion of the complete effects model, when the relationship between parental education and own education is accounted for, the model improves significantly. Finally, the urban dummy is significant and with the expected sign, indicating that those living in urban areas can expect to earn 1.55 times as much as those in rural areas all else equal[13].

Turning to the set of C variables next, it was found that ceteris paribus, Africans earn 1.67 times less than Whites, the base category (significant at the 1% level), and Coloureds earn 1.18 less than Whites (although this was not significant at even the 10% level). This is in line with the current trends in South Africa, although given the strong racial divides in South Africa it was surprising that the Coloured coefficient proves insignificant (Leibbrandt et al, 2010a).

Parents' education is significant however in determining the earnings of females (although only significant at the 10% level), and for each additional year of schooling attained by an individual's parents, the individual's earnings can be expected to increase by 3.1%, all else equal. Recalling that equation 6 has been regressed under the assumption of independence amongst the right hand side variables, this result suggests that parent's education plays a role in determining earnings independent of its effect on own education (as this is assumed to be held constant).

Intuitively, there are a number of different channels which could explain this. Firstly, networking effects could be in operation as a well-educated parent may introduce their child to the influential people needed to get a good job (Haveman, 2006: 125). In addition, a better educated parent is likely to earn more (something the model is

[12] Unfortunately no adequate instrument was found for the purposes of this paper.
[13] Note that because the dependent variable is log of earnings we have to take the antilog of the coefficient to see the partial effects (1.55= e^0.436)

not able to capture or control for as this was not recorded in the survey). Thus a well-educated parent is more likely to be able to provide the financial freedom for their child to get the best job – either supporting them through unpaid internships, paying for the transport to get them to and from work or simply by not forcing them to take the first job on offer but rather wait (financially supported) for the best option.

The model also tested for whether the difference between father and mother's education plays an important role in determining earnings – i.e. if the father is well-educated but the mother is not, will this have an effect on a child's earnings or is it sufficient that one parent has a good education (Nguyen et al, 2003:8). The results suggest that a difference between the levels of parents' education is not however significant in determining earnings for females.

Finally, with regards to the E variables, both education and education squared were included (in line with the literature on education in South Africa) to test for possible non-linearities in the returns to education. The F-test indicates that jointly, the education variables are strongly significant in determining earnings, but the individual t-tests suggest that although education is highly significant, education squared is not significant at even the 1% level. However, strong backing for non-linear returns to education has been found for South Africa thus it was decided to keep education squared in the regression (Keswell et al, 2004). The sign of the coefficient for education indicated that additional years of schooling positively affects earnings.

The migration dummy is not however significant in determining earnings. This may be due to the fact that individuals move for many reasons, not solely in their search for work opportunities – perhaps to look after a family member, or because their spouse has been transferred for their work. It could even be argued that migration is a C variable for if an individual is moved as a child this could hardly be interpreted as indicative of their efforts in searching for work. It was concluded therefore that the migration dummy is a poor proxy for an individual's efforts and it was consequently dropped from the equation. Notably, dropping this variable did not impact on the size or significance of the other variables in the model.

Males

The regression results for males once again indicated that collectively, both the C and E sets are significant in determining earnings. Focusing again on the individual coefficients, it was found that this time, both age and age squared are significant determinants of earnings all else equal (significant at either the 1% or 5% levels) with coefficients of the expected sign and magnitude. The results suggest that on average earnings peak at the age of 49½ for males, after which they decline again.[14] The urban dummy is completely insignificant however, implying that earnings in rural and urban areas do not differ significantly for males. Once again this result goes against those results found previously in South Africa (Leibbrandt et al, 2010a) indicating that perhaps some degree of endogeneity or collinearity is present and this could be distorting the coefficients.

Nevertheless, with regards to the C variables, the racial dummies are both significant at the 1% level, with African men earning 1.84 times less than White men, and Coloured men earning 1.89 times less than White men, ceteris paribus. Turning next to the parental education variables, this time mean parental education is not significant at even the 1% level, whilst the difference between mother and father's education proves significant in determining men's wages at only the 10% level. As will be shown, there may be some collinearity between own education and parental education which could be causing this distortion in the results and so discussion of these two coefficients is reserved for later.

Lastly, regarding the E variables, education and education squared are once again jointly significant, whilst education squared is insignificant at even the 10% level.

[14] This concave relationship is captured by the inclusion of the age squared variable and intuitively can be understood by the fact that after a certain age individuals either retire or work less hard as they start to enjoy the fruits of their labour. $\frac{\partial w_i}{\partial age_i} = 0.128 - (2)(0.00129) * age_i$

Nonetheless, a significant and positive relationship between years of schooling and earnings is found. The migration dummy is completely insignificant once again and as such was dropped from the male regressions too.

The results of the Partial Effects models revealed that collectively the variables contained in the sets, C and E, are jointly significant in determining earnings for both females and males. Examining each variable individually yielded some surprising results however which suggested that some issues of endogeneity may be present in the Partial Model.

(ii) Auxiliary regressions

The auxiliary regressions now provide an opportunity to address some of the endogeneity problems of the previous model – more specifically the relationship between the E variables and the C variables. Because the migration dummy was dropped from the previous regressions, the only remaining E variable is years of education, and thus the relationship between education and the other C variables was explored[15]. This analysis is akin to examining intergenerational educational mobility in South Africa. Whilst most intergenerational mobility analyses reflect on the relationship between an individual's earnings and their parents' earnings, this framework looks rather at the dependence of own education on one's parents' education.

First a correlation table was generated, testing for a relationship between own education, years of parental education and the difference between one's parents' education (see Table 5 in the appendix). The results indicate a strong correlation between own education and parent's education for both women and men (correlation coefficients of 0.6086 and 0.5788 respectively[16]) but only a very weak correlation

[15] Notably, whilst education squared can also be viewed as an E variable, because it is simply the square of education, the auxiliary analysis of education is sufficient.
[16] Note that these have been weighted to be representative for South Africa as a whole.

between one's own education and the parental education differential. This suggests that one's own education is not strongly influenced by the difference in one's parents' education.

Within racial groups a similar story was apparent, with intergenerational educational mobility lowest for Coloureds (both Coloured women and men have the highest correlation coefficients of the racial groups, 0.7423 and 0.5917 respectively). This suggests that Coloured men and women's education is strongly tied to that of their parents, making them less mobile across generations (Fields, 2000: 109). Conversely, Whites are more educationally mobile across generations with correlation coefficients of 0.3722 for White women, and 0.4024 for White men. The magnitude of intergenerational mobility for Africans is similar to that found in other countries around the world, whilst intergenerational mobility amongst Whites is high by international standards (Nimubona et al, 2007: 178 & Van Der Berg, 2006: 1). Nonetheless, even within the White category a strong relationship between own education and parent's education persists suggesting that the inclusion of parent's education and own education as right hand side variables in our partial regressions could result in multicollinearity and therefore bias our results.

Thus to gain a deeper understanding of the relationship between an individual's years of education and some of the other independent variables the following regression was run (in line with Bourguignon et al's model):

$$\begin{aligned} Education_i = {} & \gamma_1 Age_i + \gamma_2 African_i + \gamma_3 Coloured_i \\ & + \gamma_4 Parent\_education_i \\ & + \gamma_5 Education\_difference_i + \gamma_6 Urban_i + \epsilon_i \end{aligned} \tag{7}$$

The individual's education (in years attained) was regressed on years of parents' education and the difference between mother and father's education, controlling for age, the racial dummies and the urban dummy. Once again this was done separately for men and women (see Table 6 for results)

This time the results are fairly similar for men and women. In both instances the C set (race dummies and parental education variables) are jointly highly significant determinants of education. Furthermore, for both men and women, the individual variable, parent's education, is a significant determinant of their own schooling, suggesting that parents' investment in their child's human capital is similar across sons and daughters. The results suggest that for every additional year of mean parental schooling, one's own years of schooling are likely to increase by 0.46 of a year for women and by 0.39 of a year for men (all else equal). These coefficients are indicative of the extent of intergenerational education mobility. A coefficient of unity would indicate no intergenerational mobility as one extra year of parental education would translate directly into an additional year of education for the individual (Nunez, 2010: 2). Thus, because both of our coefficients are less than one, one can conclude that some degree of intergenerational educational mobility is present in South Africa which confirms the findings from the correlation tables.

