Changes in inequality in South Africa:
The effect of human capital on inequality using decomposition techniques

Julia Garlick

GRLJUL001

A minor dissertation submitted in partial fullfilment of the requirements for the award of the degree of Master of Social Science in Applied Economics

Faculty of the Humanities

University of Cape Town

2011

COMPULSORY DECLARATION

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

Signature: Signed by candidate Date: 10/02/2011

Signature removed

G:\ghu\forms\m-title-page

ABSTRACT

This paper analyses changes in the inequality of employment earnings in South Africa between 1993 and 2008. This is done through the implementation of a relatively new method of comparing the contributions of various factors to observed changes in incomes distributions. Counterfactual earnings distributions are calculated to assess the effect of changes in the returns to individual characteristics on inequality, and whether these changes operated through changing labour market outcomes or earnings for those employed. The effect of changes in the distribution of education is also calculated, and this is again decomposed into effects operating through labour market outcomes and effects operating through employment earnings. The method is designed to be used in labour markets with high unemployment rates, and incorporates analysis of labour market effects and earnings into one unified approach, which is unusual in the literature on earnings inequality. The paper finds that changes in the composition of the labour force increased inequality in earnings from regular wage employment but decreased inequality in earnings from all labour force activities. Changes in remuneration increased inequality in regular earnings, but decreased it for total labour market earnings. Changes in the distribution of education increased inequality in total labour earnings, through their effect on labour market outcomes and on the wage. Changes in education increased inequality in regular earnings through the mechanism of labour market outcomes, but decreased it through their impact on remuneration. The paper also raises the point that earnings from regular jobs behave markedly differently to the total earnings distribution, which should be of interest to future analyses of the South African earnings distribution. The data used are the first wave of the National Income Dynamics Survey, collected in 2008, and the Project for Statistics on Living Standards and Development, collected in 1993.

Table of Contents

1. INTRODUCTION

South Africa has the unfortunate combination of attributes of extreme levels of inequality and high and intense poverty. Between 1993 and 2008, earnings inequality has fallen but overall poverty levels have increased[1]. According to the international literature, inequality has a number of unfortunate effects, particularly on the poor[2]. Thus the combination of poverty and inequality in South Africa is of great interest to researchers and policy-makers alike.

Inevitably, the behaviour and determinants of inequality are a subject of much debate and research. Recent work has found that the key driver of an individual's position in the South African income distribution is their access to labour market earnings; earnings from employment also account for a disproportionate share of income inequality[3]. Income inequality is thus related to the labour market through two paths: employment and wages among the employed. Most research in South Africa has tended to focus on one or the other of these questions, without examining their joint effect[4]. This is unfortunate because it is very implausible that their effects are independent of each other – issues that affect labour market access are likely also to affect a worker's place in the earnings distribution once he has employment, though whether these effects operate to reinforce or counteract each other is open to theoretical and empirical debate.

One of the main determinants of labour market success – whether measured by employment or earnings – is human capital, and specifically education[5]. This is as true in South Africa as it is

1 Hoogeveen, J. and B. Ozler. (2005). Not separate, not equal: Poverty and inequality in post-Apartheid South Africa. William Davidson Institute Working Paper No. 739. University of Michigan Business School.
Leibbrandt, M., I. Woolard, H. McEwen and C. Koep. (2010). Inequality outcomes in South Africa. University of Cape Town. Southern Africa Labour and Development Research Unit.
Leibbrandt, M., J. Levinsohn and J. McCrary. (2010). Incomes in South Africa after the Fall of Apartheid. *Journal of Globalization and Development* Vol. 1(1) Article 2.
Seekings, J. (2007). Poverty and inequality after Apartheid. Centre for Social Science Research. CSSR Working Paper No. 200.
Van der Berg, S., R. Burger, M. Louw, R. Burger and D. Yu. (2006) Trends in poverty and inequality since the political transition. Development Policy Research Unit Working Paper 104

2 Birdsall, N. and J. Londono. (1997) Asset Inequality Matters: An Assessment of the World Bank's Approach to Poverty Reduction. In *The American Economic Review* 87(2).

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

4 An important exception is the recent paper by Leibbrandt, Levinsohn and McCrary (2010)

5 Card, D. (1999). The causal effect of education on earnings. In *Handbook of Labour Economics Volume 3.* Eds Orley Ashenfelter and David Card. Amsterdam: Elsevier.

internationally[6]. An important determinant of changes in inequality is thus changes in the educational achievement and attainment of the potential workforce. There have been substantial changes in the distribution of education between 1993 and 2008, which, by the above reasoning, would be expected to have had a marked effect on inequality[7]. However, changes in inequality have been modest. The question of what exactly education's effect on inequality has been, and through what channels it has operated, is thus of great interest. Unfortunately, this is a question which is difficult to answer – any analysis would have to take into account education's effect on both labour market outcomes and earnings. These questions are seldom answered in a unified analysis due to methodological and modelling difficulties.

This paper makes an attempt at addressing this gap in the literature. Inspiration is taken from Bourguignon, Ferreira and Leite's 2008[8] paper on decomposing changes in income distributions into their component causes, and the methodology is adapted from one proposed (and implemented in the Argentinian case) by Gonzalez-Rozada and Menendez (2006)[9] to analyse high unemployment labour market equilibria.

It is necessary to make explicit allowance for South Africa's high unemployment because extremely high unemployment is likely to act as its own reinforcer – high unemployment increases of the cost of finding jobs for workers and decreases the likelihood that they will receive a wage offer they consider fair. Most traditional economic models assume that unemployment is a relatively minor phenomenon that can be largely disregarded when agents' behaviour is being examined. This assumption is less likely to be valid when unemployment is very high, for reasons which are discussed in the model presented in Section 3.

The methodology used in this paper solves both problems – combining the analysis of employment and earnings and analysing high unemployment equilibria – by allowing for randomness in labour market outcomes when counterfactual analysis is performed. This allows for multiple counterfactual

6 Lam, D., M. Leibbrandt and J. Garlick. (2010). Investing in human capital to reduce inequality. Discussion paper prepared for Centre for Development Enterprise roundtable discussion, January 2010.

7 Lam, D., M. Leibbrandt and J. Garlick. (2010). Investing in human capital to reduce inequality. Discussion paper prepared for Centre for Development Enterprise roundtable discussion, January 2010.

8 Bourguignon, F., F.H.G. Ferreira and P.G. Leite. (2008). Beyond Oaxaca-Blinder: Accounting for differences in household income distributions. In *Journal of Economic Inequality* Vol. 6.

9 Gonzalez-Rozada, M. and A. Menendez. (2006). Why have urban poverty and income inequality increased so much? Argentina, 1991 – 2001. In *Economic Development and Cultural Change*. 55(1).

workforces to be drawn from the pool of potential labour force participants and for them to be assigned wages. Through comparison of the resulting earnings distributions, it is possible to assess the effect of changes in labour market outcomes against those of earnings on the distribution of earnings. This also allows the effect of changes in characteristics, specifically education, to be analysed and assigned to either or both of the levels of analysis of earnings.

Section 2 provides a brief summary of the international evidence on inequality and growth, and the relationship between human capital in the form of education and inequality. The South African literature on inequality and education is discussed. Section 3 presents a search model of the labour market to illustrate the role that information incompleteness can have in producing sub-optimal labour market outcomes, and to make the point that reaching an equilibrium is not a guarantee of optimality. Section 4 presents summary statistics on the data used for this analysis: the first wave of the National Income Dynamics Survey of 2008 and the Project for Statistics on Living Standards and Development from 1993. Section 5 explains the methodology used to create the counterfactual earnings distributions, and Section 6 discusses and analyses these distributions. Section 7 concludes.

2. LITERATURE REVIEW

2.1 Inequality and growth

South Africa possesses the dubious distinction of having among the highest levels of measured inequality in the world, coupled with high levels of poverty[10]. South Africa also suffers from very low intergenerational mobility: if your parents are poor, the probability is high that you will be, too[11]. The following literature review provides a brief overview of the international literature on the consequences of inequality, to motivate the research question, and then examines the relationship between education on labour market outcomes and inequality. Section 2.3 discusses education, inequality and the labour market in the South African context.

The effects of inequality are much debated, but there is evidence that its effect on the poor is particularly severe. The lower income quintiles typically experience lower growth in the presence of inequality than they would otherwise, and lower growth than the mean for their economies[12]. Thus, while growth might benefit the poor in the long run, in the short-run it contributes to increasing inequality. As inequality dampens growth, this creates a feedback loop which harms all members of the economy. This would not be such a concern from an ethical point of view if there was substantial change in the composition of the lower income quintiles. However, most developing countries have quite rigid economic stratifications. In the majority of cases, the poor remain poor and the rich remain rich. This holds true between generations, too[13].

The poor are also especially vulnerable to the composition of change in the globalised world. Though overall wage inequality between countries is falling, wage inequality within countries is rising internationally, with increases between but also within the groups of educated and uneducated workers. The gap between skilled and unskilled workers is being increased by increases in the returns to education, which can partly be attributed to education's increased value in the presence of technological change. This drives rising inequality among educated workers too, as certain types of

10 World Development Report 2006, *Equity and Development*, 2007.
World Development Report 2007, *Development and the Next Generation*, 2008.

11 Girdwood, S. and M. Leibbrandt. (2009). Intergenerational mobility: Analysis of the NIDS Wave 1 Dataset. Discussion Paper No. 15.

12 Bourguignon, F. (2004). The poverty-growth-inequality triangle. Indian Council for Research on International Economic Relations, New Delhi Working Papers 125

13 World Development Report 2006, *Equity and Development*, 2007.

education, or pure ability, are more productive[14]. However, changes in educational returns cannot explain the entire difference, nor the widening gaps among unskilled workers. Unskilled workers also face more variation in wages due to factors such as industry-specific shocks, as they are less able to shift between industries and technologies. Education provides increased flexibility, which helps to explain its additional attraction, over and above increased remuneration. However, much of the variation in wages between similar workers remains unexplained[15].

Birdsall and Londono[16] find that controlling for human capital and asset inequality removes the significance traditionally assigned to income inequality in hindering growth. This implies that the role usually attributed to income inequality may be a proxy for the effects of asset and human capital inequality. A paper by Klasen[17] focuses on gender inequality in education but arrives at similar implications. If education is distributed on any grounds other than students' ability to benefit from it, some able students will not have access to higher levels of education and their places will be taken by less able students from the favoured group, whether the latter group is chosen by taste-based discrimination – age, gender, race – or financial advantage. The less able students will become less able workers, lowering the marginal productivity of human capital. As human and physical capital are widely believed to be complements, this results in lower returns to investment in physical capital. At a minimum, this will result in lower growth for a given level of investment, and other things being equal, will decrease total investment in the economy. Further, as these less able students require more resources to achieve the same education levels, there will be a smaller equilibrium number of educated workers in the economy, again decreasing efficiency. Thus, even if we are interested only in the efficiency of overall production, the quantity and distribution of education available to the population is of concern.