In many countries, it has been found that a child's educational attainment is more closely related to their mother's education than their father's (Thomas, 2001: 332). In South Africa however, the results suggest that the difference between father and mother's education is only weakly significant (significant at only the 10% level) in determining male's education and not significant at all for females. Furthermore, the sign of the coefficient is surprising suggesting that a positive difference (i.e. that father's education is higher than mother's) has a positive effect on own education (as the sign of the coefficient is positive). In many cases, it is the mother that stays home with the child (during its school going years) and is therefore on hand to help with homework etc. It therefore does not make intuitive sense that a mother with lower education (as opposed to the same level as the father) should improve the chances of a child attaining more years of schooling. It was therefore decided that in the absence of further exploration (which goes beyond the limits of this paper), the differential between father and mother education be discounted from the analysis.

For both genders, age is significant and negative, suggesting that all else equal, older individuals can expect to have fewer years of education than those from the younger generations[17] a result confirmed by a number of other South Africa studies (Thomas, 2001: 332), (Anderson, 2001) & (McGrath, 2007)). In addition, the urban dummy was included to control for location and the results indicate that ceteris paribus, those individual's currently living in urban areas (not necessarily schooled there however) have on average more years of schooling than those in rural areas (significant at the 1% level). This result is likely to reflect the fact that individuals in urban areas are more likely to be educated than those in rural areas, as jobs in cities usually require higher standards of education than those in the country.

The results diverge somewhat for men and women for the racial dummies. For women, being African appears to have no effect on years of education attained (compared with the base category White) whilst the Coloured dummy is significant at the 1% level and indicated that all else equal, Coloureds have fewer years of education than Whites. For men, both the racial dummies are significant in determining education (at the 10% and 1% significance levels respectively), and the results suggest that ceteris paribus, Coloureds have the fewest years of education, followed by Africans and White men on average attain the most years of education. Notably, an F-test to test whether the African coefficient is significantly different to the Coloured, found there is no significant difference between Coloured and African years of education (significant at the 1% level). Thus, the regression results suggest that all else equal, Coloureds and Africans have on average fewer years of education than Whites as is expected given South Africa's past.

[17] Note that only individuals aged 25-60 were included, thus it is assumed that all individuals have finished their education.

Education by means

By examining the mean years of education within gender and race groups it is found that on average Africans and Coloureds attain between 8.11 and 8.49 years of education, whilst Whites on average attain approximately 12.8 years of education. It should be noted that the standard errors are very high within each race and gender group (ranging between 2.19 for Whites to 4.53 for Africans), implying wide variations in years of education attained within the different groups. The standard errors suggest that years of education vary most widely amongst African females. The results are presented in Table 7 in the Appendix.

A similar exercise was run for mean years of parental education, the results of which can also be seen in Table 7. This time, mean parental education varies much more starkly by race with Africans' parents on average attaining between 2.4 and 2.71 years of education and Coloureds' averaging between 4.06 and 4.17 whilst Whites' parents have on average 10.6 to 10.42 years of education. Once again the standard deviations are fairly wide (averaging at about 3.5).

The tables have three main messages. Firstly, the large variation in mean parental education across races (and within races) reveals how widely distributed circumstances are in South Africa. Secondly, it is encouraging to note that on average the number of years of education attained has increased substantially within all race groups (from the previous generation to the current one). Lastly, whilst the average years of education continues to differ substantially across the racial lines, the gap appears to be narrowing from the parent generation to the current.

Therefore the auxiliary regression analysis and examination of mean education by race suggest not only that South Africans have a widely distributed set of circumstances (proxied for by parental education) but that these circumstances are closely correlated with their efforts (i.e. education). For this reason the Complete

Effect model was run next, to capture both the direct impact of parental education on earnings as well as the indirect effect as it acts through education.

(iii) Complete Effects model

The results from the Complete Effects regressions can be seen in Table 8 of the Appendix. In contrast to the Partial Effects regressions, the residual from the auxiliary regression has been included instead of education (and similarly the residual squared instead of education squared). This in essence captures the effect of education, purged of the effects of parental education, age, race and location.

The impact this has on the results is quite marked. First, because parental education is now allowed to act through two channels (both directly and indirectly) the magnitude of the coefficient increases 3-fold for females and 6-fold for males. In addition, parental education is now highly significant (at the 1% level) for both females and males, whilst in the Partial Effects model, it is not significant at even the 10% level for males. The difference between father's and mother's education is not significant in explaining earnings in any of the racial categories except Coloured males. As such it was decided that this variable should be dropped from the regression analysis as the results suggest that it is largely not influential in determining earnings. Dropping this variable had no bearing on the size or significance of the other variables and therefore did not substantially alter the model.

Once again both education and education squared (or in this case the auxiliary residual and residual squared - education purged of the effects of parental education) are included as regressors. Whilst education squared is significant for neither females nor males in the Partial Effects models, this time both education and education squared are highly significant determinants of earnings in the Complete Effects model, suggesting that the returns to education are non-linear in South Africa and that a strong positive relationship exists between extra years of schooling and earnings (ceteris paribus).

The other coefficients (age, urban and the racial dummies) all have the expected sign and magnitude although it must be said that age remains insignificant in determining earnings for females. It is possible that some form of omitted variable bias may be causing this anomaly, which unfortunately is something that has been difficult to control for given data restrictions.

Disaggregating by race yields a very similar pattern. Education and education squared are consistently positive and significant.[18] We can once again conclude positive and non-linear returns to education within each racial category therefore. The coefficients suggest that returns to education (on earnings) are highest amongst Coloured women and men (compared with the other racial gender categories).

The results therefore reveal a number of things. Firstly, education is consistently significant in determining earnings after the effects of parental education have been purged. Furthermore education displays a non-linear relationship with earnings. Secondly, when the Partial model is compared to the Complete model (thus allowing parental education to impact on earnings both directly and indirectly) the significance of parental education in explaining earnings increases dramatically. Not only is parental education consistently significant in every racial and gender category but the magnitude of its coefficient increases markedly from the Partial model to the Complete. We can therefore conclude that the indirect channel must be of high importance in South Africa and that both one's efforts (captured by years of education) and one's circumstances (parental education) are significant determinants of one's earnings. The major question therefore is what effect these have had on inequality or the distribution of earnings in South Africa.

[18] This result was true in all but the white male category where the very small sample size of only 162 may have been distorting the results somewhat.

3.2. Inequality Decomposition

3.2.1. Fields' decomposition

The Fields' decomposition gives a first glimpse into the varying roles the C and E factors have had on inequality. By revealing the *relative* importance of C and E it helps to compare their contributions and indicate whether it is C or E that has played the dominant role in causing inequality. The decomposition was performed on both the Partial and Complete models and the results are described below and can be seen in Table 9 of the Appendix.

Examining the Partial model first, it is clear that for both genders the residual explains the largest proportion of total inequality (60% and 69% of total inequality respectively for females and males). This result is not wholly unexpected as in many cases the determinants of earnings are often either unobserved or unobservable which results in them being lumped into the residual term. As such, the residual would be expected to contribute sizably to inequality.[19]

Of the remaining variables, the only really significant contributor was education (education and education squared are viewed collectively). For females, education explains approximately 27% and for males education explains approximately 22.5% of total earnings inequality. The set of circumstance variables (race and parental education) jointly appear to explain only 8.7% for females and 5.6% for males, and individually parental education contributes only 5.9% and 1.6% for females and males respectively.