However, this is not our only or even main interest. Development economics is deeply concerned with questions of poverty and income distribution, and the role of human capital in these areas appears to be large. The literature on international development cited above, and summarised in the

14 Heckman, J., L. Lochner and C. Taber. (1998). Explaining Rising Wage Inequality: Explorations with a Dynamic General Equilibrium Model of Labor Earnings with Heterogeneous Agents. In *Review of Economic Dynamics*, 1, 1998.
Goldin, C. and L. Katz. (2008). *The Race between Education and Technology.* Cambridge, MA: Harvard University Press.

15 Gould, E., O. Moav and B. Weinberg. (2001). Precautionary Demand for Education, Inequality, and Technological Progress. In *Journal of Economic Growth*, 6(4).

16 Birdsall, N. and J. Londono. (1997) Asset Inequality Matters: An Assessment of the World Bank's Approach to Poverty Reduction. In *The American Economic Review* 87(2).

17 Klasen, S. (2002). *Low Schooling for Girls, Slower Growth for All? Cross-Country Evidence on the Effect of Gender Inequality in Education on Economic Development*. *The World Bank Economic Review*, 16(3).

World Development Reports of 2006 and 2007[18], makes a strong case for the fact that inequality in education plays a central role in perpetuating and generating inequality in labour market earnings and in income inequality more generally. This case has strengthened in the increasing globalised world. Workers now compete on international labour markets. Even individuals who are not mobile are often competing for jobs which are. This increase in competition has resulted in a steeper wage scale, with the ratio of returns at the top of the earnings distribution to those at the bottom being far higher than was generally the case in geographically segmented labour markets. A strong predictor of one's place in this distribution is one's education level[19]. Given this, it is worth discussing the precise mechanisms through which education affects labour market outcomes before moving on to the South African results.

2.2 Education and the labour market

The theoretical literature on education's labour market impacts splits into two streams – the human capital model and the signalling approach[20]. In the former view, education is one way of increasing workers' human capital, and thus their productivity. Increased wages are associated with higher levels of education because the marginal product of the worker has increased and thus her value to the firm is higher, and to compensate the worker for costs associated with acquiring her human capital (both direct costs – fees, training materials and mental costs – and indirect costs such as foregone wages). Human capital in these models includes characteristics other than education – health, skills gained through work experience and inherent ability[21]. However, the focus of this paper is on education, though ability is often assumed to be correlated with education.

The second school of thought argues that education's effect on wages is not causal, through increased productivity, but rather an information device. There is always an information asymmetry between employer and potential employee about the employee's quality. Workers cannot simply tell firms their productivity level, as there is a strong incentive for all workers to overstate their value

18 World Development Report 2006, *Equity and Development*, 2007 and World Development Report 2007, *Development and the Next Generation*, 2008.

19 Card, D. (1999). The causal effect of education on earnings. In *Handbook of Labour Economics Volume 3.* Eds Orley Ashenfelter and David Card. Amsterdam: Elsevier.

20 Berndt, E.R. (1991). *The Practice of Econometrics: Classic and Contemporary.* Massachusetts: Addison-Wesley Publishing Company.

21 Berndt, E.R. (1991). *The Practice of Econometrics: Classic and Contemporary.* Massachusetts: Addison-Wesley Publishing Company.

and acquire higher wages, at least temporarily. However, education provides a screening device for employers and a signalling device for workers. Educational attainment is costly – in terms of time, effort, fees and foregone wages. However, according to this argument, it is less costly for more able individuals. They will need to expend less time and effort, and thus incur fewer direct and indirect costs, to obtain any given level educational qualification. Thus, education gives a 'hard to fake' signal of ability, and is valuable to employers and workers for that reason[22].

In practice, the models give similar empirical results – higher labour market returns to higher education levels – and thus provide few ways to choose between them. This is unfortunate, as their policy implications are substantially different. The human capital framework predicts that increasing average education levels will improve average worker productivity, while the signalling framework predicts that rising education will change employers' interpretation of workers' education levels but will not affect the real economy. The former would suggest that increased investment in education is valuable, while the latter implies that this would simply dilute the signal employers receive from education[23]. Testing these competing hypotheses is beyond the scope of this paper, however. In the international literature problems with the signalling hypothesis have emerged, but these have not proven fatal, and the theory still has useful insights to offer. However, in the South African context, the mechanisms proposed by both theories may be somewhat blurred.

A major theme that will emerge from the discussion of the South African labour market that follows is that there is a growing disjuncture between the quality and quantity of education received in the South African school system[24]. In such a situation, the quality of the signal given by the level of education obtained will be diluted. Employers (and researchers) need additional information about the qualifications of workers to make informed assessments of their ability. Unfortunately, while this information may be available to employers, it is less often available to researchers. Thus, the impact of education on wages may be subject to substantial mismeasurement.

This problem also affects the human capital model of education. If the same qualification confers different increases in human capital on different people, measuring the impact of a particular level

22 Berndt, E.R. (1991). *The Practice of Econometrics: Classic and Contemporary.* Massachusetts: Addison-Wesley Publishing Company.

23 Card, D. (1999). The causal effect of education on earnings. In *Handbook of Labour Economics Volume 3.* Eds Orley Ashenfelter and David Card. Amsterdam: Elsevier.

24 World Development Report 2006, *Equity and Development*, 2007.

of educational achievement becomes more difficult. The standard result of measurement error is attenuation bias – lower measured effects. When combined with omitted variable bias – which tends to increase the magnitude of coefficients – these effects tend to cancel out in many cases[25]. In South African education studies, as in most examinations of education's effects, omitted variable bias is inevitably present due to the interaction of education and ability, which is difficult to measure. However, with the additional measurement error created by variance in education quality, the two effects may not fully offset each other in the South African case, leading to attenuated estimates. It is important to bear this constraint in mind and understand that the estimated relationships between education and employment or wages discussed in the next section may be lower bounds on the true relationship.

2.3 South African evidence

In South Africa, as in many other countries, the dominant driving force of inequality is the labour market. Leibbrandt, Woolard, Finn and Argent review the post-apartheid empirical work on the relative impact on household income inequality of different income sources including wage income, state transfers and remittances[26]. They then bring this work up to date using data from the 2008 National Income Dynamics Study. Two key points emerge. First, from the start of the post-apartheid period onwards, the relative success (or lack thereof) of household members in the labour market is the dominant driver of their position in the household income distribution, both absolute and relative. Wage income comprises 70% of overall income, but accounts for almost 85% of measured inequality. Second, there are two prongs to the labour market's role; namely, whether the members of the household have employment at all and then, for those with such earnings, their position within the distribution of labour market earnings.

Returns to education in South Africa are somewhat atypical. Earnings as a function of education are usually found to be concave, with decreasing rates of return between secondary and tertiary levels, in studies using developed world data. In South Africa, the earnings function appears to be convex. Marginal returns to education increase with education level, rather than decreasing. This

25 Ashenfelter, O. and C. Rouse. (1998). Income, Schooling, And Ability: Evidence From A New Sample Of Identical Twins. In *The Quarterly Journal of Economics*. 113(1)

Card, D. (2001). Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems. In *Econometrica*. 69.

Ashenfelter, O., C. Harmon and H. Oosterbeek. (1999). A review of estimates of the schooling/earnings relationship, with tests for publication bias. In *Labour Economics*. 6(4).

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

result holds for the employment function too – the marginal effect of education on the probability of employment increases at higher education levels This is found to be the case whether these returns are measured for employment or remuneration[27].

These extremely high returns were at one stage attributed to the deliberate restriction of education to whites that took place under Apartheid. Mwabu and Schultz's 2000 paper[28], using data from the 1993 Project for Statistics on Living Standards and Development (PSLSD)[29], argued that once education became more widely available to the population, these returns could be expected to decrease and fall in line with international norms. However, this has not occurred. Access to education has expanded considerably in the last 16 years, but authors continue to find unusually high returns to education. Studies of South African data over the past decade have revealed marked increases in years of complete secondary and tertiary education. More South Africans are completing school and going on to tertiary studies, which might have been expected to go some way towards decreasing the high measured returns to schooling. However, this was not the case. Recent research[30] found that the returns to complete secondary schooling, and even incomplete tertiary schooling, remain incredibly high; if anything, they appear to have increased over time. In 1997, an African man with a matric could expect to earn 26% more than one with only grade 11, while in 2007 this figure was 29%. For all men, a complete university qualification, relative to a matric, yielded an earnings premium of 218% in 2007. These changes were mirrored in the effect of education on employment: between 2000 and 2007, the probability of employment for a tertiary graduate went from double to triple relative to a secondary graduate[31].

Peculiarly, in the same time period that returns to higher education increased, unemployment among tertiary graduates increased. Various authors find that unemployment rates increased markedly for those with complete secondary or tertiary schooling. There is some evidence that this may be due to quality problems in the South African education system. Graduates from historically black

27 Keswell, M. and L. Poswell. (2004). Returns to education in South Africa: A retrospective sensitivity analysis of the available evidence. In *The South African Journal of Economics*. 72(4).
Lam, D., M. Leibbrandt and J. Garlick. (2010). Investing in human capital to reduce inequality. Discussion paper prepared for Centre for Development Enterprise roundtable discussion, January 2010.

28 Mwabu, G. and T.P. Schultz. (2000). Wage premiums for education and location of South African workers, by gender and race. In *Economic Development and Cultural Change,* 48(2).

29 The same dataset will be used for part of the analysis of this paper, making their results particularly relevant.

30 Lam, D., M. Leibbrandt and J. Garlick. (2010). Investing in human capital to reduce inequality. Discussion paper prepared for Centre for Development Enterprise roundtable discussion, January 2010.

31 Branson, N., M. Leibbrandt and T. Zuze. (2009). The demand for teriary education in South Africa. Final report to the Centre for Higher Education Transformation, September 2009.

institutions, or those with diplomas as opposed to degrees, experience more difficulty finding employment than do those from historically white institutions.[32] Quality at the secondary school level is also problematic, both in terms of level and dispersion, according to many authors[33].

The combination of extremely high returns to advanced education and relatively limited advanced education makes South Africa's inequality situation somewhat easier to understand. The South African literature is united in its assessment of the direction of South African inequality trends, if not their magnitude – slight increases in overall income inequality from 1993 onwards. Inequality between races appears to have decreased somewhat, but these changes were offset by increased inequality within race groups. A major driver of this is agreed to be the increasing income disparity within the black racial group[34]. This is coupled with continued deep and widespread poverty and declining real average incomes. The overall effect is that the wage distribution has shifted to the left, with a substantial tail of high earners[35]. The combination of high inequality with extreme poverty is an unpalatable one: those individuals at the bottom of the income distribution are not only relatively in trouble, but face real income constraints. The provision of social assistance grants and government-provided services may mitigate the level of deprivation experienced by the poor[36],

32 Dias, R. and Posel, D. (2007). Unemployment, education and skills constraints in post-apartheid South Africa. Development Policy Research Unit Working Paper No. 07/120.
Bhorat, H. (2004). Labour market challenges in the post-apartheid South Africa. In *The South African Journal of Economics* 72(5).
Oosthuizen, M. (2006). The post-apartheid labour market: 1995 – 2004. Development Policy Research Unit Working Paper No. 06/103.