A very similar trend was apparent within the racial groups in the Partial Effects model with education consistently outstripping the other variables in its contribution to within race earnings inequality. The contribution from education varies markedly between the races however. For Coloured males it contributes a huge 62% and for White

[19] This is in fact a result that has been found by many academics that have utilized Fields' decomposition (Leibrrandt et al, 2010a).

females a similarly high value of 60%. For Coloured females and African females and males education contributes roughly 30% and for White males it accounts for only 9.4%. In every case, education (interpreted here as efforts) is the largest contributor to earnings inequality.

Circumstances (parental education) once again play a relatively minor role in determining inequality. For African males the contribution is almost negligible and although it is more significant for the other races (most notably amongst Whites) its contribution never exceeds that of own education. Thus, if one were to base an hypothesis on these results one would no doubt say that one's education plays the highest role in determining earnings inequality and that circumstances have a small (to negligible) effect.

Jumping to this conclusion would be premature however as these results are somewhat misleading. By performing the Fields' decomposition on the Complete Effects model one can see the importance of allowing parental education to act through the dual channels.

Examining overall inequality first (amongst all women and all men in South Africa) it is clear that although the residual continues to play the largest role, the contribution of education has decreased dramatically (contributing only 14.64% for females and 13.56% for males in the Complete Effects model). Conversely, parental education now plays an important role with a 16.55% contribution for females and a 10.15% contribution for males. Collectively circumstances (race and parental education) account for 21% for females and 15.5% for males, in both cases exceeding the effect of efforts (proxied for by education).

Once again, decomposing the within race inequality revealed very similar results. In every instance education's contribution falls compared with the Partial Effects model, whilst the contribution of parental education increases. For White females, White males and Coloured females parental education contributes more to inequality than own education, whilst for the remaining race categories education remains the largest contributor.[20]

The comparison presented above highlights an important characteristic of earnings inequality in South Africa. The direct contribution of parental education to earnings inequality is relatively small (as indicated by the Partial Effects model), however when one broadens the model to allow parental education to affect earnings inequality indirectly (through its impact on education), the true contribution of parental education is revealed. The Complete Effects results therefore suggest that both circumstances and efforts are important contributors to total earnings inequality in South Africa and to within race inequality.

Having examined the relative contributions of C and E through the Fields decomposition, the next step is the Bourguignon et al decomposition. The Bourguignon et al method reveals not only the importance of C and E in determining inequality but the exact contribution C has had on the GINI coefficient (an inequality indicator) in South Africa. One is therefore able to pinpoint the effect that an equalisation of circumstances is likely to have on earnings inequality both overall in South Africa and within race groups. The results are presented below.

[20] Because the relative sample sizes of the different racial gender categories are fairly small, the discussion of the within race inequality contributions has been limited in this paper. Performing the Fields' decomposition on a larger sample would help to generate further insight however.

3.2.2. Bourguignon et al decomposition

Following the Bourguignon et al decomposition, the GINI coefficient (GINI 1) for the actual distribution of earnings was derived first. Thereafter the C variables were equalised and assigned to every individual and a second GINI (GINI 2) was calculated from the associated earnings distribution. Comparing the GINI coefficients thus made it possible to discern what role circumstances have had in determining inequality, having removed the impact of efforts and the residual. Because mean levels of parental education differ so drastically across races, it is expected that assigning everyone the same level thereof should lower inequality substantially. This decomposition was run for the distribution of earnings for South Africa as a whole and also within each race group for both the Partial and Complete Effects models. The results are presented below and can be seen in Table 10 of the Appendix.

At a glance, the GINI coefficients indicate that earnings inequality is extremely high amongst both women and men across South Africa, as their respective GINI's are 0.600 and 0.627. Earnings inequality appears slightly more severe amongst men than women. Disaggregating by race reveals that the highest within race inequality is amongst African females as their GINI is 0.588 and African males are only 2.3 percentage points behind with a GINI of 0.565. Inequality is also notably high amongst Coloureds, although this time inequality is higher between Coloured males than between Coloured females. Amongst Whites, inequality is substantially lower than the other races with White females attaining a GINI of only 0.382 and males that of 0.490. This quick analysis therefore reveals that South Africa clearly has both high overall earnings inequality and high within race inequality which is most acute for Africans and Coloureds.

The decomposition of the Partial Effects model was examined next. To examine the effect on total inequality (across all females and males regardless of race) the national mean parental education was assigned to every individual in the sample and in addition the racial dummies were excluded from the regression (in essence assigning everyone the same race).

The results vary quite markedly by gender. Inequality amongst females decreases quite substantially after the equalisation of circumstances as their GINI falls by 6.4 percentage points from 0.600 to 0.536. The result for males is less dramatic however as their GINI falls by only 2 % points from 0.627 to 0.607. This suggests that circumstances (captured by one's race and parental education), acting through the direct channel, play a relatively small role in determining inequality amongst males in South Africa (given that levelling the circumstance playing field had little impact). Inequality amongst females however is significantly more responsive to the equalisation of circumstances.

Further insight was gained by looking at within race inequality. In this case, because everyone within each category was of the same race, the only circumstance variable to be equalised was parental education and everyone was assigned their race and gender group's mean parental education (naturally these differed across categories). The impact on within race inequality amongst Africans is fairly small with only a 2.4% drop in the African female GINI (from 0.588 to 0.564) and a tiny 0.3% drop for the African male GINI. The results for Whites reveal a very similar situation with both the White female and White male GINI's falling by only 1.7% and 1.6% respectively.

The within race inequality decomposition for Coloureds generates some surprising results however. Whilst the equalisation significantly reduces inequality amongst Coloured women, lowering their GINI by 7.6 percentage points, it in fact worsens inequality amongst Coloured males as their GINI increases by 2.3 percentage points. This would suggest that levelling the playing field of parental education actually worsens inequality amongst Coloured males, a result that remains hard to justify but may be linked to the small sample sizes used in the gender-race sub groups. However, recalling that parental education was completely insignificant in determining earnings amongst Coloured males (in the Partial regressions) suggests that standardising parental education may have little significant effect on inequality. Naturally, this prompted further study and the Complete Effects model was decomposed next.

The Complete Effects model, as afore mentioned allows parental education to act both directly and indirectly on earnings. Thus, equalising parental education will affect each person's own level of education (which in turn will impact on earnings and inequality) and also affect earnings in its own right. Allowing parental education to act through both channels has a dramatic effect on the inequality decomposition as the results below reveal.

Overall inequality (between all women and all men in South Africa) falls significantly after the equalisation for both genders, with the female GINI dropping 14.25% points and the male GINI dropping 14.13% points. When compared with the Partial Effects model, inequality falls by an additional 7.9 % points for females and 12.1% points for males in the Complete Effects model. This comparison illuminates the importance of the indirect channel in South Africa. It is parental education's role in determining own education, which in turn affects earnings that has elevated parental education to be a key driver of earnings and thus of inequality. This has been most noticeable for men in South Africa given that their GINI coefficient changes most dramatically between the Partial and Complete Effects models This suggests that the indirect channel is strongest amongst the men. The true importance of one's circumstances (such as parental education) therefore becomes clear only in the Complete Effects model revealing how critical quality homework time with an educated parent is, or how the financial stability of more educated parents contributes to a child's education.

Looking next at within race inequality, the results reveal that the equalisation of parental education helps to lower inequality in every racial group (this time including Coloured males). In addition, the comparison of the Partial with the Complete Effects model shows that the indirect channel is consistently important in every race and gender group. The indirect channel appears to have the largest impact on inequality amongst Africans (as the absolute decrease in the African female and male GINI's are 10.1% and 9.1% respectively when comparing the Partial with the Complete model). Significant gains are seen for Coloured males too, whose GINI falls 7.09% points compared with the Partial model.