33 Moll, P. (1998). Primary Schooling, Cognitive Skills and Wages in South Africa. In *Economica*, 65(258).
Oosthuizen, M. (2006). The post-apartheid labour market: 1995 – 2004. Development Policy Research Unit Working Paper No. 06/103.
Pauw, K., M. Oosthuizen. and C. Van der Westhuizen, C. (2006). Graduate unemployment in the face of skills shortages: A labour market paradox. Development Policy Research Unit Working Paper No. 06/114.
Van der Berg, S. and R. Burger. (2003). Education and socio-economic differentials: A study of school performance in the Western Cape. Development Policy Research Unit Working Paper No. 03/73.
Altman, M. and D. Lee. (2004). Meeting equity targets: Are there enough graduates?. Employment & Economic Policy Research Programme, Human Sciences Research Council

34 Hoogeveen, J. and B. Ozler. (2005). Not separate, not equal: Poverty and inequality in post-Apartheid South Africa. William Davidson Institute Working Paper No. 739. University of Michigan Business School.
Leibbrandt, M., I. Woolard, H. McEwen and C. Koep. (2010). Inequality outcomes in South Africa. University of Cape Town. Southern Africa Labour and Development Research Unit.
Seekings, J. (2007). Poverty and inequality after Apartheid. Centre for Social Science Research. CSSR Working Paper No. 200.
Van der Berg, S., R. Burger, M. Louw, R. Burger and D. Yu. (2006) Trends in poverty and inequality since the political transition. Development Policy Research Unit Working Paper 104

35 Leibbrandt, M., J. Levinsohn and J. McCrary. (2010). Incomes in South Africa after the Fall of Apartheid. *Journal of Globalization and Development* Vol. 1(1) Article 2.
Leibbrandt, M., I. Woolard, H. McEwen and C. Koep. (2010). Inequality outcomes in South Africa. University of Cape Town. Southern Africa Labour and Development Research Unit.

36 Van der Berg, S., R. Burger, M. Louw, R. Burger and D. Yu. (2006) Trends in poverty and inequality since the

but does not remedy the problems described above.

This suggests that income inequality in South Africa requires closer examination. While all studies agree that it is rising, there has been less research into the causes of this increase. One of the key approaches in the literature to inequality analysis is to use counterfactual techniques. These techniques provide a powerful set of tools with which to compare two income distributions and to analyse the contribution of various factors to differences between the distributions. This is done by assigning one (original) distribution certain characteristics of the other (alternate) distribution; which characteristics or structures are changed depends on the specific decomposition approach used. The result is a distribution (or series of distributions) that lies between the two observed distributions, with some characteristics of each, which is called a counterfactual distribution. The differences between the measures of interest (means, variance, or in this case, inequality measures) for the observed and counterfactual distributions allow the influence of the characteristics being examined to be assessed. The closer the counterfactual distribution is to the alternate observed distribution, the greater the contribution of the changed characteristics to the differences between the two observed distributions[37].

Unfortunately, counterfactual techniques are typically unable to integrate multiple sources of change into one analysis. When examining changes in an income distribution, the problem is that (at least) three separate sources of change might be affecting the distribution, and may be driving it in different directions. The three potential causes of observed changes are characteristics of the labour force, returns to particular characteristics, and changes in selection into the labour force. Standard decomposition analysis has to treat the third area as a separate question, which is seldom integrated into the former two analyses. In the South African case, the three areas have not typically been examined in combination. The paper by Leibbrandt, Levinsohn and McCrary (2010) goes some way to addressing this, with its careful analysis of different approaches to the counterfactual question, and an assessment of all three potential sources of change in the income distribution. But in general, few papers using South African data address more than one aspect of the problem, and almost none have considered all three.

This is unfortunate. There are likely to be two distinct sources of changes in inequality in South Africa – remuneration of the labour force, and selection of the labour force. Traditional measures of

political transition. Development Policy Research Unit Working Paper 104

37 Fortin, N., T. Lemieux and S. Firpo. (2010). Decomposition Methods in Economics. In *Handbook of Labour Economics Volume 4.* Eds Orley Ashenfelter and David Card. Amsterdam: Elsevier / North Holland.

earnings inequality discard zero earners from the distribution, which would tend to understate inequality, particularly by any measure that places more weight on lower earners. Most analyses of inequality have avoided dealing with this issue, and have simply analysed earnings inequality among the employed, rather than including the unemployed in inequality measures[38]. In the South African case, with extremely high unemployment, this is particularly unfortunate as such a large portion of the population is lost. One approach that has been used is to add zero earners back into the wage distribution[39]. While this is likely to result in more realistic inequality measures it cannot control for possible interactions between employment and earnings in terms of their effects on inequality and it makes extremely strong distributional assumptions due to the combination of continuous and discrete distributions in the wage (piling the unemployed at zero amounts to appending a discrete distribution to the continuous distribution of wage earners).

A new approach, pioneered by Gonzalez-Rozada and Menendez (2006), offers an alternative. These authors integrate the questions of selection and labour market returns through the use of micro-simulations and decomposition analysis. Their technique is designed for a labour market in which there is substantial unemployment – Argentina during the 1990s and early 2000s – and is thus highly suited to the high-unemployment situation of South Africa. The following sections will briefly discuss the theoretical model underpinning their approach, and Section 5.1.1 explains in detail how their model was adapted to suit the question of how inequality has changed in South Africa between 1993 and 2008, and more specifically, what the effect of changes in the distribution of education has been.

38 Leibbrandt, M., J. Levinsohn and J. McCrary. (2010). Incomes in South Africa after the Fall of Apartheid. *Journal of Globalization and Development* Vol. 1(1) Article 2.

39 Seekings, J. (2007). Poverty and inequality after Apartheid. Centre for Social Science Research. CSSR Working Paper No. 200.