A different scenario is apparent amongst Coloured females. Coloured females had already seen a substantial fall in their within group inequality after the equalisation of parental education in the Partial model. The opening up of the indirect channel in the Complete model appears to have little impact on inequality as their GINI falls by only an additional 1.6% points when the Partial and Complete models are compared. For Whites, whilst the GINI falls by only 3.5 and 4 absolute percentage points respectively for females and males, in relative terms (the change relative to their GINI in the Partial model) the gains are more substantial. The analysis by race therefore reveals that the indirect channel plays an important role in determining inequality in all racial categories (although the effect is relatively small amongst Coloured females).

The inequality decomposition of the Partial and Complete Effect models has helped to highlight the key importance of the direct and indirect way that parental education affects earnings. Parental education can affect one's earnings directly through networking effects or because highly educated parents are often richer and thus can support their child to get the best job. This channel seems to be most important in determining inequality amongst females in South Africa (both overall and in the within race inequality of women). Conversely, amongst men it appears to play a relatively small role and indeed the within race inequality amongst Coloured males actually responded adversely to the equalisation of parental education.

Parental education can also affect earnings indirectly as it has a strong positive correlation with an individual's own education (which in turn has a positive relationship with their earnings). The results above suggest that this channel is highly important in South Africa for both men and women. When both the direct and indirect channels were accounted for, inequality amongst men (both overall and within race groups) fell markedly after circumstances were equalised. This was true for African and White women too but inequality amongst Coloured women appeared to react relatively little to the inclusion of the indirect effects.

Section 4: Policy Implications

The results of this paper present important evidence on the nuances of South Africa's earnings and education markets and will prove useful for any policy maker seeking to decrease earnings inequality in South Africa. In every model it is clear that parental education is an important determinant of a child's education and thus is helping to drive high earnings inequality in South Africa (as earnings are closely linked to education). Policy makers therefore need to reduce the dependence of a child's education on their parents' (and on other circumstances) to ensure that every child gets a fairer shot at a good education and thereby give them a better chance in the labour market.

This suggests that a simple cash transfer from rich to poor may have limited effect on earnings inequality and any inequality reducing policy needs to be carefully designed. Furthermore, because school enrolment is already fairly high in South Africa (particularly at primary school level) a conditional cash transfer, conditional on school attendance, may also prove ineffective. Instead policy makers need to incentivise children to stay in school for longer and focus on improving the quality of the schools that children from lesser educated families attend.

Lengthening the time that a child spends in school has significant returns on their future earnings and various incentives such as student loans for high school students or grants which are school specific (to be spent on textbooks for instance) may help to keep children from poorly educated families (who are more likely to leave school before completion) in school for longer.

Secondly, South Africa has very wide variations in the quality of education in its schools. This phenomenon is of particular interest in the context of the Inequality of Opportunities given that school quality is strongly linked to the circumstances of the pupils. This is in part a legacy of apartheid which promoted world class schooling for whites and sub-standard schooling for the other races, but also reflects the inability of the post-apartheid government to successfully address the education problem. For the most part education is not free in South Africa and this means that private schools and the top government schools can charge higher fees than other schools, thereby attracting children from more wealthy families and allowing the school to hire the best quality teachers and maintain better facilities. Furthermore, many government schools offer preferential access to children who live nearby in an effort to reduce the cost and time hassle of parents. Unfortunately, because wealth is closely tied to residential location (as a consequence of the spatial segregation patterns of apartheid) this means that children from wealthy families tend to go to the same schools and a self perpetuating pattern is created with schools in relatively wealthy areas able to charge higher fees and thereby afford to provide higher levels of education quality.

Thus, improvements to schools in the poorest areas (where pupils often come from poorly educated households) will help to weaken the link between a child's circumstances and the quality of their schooling. Many of these schools would benefit significantly from government grants to maintain school facilities or the subsidising of teachers' wages. Improving the quality of education for children from poorly educated families should contribute to higher returns to education for these pupils in terms of their future earnings and also help the students to complete more years of schooling (thus increasing the likelihood of tertiary education). This will ultimately help to lower earnings inequality in South Africa by raising the earnings potential of those who previously had the weakest education levels.

Notably, before implementing any of the policy proposals stated above, additional research is recommended. As previously stated, this study would benefit substantially from the inclusion of additional proxies for circumstances and efforts. Furthermore, extensions of the model to include non-labour force participants and unemployed individuals would yield significant insight give the nature of South Africa's labour market. A number of new surveys under way in South Africa offer exciting prospects for new data which would help address some of the challenges encountered in this paper. Thus whilst this study has furthered our understanding of the Inequality of Opportunities in South Africa, a number of data and structural issues need to be overcome before any policy changes are enacted.

Section 5: Conclusion

South Africa's persistently high income inequality has generated considerable academic interest and the literature on inequality in South Africa has expanded significantly over the past two decades. This paper augments the current body of work by exploring the Inequality of Opportunities for the first time in South Africa, thereby revealing whether inequality is largely a result of variations in efforts exerted by the population or whether it is more closely related to the varying circumstances that they face. History has shown that without carefully designed redistributive policies, inequality can perpetuate unabated for years which reiterates the importance of understanding the roots of inequality before a policy is implemented.

A quick examination of parental education (a proxy for circumstances in this paper) and own education (an efforts proxy) in South Africa reveals a wide distribution of circumstances and efforts, largely as a result of apartheid and South Africa's past. By adapting the decompositions of Fields (Fields, 2003) and Bourguignon et al (Bourguignon et al, 2003) it has been possible to construct an Inequality of Opportunities framework for South Africa that not only identifies the relative importance of these circumstances and efforts in determining inequality but also illuminates the channels through which circumstances and efforts act.

Fields and Bourguignon et al's methodologies were applied to two regression models in this paper – the Partial Effects model and the Complete Effects model. In the Partial Effects model, circumstances and efforts are restricted such that they can only impact directly on earnings. Conversely, in the Complete Effects model, whilst efforts continue to impact directly, circumstances can affect earnings both directly and indirectly (through their effect on efforts). A comparison of the models therefore reveals the importance of the indirect channel of circumstances.

The results suggest that in the Partial Effects model, efforts (relative to circumstances) play the most important role in determining inequality. Equalising circumstances amongst men in South Africa has a very small effect on male inequality and whilst inequality amongst women is more responsive to the equalisation, education consistently outstrips parental education in its contribution to inequality.

This prompted an investigation of intergenerational education mobility in South Africa and it was found that an individual's own education is highly correlated with that of their parents'. Furthermore, in the Complete Effects model, when parental education was free to act both directly and indirectly on earnings, its relative importance increases dramatically with circumstances even outstripping the contribution of efforts in some instances. It is therefore clear that for both men and women in South Africa the indirect channel is highly important, and that circumstances play a significant part in determining earnings inequality.

The results of this paper suggest that if the South African government is serious about reducing earnings inequality then the indirect channel of parental education on their child's earnings needs to be fully investigated. Given that primary school attendance is already high in South Africa, policies need to focus on incentives to keep children in school for longer as well as improve the quality of schools in poorer areas. This will help to weaken the strong link between parental education and own education and a child's earnings will therefore depend less on circumstances and more on efforts.

Because variations in wages play such an important role in determining South Africa's overall inequality, this paper has focused solely on earnings and their determinants. A worthy extension would however be to apply the Inequality of Opportunities methodology outlined herein to household level inequality, thereby expanding the analysis to include the unemployed and non-labour force sectors of the population. South Africa's high rate of unemployment makes this particularly pertinent for a comprehensive understanding of inequality in South Africa.