3. THEORETICAL MODEL

The model underlying the analysis in this paper is a fairly standard job search specification. This model, first advanced by Stigler (1961, 1962) and McCall (1970)[40], presents the idea that activities by individuals not currently employed in the labour market can still be productive. Specifically, the model argues that individuals must devote effort to the search for employment. Thus, even if an individual is not in wage employment, their efforts to find such employment are economically productive. The underlying implications of the model are that (1) job search is costly, at some level (financial, physical or psychic), (2) information can be imperfect, leading to uncertainty and (3) even in the presence of imperfect information, individuals can make optimal decisions, *provided* the uncertainty is not too significant relative to meaningful signals.

McCall's fundamental insight was that the stopping condition of the job search is not an optimal number of searches for a particular individual; it is a wage. Each individual who is engaged in the job search faces various costs and has trade-offs to make. He incurs a cost each time he searches, c; he presumably has access to some sort of unemployment support, u; and he knows his own discount rate (β) and his characteristics and their expected return in the labour market. Thus, he has a belief set Ω about the various parameters influencing his decision to search: $\Omega = \{c, u, \beta, \omega\}$. $\omega : X \rightarrow \mathbb{R}_+$ is a function of the individual's characteristics that gives rise to a wage $\omega(X)$ with probability density function $q(\omega)$. Individuals' wages are not perfectly determined by their characteristics for two reasons. First, the search model assumes imperfect information – the worker may receive different wage offers from different employers, due to the employers assessing his ability differently. Second, different potential jobs may simply pay different wages. If markets are imperfectly integrated, this may affect even very similar jobs. Finally, we assume that all individuals receive utility from income, and that utility from leisure is negligible (this is purely to simplify the model).

In the South African case, an individual's labour market problem can be represented as a two-stage optimal stopping problem. In the first stage, the individual must decide whether or not to enter the labour market. If he does not, he receives his unemployment support for the rest of his life[41]. If he

40 McCall, J. J. (1970). Economics of Information and Job Search. In *The Quarterly Journal of Economics*. 84(1)
Stigler, G.J. (1961). The Economics of Information. In *Journal of Political Economy*. 69(3)
Stigler, G.J. (1962). Information in the Labour Market. In *Journal of Political Economy*. 70(5)

41 This could be support from relatives or other household members, personal savings, or simply zero. If the last

does enter the labour market, he pays a search cost c and faces a draw from the wage distribution, conditional on his characteristics X. As the South African labour market is a high unemployment scenario, we allow for the possibility that the individual does not receive an offer, conditional on his characteristics X, with probability $\theta: X \to [0,1]$. If he does not receive a wage offer, he simply draws his unemployment support for another period. If he does receive an offer, he chooses between accepting it – in which case this will be his lifetime wage – or rejecting it and choosing to search another period. In the latter case, he pays another search cost and receives a wage offer with probability θ. In this formulation, the individual's maximisation problem can be represented as:

$$\begin{aligned} V(\omega) = max \Bigg\{ & \frac{u}{1-\beta}, -c \\ & + max \left[\frac{\omega(X)}{1-\beta} \theta(X) + \big(1-\theta(X)\big)u, -c \right. \\ & \left. + \theta(X)\beta \int v(\omega'(X)) q(\omega'(X)) d(\omega(X)) + \big(1-\theta(X)\big)\beta u \right] \Bigg\} \end{aligned}$$

In the ordinary course of events, the individual knows his wage distribution and can form assessments of his expected utility in each scenario, thus allowing him to make decisions that maximise his utility. McCall (1970) discusses the possibility of an individual's assessment of his wage distribution being fundamentally wrong, and shows that the result is an inefficient decision from the point of view of both the worker and society.

The argument made in this paper is that, in the South African case, structural problems in the labour market make it extremely difficult for any worker to form an accurate assessment of his wage distribution. Information flows so imperfectly that there is systematic misestimation of $q(\omega)$ and θ. This results in the appropriate specification of the individual's problem being:

$$\begin{aligned} V(\omega) = max \Bigg\{ & \frac{u}{1-\beta}, -c \\ & + max \left[\frac{\omega(X+\zeta_1)}{1-\beta} \theta(X+\zeta_2) + \big(1-\theta(X+\zeta_2)\big)u, \right. \\ & -c + \theta(X+\zeta_2)\beta \int v\big(\omega'(X+\zeta_1)\big) q\big(\omega'(X+\zeta_1)\big) d\big(\omega(X+\zeta_1)\big) \\ & \qquad \left. + \ (1-\theta(X+\zeta_2))\beta u \right] \Bigg\} \end{aligned}$$

---

scenario holds, the worker clearly has very strong incentives to choose to search.

where $\zeta_1 \sim N(\mu_1, \sigma_1^2)$ and $\zeta_2 \sim N(\mu_2, \sigma_2^2)$asymptotically, both μ and σ^2 unknown – perturbation terms representing the noise in the distribution created by information problems. ζ_1 and ζ_2 may be correlated, but this is not important to the model discussion.

In practical terms, this specification differs from the one given previously due to the inclusion of this noise term. The distribution of wages faced by any particular individual will have a higher variance and potentially a different mean. This makes any expectation the worker forms about his wage (1) likely to be different from the true mean wage, (2) less responsive to variations in his characteristics, and (3) less stable due to fluctuations in the noise term between periods.

Wage offers to the potential worker are still based on his observed characteristics, X, but these characteristics do not fully determine the distribution. This is due to the presence of information problems. The two most likely sources of information problems are that the worker's characteristics, X, are not well-defined, or that they are no longer fully observable. The former situation would occur if the same X did not give clear information. In the real world sense, this describes a situation in which there is widely varying quality attached to one explicit signal. In the South African education system, as is discussed in section 2.3, a matric certificate is not a clear indicator of ability, or even educational attainment, due to the wide variance in school quality. Thus, faced with a worker with a matric certificate, the employer cannot make an informed assessment of his skills.