Nonetheless, this paper has laid the foundations for understanding the Inequality of Opportunities in South Africa through its comprehensive study of earnings inequality. The examination has revealed that circumstances continue to play a significant role in earnings inequality in South Africa and most importantly that the strong link between parental education and an individual's own education is highly important in determining earnings inequality in South Africa.

Bibliography

Anderson, K., Case, A. & Lam, D., (2001), Causes and consequences of Schooling Outcomes in South Africa: Evidence from Survey Data, *Social Dynamics 27:1*, pp37-59.

Barros, R., De Carvalho, M., Franco, S. & Mendonca, R. (2010). Markets, the State, and the Dynamics of Inequality in Brazil. *Declining Inequality in Latin America,* Brookings Institution Press, Baltimore, USA.

Belhaj Hassine, N. (2010). Inequality of Opportunity in Egypt. *Working paper No. 549, Economic Research Forum.* Retrieved from: http://scholar.google.co.za/scholar?q=Inequality+of+Opportunity+in+Egypt&hl=en&as_sdt=1%2C5&as_sdtp=on (Jan 2010).

Bhorat, H., Leibbrandt, M. & Woolard, I., (2000). Understanding Contemporary Household Inequality in South Africa. *Studies in Economics and Econometrics,* vol.24, No.3, pp.31-52.

Bourguignon, F., Ferreira, F., & Menendez, M. (2003). Inequality of Outcomes and Inequality of Opportunities for Brazil, *World Bank Working Paper No. 3174.* Retrieved from: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=497242 (Nov 2010).

Deaton, A. (1997). *The Analysis of Household Surveys, A Microenometric approach to Development Policy*. World Bank Publishers, USA.

Fields, G., (2003). Accounting for income inequality and its change: a new method, with application to the distribution of earnings in the United States. *Worker Well-Being and Public Policy (Research in Labor Economics, Volume 22)* pp.1-38.

Fields, G., (2001). Ch. 6. The Meaning and Measurement of Income Mobility. *Distribution & Development: A New Look at the Developing World*, Russell Sage Foundation, USA.

Haveman, R. & Smeeding, T. (2006) The Role of Higher Education in Social Mobility. *The Future of Children,* Vol. 16, No. 2, pp. 125-150.

Hoogeveen, J.G. & Özler, B. (2006) Ch. 2. Poverty and Inequality in post-Apartheid South Africa: 1995-2000. in H Bhorat & S. M. Ravi Kanbur (eds) *Poverty and Policy in Post-Apartheid South Africa,* HSRC Press, Cape Town.

Keswell, M. & Poswell, L. (2004) Returns to Education in South Africa: A retrospective sensitivity analysis of the available evidence. *South African Journal of Economics,* Vol. 72, Issue 4, Sep 2004. pp.834-860.

Lam, D., & Leibbrandt, M. (2009) Schooling Inequality and Earnings Inequality in South Africa since The End of Apartheid. unpublished paper.

Leibbrandt, M., Levinsohn, J. & McCrary, J. (2010a). Incomes in South Africa after the Fall of Apartheid. *Journal of Globalization and Development,* Vol. 1, Issue 1, pp. 1-60.

Leibbrandt, M., Woolard, I. & Bhorat, H. (2001) Understanding contemporary household inequality in South Africa. in Bhorat, H., Leibbrandt, M., Maziya, M., Van der berg, S. & Woolard, I. (eds) *Fighting Poverty, Labour Markets and Inequality in South Africa*, UCT Press, Cape Town.

Leibbrandt, M., Woolard, I., Finn, A. & Argent, J. (2010b) Trends in South African Income Distribution and Poverty since the Fall of Apartheid. *OECD Social, Employment and Migration Working Papers, No. 101,* OECD Publishing, OECD. doi:10.1787/5kmms0t7p1ms-en.

McGrath, S., & Akoojee, S., (2007) Education and skills for development in South Africa: Reflections on the accelerated and shared growth initiative for South Africa, *International Journal of Educational Development* Vol. 27, Elsevier Ltd., pp.421–434.

Murray, M. (2006) Avoiding Invalid Instruments and Coping with Weak Instruments. *The Journal of Economic Perspectives,* Vol. 20, No 4, American Economic Association. pp. 111-132.

NIDS, National Income Dynamic Study, Wave 1, July 2009, User Document, SALDRU, University of Cape Town, Retrieved from: http://www.datafirst.uct.ac.za/catalogue3/index.php/ddibrowser/53 (Jan 2010).

NIDS, National Income Dynamic Study, Wave 1, July 2009, Technical paper no. 1, SALDRU, University of Cape Town, Retrieved from: http://www.datafirst.uct.ac.za/catalogue3/index.php/ddibrowser/53 (Jan 2010).

NIDS, National Income Dynamic Study, Wave 1, July 2009, Introduction to NIDS data, SALDRU, University of Cape Town, Retrieved from: http://www.datafirst.uct.ac.za/catalogue3/index.php/ddibrowser/53 (Jan 2010).

NIDS, National Income Dynamic Study, Wave 1, July 2009, Report on NIDS Wave 1 Technical Paper no. 2, SALDRU, University of Cape Town, Retrieved from: http://www.datafirst.uct.ac.za/catalogue3/index.php/ddibrowser/53 (Jan 2010).

Nguyen, A. & Haile, G. (2003) Intergenerational mobility in educational and occupational status: evidence from the U.S. *Lancaster University Management School Working Paper 2003/043*

Nimubona, A. & Vencatachellum, D. (2007) Intergenerational education mobility of black and white South Africans. Published online, Springer-Verlag. Retrieved from: http://www.springerlink.com/content/h07x783760774152/ (Jan 2011).

Núñez, J. & Miranda, L. (2010) Intergenerational Income Mobility in a Less-Developed, High-Inequality Context: The Case of Chile. *The B.E. Journal of Economic Analysis & Policy,* Volume 10, Issue 1, Article 33, pp.1-15. Retrieved from: http://scholar.google.co.za/scholar?hl=en&q=Intergenerational+Income+Mobility+in+a+Less-Developed%2C+High-Inequality+Context%3A+The+Case+of+Chile&btnG=Search&as_sdt=0%2C5&as_ylo=&as_vis=0 (Jan 2010).

Núñez, J. & Tartakowsky, A. (2007) Inequality of outcomes vs. inequality of opportunities in a developing country. An exploratory analysis for Chile. *Estudios de Economia, University of Chile, Department of Economics,* Vol. 34(2 Year 20), pp.185-202. Retrieved from: http://scholar.google.co.za/scholar?hl=en&q=Inequality+of+outcomes+vs.+inequality+of+opportunities+in+a+developing+country.+An+exploratory+analysis+for+Chile&as_sdt=0%2C5&as_ylo=&as_vis=0 (Jan 2010).

Oaxaca, R. (1973) Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review,* Vol. 14, No. 3 (Oct., 1973), pp. 693-709. Retrieved from: http://www.jstor.org/stable/pdfplus/2525981.pdf , (Jan 2010).

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

Singh, A. (2010) The Effect of Family Background on Individual Wages and an Examination of Inequality of Opportunity in India. *Journal of Labor Research,* Springer, vol. 31(3), pp.220-246. Retrieved from: http://econpapers.repec.org/RePEc:spr:jlabre:v:31:y:2010:i:3:p:230-246 (Jan2010) .

South African Government Information, "The New Growth Path", Retrieved from: http://www.info.gov.za/aboutgovt/programmes/new-growth-path/index.html (Jan 2010).

Van der Berg, S., Burger, R., Burger, R., Louw, M. & Yu, D. (2006), Trends in Poverty and Inequality since the Political Transition. *Development Policy Research Unit Working Paper* 06/104. Retreived from: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=982093 (Jan 2010).