The latter scenario describes a situation that is due to poor information flows. Employers may be unable to observe a worker's characteristics simply because they do not observe him. In a labour market with high unemployment, many workers may never be able to reach potential employers due to lack of information about job opportunities, or because the employers have to process such high volumes of applicants. The information problems could thus operate from two directions: individuals could form inaccurate beliefs about what their wage distributions ought to be; or the market could have inaccurate beliefs about what wage the individual should receive.

The important point here is that, as much as we might theorise about the source of the information problems, there must be a substantial component of noise experienced by each individual that is not correlated with their observed characteristics. The net result is that individuals must make a decision about their labour market status based on a 'noisy' wage distribution. If the distortion caused by these information problems is large enough, individuals can allocate themselves into the 'wrong' categories.

This provides the justification for the methodology discussed in the next section. In this scenario, individuals are selected into labour market participation and employment becomes somewhat random. While characteristics still matter, they do not fully determine the questions of participation, employment and wages. In a different world, a slightly different set of individuals might have become employed. The observed wage distribution, and the inequality measures calculated from it, depends on who the employed are, as their characteristics influence not only whether they become employed, but what wage they receive once they are employed. Counterfactual analysis ought to take this noisiness into account, by explicitly allowing for increased randomness in labour force outcomes.

The methodology used in this paper (discussed in detail in section 5) attempts to accommodate the incomplete information by adding disturbances to who gets selected into particular labour market categories. Additionally, it assesses the effect of changes in education on the wage distribution by looking at how counterfactual education levels would have affected inequality, given the disturbed selection probabilities. Adding disturbance terms to these probabilities, by the nature of the exercise, makes the distribution somewhat unstable. Thus Monte Carlo techniques are used to test the stability of the inequality estimates derived under the hypothetical scenarios created by counterfactual analysis.

4. SUMMARY STATISTICS

4.1 Data

This analysis uses two separate surveys, the Project for Statistics on Living Standards and Development (PSLSD) from 1993 and the National Income Dynamics Survey (NIDS) from 2008. Although these datasets were not collected as part of the same project, analyses have shown that they provide roughly comparable data – in terms of information elicited and sample considered – and produce consistent results. Each dataset was collected to form a sample representative of South Africa, using post-sampling weights to correct for sampling error and ensure that the working sample is nationally representative. The PSLSD was conducted in 1993 by the Southern Africa Labour and Development Research Unit, to obtain information on the economic standing and well-being of South Africa, sampling over 40 000 individuals. While not representative for any analysis at the provincial level or for small groups, it is nationally representative and, with use of sampling weights, provides a proportional sample of the national population. It contains information on labour market status, income, expenditures, demographics and household relationships, and is thus ideal for the type of analysis undertaken in this paper. It is also the seminal survey referenced in the literature, as it was a snapshot of a nation about to undertake substantial change. This makes it ideal for comparison with later datasets, to examine the effects of labour market changes during a period of rapid change[42].

The NIDS dataset used is the first wave of a project to collect information on the income, well-being and household formation processes of the South African population. It was collected in 2008 under the auspices of the Southern Africa Labour and Development Research Unit, in two waves (February to July and September). This produced a sample of 7305 households containing a total of 28355 individuals. These households represent all provinces and location types – urban formal, urban informal, metropolitan and rural – in South Africa, as well as all racial groups. The data are not representative at the provincial level, so we cannot perform analyses for the different provinces. As is common in South Africa, the White and Indian/Asian racial groups are under-represented. As these race groups are often correlated with higher incomes, the representation issue may affect the analysis of earnings done in this paper. Even if this makes a substantial difference to to the representivity of earnings, this should result in the measures presented here forming a lower bound

42 South Africa Labour Development Research Unit. (1994). South Africans Rich and Poor: Baseline Household Statistics. University of Cape Town, South Africa.

for inequality measures, as losing part of the top of the distribution will narrow the observed variance. In addition to information on earnings and labour market status, the dataset allows household relationships to be examined; and important component in any analysis of labour market behaviour.[43]

4.2 Demographic characteristics

Table 1 and Table 2 present weighted demographic characteristics of the sample in 1993 and 2008, respectively. Each table contains statistics on three different sections of the population. The first and second column contain numbers and percentages for the sample of individuals who have given basic demographic information – age, sex, race, education levels, location and marital status ("full sample"), and is reported for background information only. The third and fourth columns give information on a smaller sample of individuals for whom we have more data, including labour market status and household information, such as whether they live with one of their biological children ("estimation sample"). This sample will be used for the majority of analyses in this paper. The final sample is restricted to individuals for whom we have all the above information, in addition to labour market earnings ("estimation sample with reported income", or "with income" sample). Obviously, this sample is thus restricted to employed individuals. The initial discussion of earnings distribution uses these individuals, as they are the only group for which we observe earnings. They will also be used to calculate returns to characteristics in Section 5.2.

These tables provide some striking conclusions. In both years, the initial sample contained a majority of women. As soon as the sample was restricted to wage earners, males become the majority. The difference in proportions becomes smaller between 1993 and 2008. Similarly, although population proportions remained quite similar among races between the two years, and Africans remained underrepresented in the employed sample, the difference between the proportions of Africans in the overall and the employed sample became far smaller between 1993 and 2008 – decreasing from almost 20% to 7%.

The distribution of education in the population has changed considerably between the two years. In 1993 in the overall sample, only 12% of the population had completed secondary school or obtained higher qualifications. In 2008, this group represented almost 32% of the population, while another

43 Leibbrandt, M., I. Woolard and L. De Villiers. (2009). Methodology: Report on NIDS Wave 1. Technicial Paper No. 1. July 2009.