Van der Berg, S., Louw, M., Yu, D. (2006) Educational attainment and intergenerational social mobility in South Africa. *Stellenbosch Economic Working Papers: 09/06* (2006). Retrieved from: http://scholar.google.co.za/scholar?hl=en&q=Educational+attainment+and+intergenerational+social+mobility+in+South+Africa&as_sdt=0%2C5&as_ylo=&as_vis=0 (Jan 2010).

Willis, R.J. (1986) Wage Determinants: A Survey and Reinterpretation of Human Capital Earnings Functions. in Ashenfeller, O. And Layard, R. (eds) *Handbook of Labor Economics, Volume 1*, Netherlands.

Appendix: Tables

Table 1: VIF tables

Females

Variable	VIF	1/VIF
Age	85.76	0.011661
Age Squared	85.14	0.011746
Education	12.88	0.077649
Education Squared	12.72	0.078619
Occupation 6	5.02	0.199176
Occupation 7	5	0.200158
Occupation 8	4.48	0.222986
Occupation 9	4.46	0.224085
Occupation 5	3.85	0.260006
Occupation 2	3.4	0.294139
African	3.15	0.317426
Parental Education	2.58	0.387236
Coloured	2.46	0.406841
Occupation 4	2.12	0.471555
Father Occupation 6	1.99	0.50276
Father Occupation 7	1.77	0.565067
Occupation 3	1.73	0.57854
Father Occupation 8	1.67	0.599407
Father Occupation 2	1.63	0.614599
Urban	1.56	0.642065
Father Occupation 5	1.52	0.656939
Father Occupation 3	1.27	0.786181
Father Occupation 4	1.26	0.795116
Parental Education difference	1.09	0.9169
Migrate	1.08	0.930002

Males

Variable	VIF	1/VIF
Age	90.73	0.011022
Age Squared	89.09	0.011224
Education	12.02	0.08316
Education Squared	11.92	0.08392
Occupation 9	8.13	0.122933
Occupation 2	6.62	0.151041
Occupation 4	5.14	0.194458
Parental Education	3.35	0.298327
Occupation 3	3.34	0.299593
African	3.32	0.301636
Occupation 5	3.27	0.305664
Occupation 6	3.05	0.328084
Occupation 1	2.32	0.431237
Coloured	2.09	0.479413
Father Occupation 6	1.82	0.550789
Father Occupation 2	1.76	0.568321
Father Occupation 7	1.72	0.582051
Father Occupation 8	1.58	0.632738
Father Occupation 5	1.5	0.667729
Occupation 8	1.41	0.707502
Father Occupation 3	1.38	0.72648
Urban	1.31	0.764715
Father Occupation 4	1.31	0.765466
Migrate	1.13	0.888739
Parental Education difference	1.1	0.911376

The 9 occupation categories are:
1.Legislators, senior officials and managers, 2. Professionals, 3. Technicians and associate professionals, 4. Clerks, 5. Service workers and shop and market sales workers, 6. Skilled agricultural and fishery workers, 7. Craft and related trades and workers, 8. Plant and machinery operators and assemblers, 9. Elementary occupations. These are in line with the categories delineated International Standard Classification of Occupations (ISCO-08)

Table 2: Correlation table

Correlation Between education and primary occupation

	Occupation1	Occupation2	Occupation3	Occupation4	Occupation5	Occupation6	Occupation7	Occupation8	Occupation9
Education	0.1298	0.3962	0.1101	0.1883	0.1151	-0.3027	-0.0347	-0.0834	-0.329

Table 3: Multinomial Logit Regressions

Adjusted Wald Test: Testing whether Education and Education Squared are jointly significant in determining one's Primary Occupation in the Multinomial Logit Regressions

	Females
F (16, 343):	15.00
P-value	0.0000

	Males
F (16, 323):	7.65
P-value	0.0000

Please contact the author for more detail on the Multinomial Logit regressions.

Table 4: Partial Regression Results for females and males

Dependent variable: Log earnings

Independent variables	**Female**	**Male**
Age	0.0851 (0.0577)	0.128*** (0.0444)
Age Squared	-0.000895 (0.000687)	-0.00129** (0.000559)
Urban	0.436*** (0.127)	0.168 (0.124)
African	-0.513*** (0.196)	-0.613*** (0.171)
Coloured	-0.168 (0.164)	-0.636*** (0.221)
Parental Education	0.0310* (0.0188)	0.0184 (0.0146)
Parent Education Difference	0.0185 (0.0227)	-0.0184* (0.0104)
Education	0.00446 (0.0445)	0.0212 (0.0346)
Education Squared	0.00825*** (0.00264)	0.00611*** (0.00171)
Migrate	0.0976 (0.107)	0.0629 (0.115)
Constant	4.408*** (1.158)	4.406*** (0.878)
Observations **R-squared**	724 0.515	813 0.492

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Base category: White

Table 5: Correlations of education with parental education

5a: Correlations for SA as a whole

Females

	Education	Parent Education	Difference between parent's education
Education	1		
Parent Education	0.6086	1	
Difference between parent's education	0.0475	0.039	1

Males

	Education	Parent Education	Difference between parent's education
Education	1		
Parent Education	0.5788	1	
Difference between parent's education	0.109	0.0998	1

5b: Correlations within races

African

Females

	Education	Parent Education	Difference between parent's education
Education	1		
Parent Education	0.493	1	
Difference between parent's education	0.0369	0.0509	1

Males

	Education	Parent Education	Difference between parent's education
Education	1		
Parent Education	0.4931	1	
Difference between parent's education	0.0261	0.0652	1

Coloured

Females

	Education	Parent Education	Difference between parent's education
Education	1		
Parent Education	0.7423	1	
Difference between parent's education	0.1069	0.0698	1

Males

	Education	Parent Education	Difference between parent's education
Education	1		
Parent Education	0.5917	1	
Difference between parent's education	0.1093	0.1398	1

White

Females

	Education	Parent Education	Difference between parent's education
Education	1		
Parent Education	0.3722	1	
Difference between parent's education	0.19	0.1903	1

Males

	Education	Parent Education	Difference between parent's education
Education	1		
Parent Education	0.4024	1	
Difference between parent's education	0.4809	0.1918	1

Table 6: Auxiliary regressions for females and males

Dependent variable: Years of education

Independent Variables	Female	Male
Age	-0.113***	-0.0913***
	(0.0157)	(0.0165)
African	-0.418	-1.359*
	(0.464)	(0.692)
Coloured	-1.567***	-2.134***
	(0.494)	(0.793)
Parental Education	0.460***	0.391***
	(0.0355)	(0.0372)
Parental education difference	0.0482	0.0887*
	(0.0321)	(0.0514)
Urban	1.186***	1.921***
	(0.340)	(0.351)
Constant	11.74***	11.05***
	(0.779)	(0.917)
Observations	1,685	1,592
R-squared	0.442	0.405

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Base category: White

Table 7: Mean Education and Mean Parental Education by race and gender

	African		Coloured		White	
	Females	**Males**	**Females**	**Males**	**Females**	**Males**
Mean Years of Education	8.112078	8.145022	8.502193	8.495127	12.87137	12.88172
Standard deviation of Years of Education	4.531443	4.372673	3.729301	3.873043	2.193753	2.947439
Mean Years of Parental Education	2.437898	2.712132	4.176375	4.064024	10.605	10.42417
Std Deviation of Parental Education	3.208219	3.391882	3.713628	3.574526	3.156946	3.445692

Table 8: Complete Effects Regression for females and males (overall and within race groups)

Dependent Variable: Log Earnings

Independent Variables	Female Complete	Male Complete	African Female Complete	African Male Complete	Coloured Female Complete	Coloured Male Complete	White Female Complete	White Male Complete
Age	0.0640	0.127***	0.00468	0.0912**	0.135	0.187	0.162**	0.178
	(0.0461)	(0.0456)	(0.0523)	(0.0377)	(0.0836)	(0.125)	(0.0678)	(0.174)
Age Squared	-0.000827	-0.00138**	-0.000160	-0.00101**	-0.00138	-0.00206	-0.00180**	-0.00178
	(0.000547)	(0.000578)	(0.000617)	(0.000454)	(0.000982)	(0.00151)	(0.000807)	(0.00215)
Urban	0.574***	0.473***	0.579***	0.514***	0.582**	0.444***	0.306	0.320
	(0.101)	(0.0945)	(0.106)	(0.0995)	(0.243)	(0.150)	(0.220)	(0.387)
African	-0.640***	-0.560**						
	(0.151)	(0.254)						
Coloured	-0.532***	-0.753***						
	(0.142)	(0.241)						
Parental Education	0.115***	0.0760***	0.115***	0.0606***	0.174***	0.113***	0.0951**	0.127***
	(0.0127)	(0.0109)	(0.0156)	(0.0107)	(0.0214)	(0.0212)	(0.0389)	(0.0255)
Parental Education Difference	0.0168	-0.000419	0.0244	-0.00496	0.00478	0.115***	-0.0233	-0.0261
	(0.0164)	(0.0101)	(0.0184)	(0.0104)	(0.0463)	(0.0362)	(0.0301)	(0.0348)
Education (Aux Residual)	0.151***	0.139***	0.158***	0.146***	0.226***	0.207***	0.192***	0.119***
	(0.0141)	(0.0154)	(0.0134)	(0.0159)	(0.0566)	(0.0285)	(0.0251)	(0.0437)
Education squared (Aux Residual Squared)	0.0118***	0.0112***	0.0156***	0.0145***	0.0306***	0.0145***	-0.00800	0.00727
	(0.00367)	(0.00267)	(0.00345)	(0.00264)	(0.0112)	(0.00441)	(0.00485)	(0.00502)
Constant	5.712***	4.769***	6.264***	4.991***	2.801	2.524	4.012**	3.049
	(0.938)	(0.838)	(1.067)	(0.769)	(1.760)	(2.460)	(1.640)	(3.305)
Observations	1,202	1,301	880	962	185	193	137	146
R-squared	0.514	0.383	0.365	0.317	0.456	0.459	0.372	0.172

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Base category: White

Table 9: Fields' Decomposition: Relative contributions to earnings inequality

9a: Partial Model

(i) Total female and male (for South Africa as a whole)

Female

Age	-0.0012
Age Squared	0.0043
Urban	0.0385
African	0.0256
Coloured	0.0018
Parental Education	0.0593
Education	-0.0485
Education Squared	0.3204
Residual	0.5998

Male

Age	-0.0022
Age Squared	0.0082
Urban	0.0257
African	0.0331
Coloured	0.0071
Parental Education	0.016
Education	0.0399
Education Squared	0.1856
Residual	0.6865

(ii) By race and gender group (with race inequality)

African

Female

Age	-0.0005
Age Squared	0.0011
Urban	0.0392
Parental Education	0.0388
Education	-0.2121
Education Squared	0.5513
Residual	0.5823

Male

Age	-0.0017
Age Squared	0.006
Urban	0.0235
Parental Education	-0.0037
Education	-0.045
Education Squared	0.3207
Residual	0.7002

Coloured

Female

Age	-0.002
Age Squared	0.0071
Urban	0.0359
Parental Education	0.0753
Education	-0.187
Education Squared	0.4561
Residual	0.6146

Male

Age	-0.0043
Age Squared	0.0174
Urban	0.0046
Parental Education	0.0437
Education	-0.2393
Education Squared	0.6309
Residual	0.5471

White

Female

Age	-0.002
Age Squared	0.008
Urban	0.019
Parental Education	0.0905
Education	0.9652
Education Squared	-0.3965
Residual	0.3159

Male

Age	-0.0033
Age Squared	0.0123
Urban	0.0355
Parental Education	0.0643
Education	-0.0894
Education Squared	0.1837
Residual	0.7968

Table 9: Fields' Decomposition: Relative contributions to earnings inequality

9b: Complete Model

(i) Total female and male (for South Africa as a whole)

Female

Age	-0.0007
Age Squared	0.0031
Urban	0.0585
African	0.0389
Coloured	0.0057
Parental Education	0.1655
Education (Aux Residual)	0.1548
Education Squared (Aux Residual Sqrd)	-0.0084
Residual	0.5825

Male

Age	-0.0019
Age Squared	0.0077
Urban	0.0469
African	0.0443
Coloured	0.0096
Parental Education	0.1015
Education (Aux Residual)	0.1338
Education Squared (Aux Residual Sqrd)	0.0018
Residual	0.6562

(ii) By race and gender group (with race inequality)

African

Female

Age	0.0002
Age Squared	-0.0002
Urban	0.0601
Parental Education	0.1595
Education (Aux Residual)	0.2033
Education Squared (Aux Residual Sqrd)	0.0063
Residual	0.5709

Male

Age	-0.0013
Age Squared	0.0053
Urban	0.0479
Parental Education	0.0852
Education (Aux Residual)	0.1796
Education Squared (Aux Residual Sqrd)	0.0056
Residual	0.6777

Coloured

Female

Age	-0.0008
Age Squared	0.0029
Urban	0.0607
Parental Education	0.2065
Education (Aux Residual)	0.1246
Education Squared (Aux Residual Sqrd)	-0.0117
Residual	0.6178

Male

Age	-0.0035
Age Squared	0.0151
Urban	0.0655
Parental Education	0.1546
Education (Aux Residual)	0.1563
Education Squared (Aux Residual Sqrd)	0.0213
Residual	0.5907

White

Female

Age	-0.002
Age Squared	0.0078
Urban	0.024
Parental Education	0.1755
Education (Aux Residual)	0.1514
Education Squared (Aux Residual Sqrd)	-0.0161
Residual	0.6594

Male

Age	-0.0033
Age Squared	0.0124
Urban	0.0344
Parental Education	0.1398
Education (Aux Residual)	0.0643
Education Squared (Aux Residual Sqrd)	0.0072
Residual	0.7451

Table 10: Bourguignon et al decomposition

10a: Partial Effects model

(i) Total (for South Africa as a whole)

	Female	Male
GINI 1: Derived from Actual Distribution of Earnings	0.60013	0.62788
GINI 2: Derived from the distribution after equalising circumstances	0.53644	0.60767

(ii) By race and gender group (within race inequality)

	African		Coloured		White	
	Female	**Male**	**Female**	**Male**	**Female**	**Male**
GINI 1: Derived from Actual Distribution of Earnings	0.58844	0.56511	0.52581	0.5757	0.38215	0.48977
GINI 2: Derived from the distribution after equalising circumstances	0.56429	0.56236	0.44965	0.59894	0.3651	0.4743

Table 10: Bourguignon et al decomposition

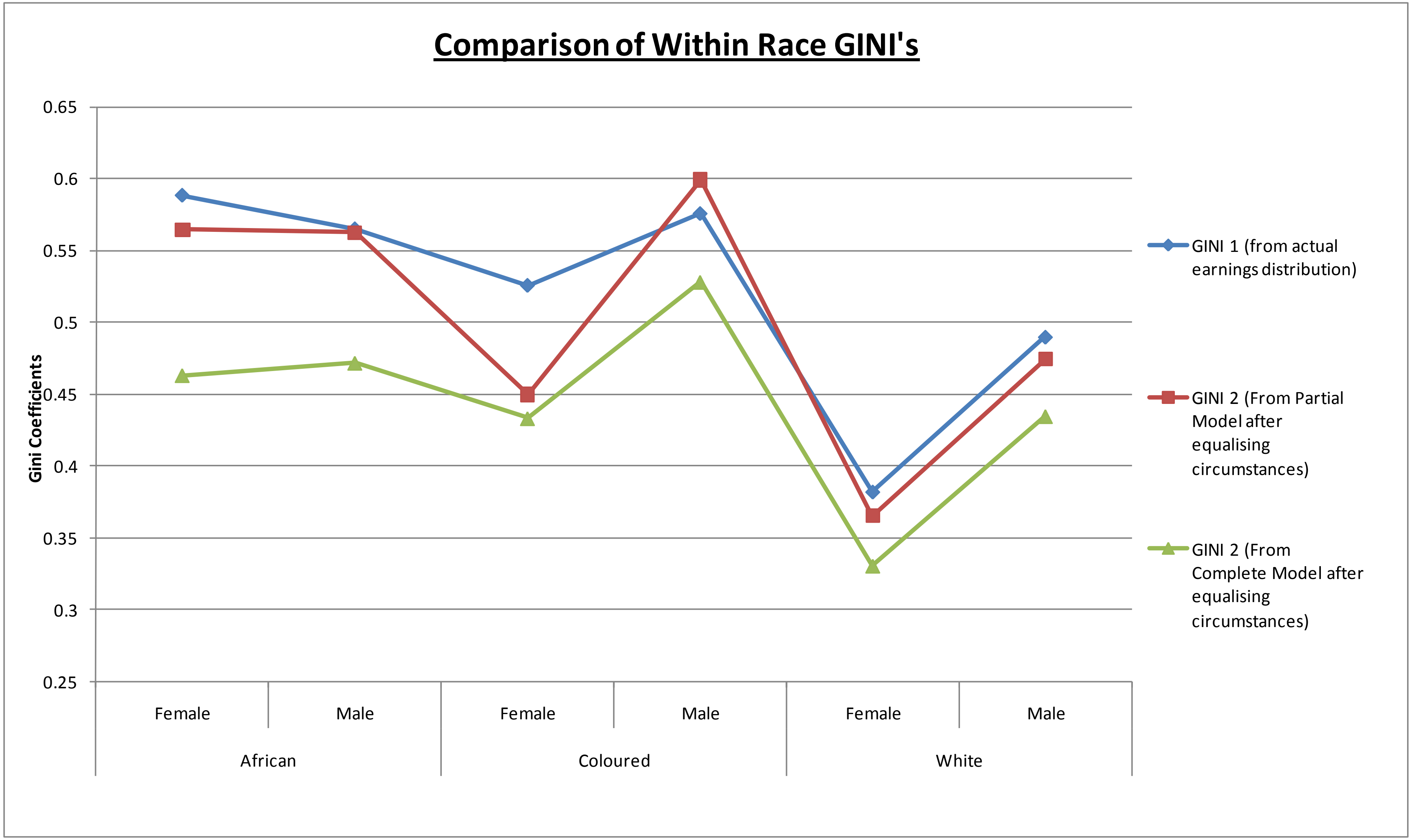
10b: Complete Effects model

(i) Total (for South Africa as a whole)

	Female	Male
GINI 1: Derived from Actual Distribution of Earnings	0.60013	0.62788
GINI 2: Derived from the distribution after equalising circumstances	0.45761	0.48657

(ii) By race and gender group (within race inequality)

	African		Coloured		White	
	Female	**Male**	**Female**	**Male**	**Female**	**Male**
GINI 1: Derived from Actual Distribution of Earnings	0.58844	0.56511	0.52581	0.5757	0.38215	0.48977
GINI 2: Derived from the distribution after equalising circumstances	0.4631	0.4717	0.43312	0.52802	0.33045	0.43457


Comparison of Within Race GINI's
Gini Coefficients
0.65
0.6
0.55
0.5
0.45
0.4
0.35
0.3
0.25
Female
Male
Female
Male
Female
Male
African
Coloured
White
GINI 1 (from actual earnings distribution)
GINI 2 (From Partial Model after equalising circumstances)
GINI 2 (From Complete Model after equalising circumstances)

Appendix: Fields' Derivation

Derivation of Gary Fields' Decomposition (Fields, 2003)

The starting point of the Fields' decomposition is the Mincerian earnings regression which can be expressed by equation 1 below:

$$\ln Y_{it} = a_t{}' Z_{it} \tag{1}$$

Above, the dependent variable is log of earnings ($\ln Y_i$), a_t is the vector of coefficients, Z_{it} is the vector of variables which determine earnings (including a constant) and ε_{it} is the i.i.d. error term.

$$a_t = [\alpha_t\, \beta_{1t}\, \beta_{2t} \ldots \beta_{jt}\, 1] \tag{2}$$

$$Z_{it} = [1\ x_{i1t}\, x_{i2t} \ldots x_{iJt}\, \varepsilon_{it}] \tag{3}$$

The log-variance decomposition

To attain the inequality decomposition, the variance of both sides of equation 1 are taken. For the left-hand side (LHS) one attains the log-variance of earnings (a simple earnings inequality measure). On the right-hand side (RHS) it is possible through a series of manipulations (which are explained below) to derive the percentage contributions of each RHS variable to the earnings inequality of the LHS.

Theorem (Mood, Graybill and Boes): Following the theorem of Mood, Graybill and Boes, let $A_1 \ldots A_p$, and $B_1 \ldots B_q$ be two sets of random variables. Furthermore, let $a_1 \ldots a_p$, and $b_1 \ldots b_q$ be two sets of constants. Then:

$$cov[\sum_{p=1}^{P} a_p A_p, \sum_{q=1}^{Q} b_q, B_q] = \sum_{p=1}^{P} \sum_{q=1}^{Q} a_p b_q cov[A_p, B_q] \tag{4}$$

If equation 4 is applied to the single random variable, ln Y where,

$$\ln Y = \sum_{j=1}^{J+2} a_j Z_j$$

then,

$$cov[\sum_{j=1}^{J+2} a_j Z_j , \ln Y] = \sum_{j=1}^{J+2} cov[a_j Z_j, \ln Y] \tag{5}$$

However, on the LHS of equation 5 is simply the covariance of ln Y with itself which is the variance of ln Y. We therefore attain:

$$\sigma^2(\ln Y) = \sum_{j=1}^{J+2} \mathrm{cov}[\,a_j Z_j, \ln Y] \tag{6.a.}$$

By dividing through by $\sigma^2(\ln Y)$ the relative contributions become clear, as seen below:

$$100\% = \frac{\sum_{j=1}^{J+2} \mathrm{cov}[\,a_j Z_j, \ln Y]}{\sigma^2(\ln Y)} \equiv \sum_{j=1}^{J+2} S_j\,(\ln Y) \tag{6.b.}$$

Where each $S_j\,(\ln Y)$ are the "relative factor inequality weights" which are defined below.

$$S_j\,(\ln Y) = \frac{cov[a_j Z_j, \ln Y]}{\sigma^2(\ln Y)} \tag{6.c.}$$

The relative factor inequality weights sum exactly to $R^2(\ln Y)$

Finally, because the ordinary correlation coefficient is related to the covariance as follows:

$$cor\,[a_j Z_j, \ln Y] = \frac{cov[a_j Z_j, \ln Y]}{\sigma(a_j Z_j)\sigma(\ln Y)} \tag{7}$$

We can combine the above equations and represent them by equation 8 below:

$$S_j\,(\ln Y) = \frac{cov[a_j Z_j, \ln Y]}{\sigma^2(\ln Y)} = \frac{a_j * \sigma(Z_j) * \mathrm{cor}[Z_j, \ln Y]}{\sigma(\ln Y)} \tag{8}$$

where

$$\sum_{j=1}^{J+2} s_j(\ln Y) = 100\%$$

and

$$\sum_{j=1}^{J+2} s_j(\ln Y) = R^2(\ln Y)$$

Such that the proportion explained by the j'th explanatory variable, $P_j\,(\ln Y)$ is:

$$P_j\,(\ln Y) = \frac{s_j(\ln Y)}{R^2(\ln Y)}$$