40% had incomplete secondary education, compared to 25% in 1993. Those with no education decreased from 25% to 9%, though this is partially attributable to members of this group being predominantly older adults in 1993.

The location of the population has also altered substantially between 1993 and 2008. In 1993, 53% of the sample lived in metropolitan locations, with another 21% located in urban areas. In 2008, 30% lived in metropolitan areas, and 63% in urban regions. Only 7% lived in rural areas in 2008, which seems like a surprisingly low number.

Table 1: Demographic characteristics in 1993

	Full sample N = 43173		Estimation sample N = 17934		"With Income" sample N = 7436	
Male	20931	48.48	8415	46.92	4248	57.13
Female	22242	51.52	9519	53.08	3188	42.87
African	33189	76.88	12592	70.21	4297	57.79
Coloured	3542	8.20	1811	10.10	940	12.65
Asian/Indian	1052	2.44	654	3.65	321	4.31
White	5390	12.48	2877	16.04	1878	25.25
0-16	15400	35.67				
16-19	3737	8.66	2633	14.68	154	2.07
20-24	4458	10.33	3143	17.53	793	10.66
25-34	7004	16.22	4651	25.93	2358	31.71
35-49	6953	16.10	4913	27.40	2986	40.16
50-59	2595	6.01	1858	10.36	974	13.09
60+	3027	7.01	735	4.10	172	2.31
None	10934	25.33	2108	11.75	880	11.84
Incomplete primary	12940	29.97	3335	18.59	1233	16.58
Complete primary	3136	7.26	1564	8.72	555	7.46
Incomplete secondary	10859	25.15	7087	39.52	2499	33.60
Complete secondary	3376	7.82	2443	13.62	1253	16.85
Some tertiary/diploma+incomplete secondary	231	.53	161	.90	84	1.13
Diploma	1208	2.80	879	4.90	645	8.68
Degree	490	1.14	358	1.99	287	3.86
Metro	22715	53	7218	40	2219	30
Urban	8976	21	4329	24	1935	26
Rural	11482	27	6388	36	3282	44

Table 2: Demographic characteristics in 2008

	Full sample N = 18046		Estimation sample N = 11815		"With income" sample N = 3049	
Male	8456	46.86	5297	44.83	2059	53.3
Female	9590	53.14	6518	55.17	1804	46.7
African	13906	77.06	9179	77.69	2718	70.35
Coloured	1658	9.19	1059	8.96	423	10.95
Asian/Indian	517	2.87	381	3.23	182	4.7
White	1965	10.89	1195	10.12	541	14
0-16	520	2.88				
16-19	2328	12.9	1993	16.87	145	3.75
20-24	2610	14.46	1981	16.77	461	11.94
25-34	4522	25.06	2928	24.79	1193	30.89
35-49	4243	23.51	3068	25.96	1437	37.2
50-59	1899	10.52	1389	11.76	534	13.83
60+	1924	10.66	456	3.86	92	2.39
None	1628	9.02	848	7.17	230	5.96
Incomplete primary	2261	12.53	1400	11.85	445	11.52
Complete primary	1151	6.38	714	6.05	215	5.57
Incomplete secondary	7247	40.16	5017	42.46	1238	32.06
Complete secondary	3504	19.42	2357	19.95	893	23.11
Some tertiary/diploma+incomplete secondary	263	1.46	160	1.36	81	2.09
Diploma	1402	7.77	957	8.1	545	14.11
Degree	589	3.26	362	3.06	215	5.57
Metro	5438	30.13	3756	31.89	793	20.53
Urban	11367	62.99	7210	61.23	2700	69.9
Rural	1241	6.88	809	6.87	370	9.57

4.3 Employment, unemployment and participation[44]

Table 3 summarises the employment rates in 1993 and 2008 for the individuals for whom we had data including labour market status, and breaks employment down by gender. The employment rate is defined as the percentage of the population 16 years or older that identify themselves as

44 The results of Sections 4.3 and 4.4 are the same as those found in the initial discussion paper on the NIDS dataset. This gives support to the labour market and earnings analyses in Sections 4 and 6. (Ranchhod, 2009).

employed. The participation rate is the percentage of the population 16 years or older who are employed or unemployed. The individuals in this group – 16 years or older and employed or unemployed – are called the labour force. The unemployment rate is the percentage of the labour force without jobs. In all the analyses below, unemployment is defined broadly – adults who are actively looking for work are unemployed, as are adults who describe themselves as not actively looking for work but willing to accept work if offered.

Table 3: Employment rates in each year

Estimation sample	1993	2008
Employment rate	43.82	43.56
Male	54.15	52.93
Female	34.69	35.87

In both 1993 and 2008, not quite 44% of adults were employed. Very small shifts were observed in employment rates by gender: in 1993, 54% of adult males and 35% of adult females were employed, compared to 53% and 36% in 2008, respectively. Women are thus still far less likely to be employed than are men. Tables 4 and 5 illustrate that this effect is due to lower probabilities of participation and employment, and that the relative impact of these seems to have changed over time.

In 1993, 35% of women active in the labour market were unemployed, compared to 42% in 2008. In the same period, unemployment rates for men decreased from 26% to 24%. Participation rates went up for women (from 62% to 70%) and down for men (73% to 70%). A preliminary explanation for the changes could be that less-employable women were entering the labour market, while less-employable men were exiting it. This will be examined in Section 5.2.

The proportion of prime-age adults (25 to 49) who were unemployed increased from 1993 to 2008, while the proportions of labour market participants in this age group increased. Young adults (16 to 24) represent a larger share of labour market participants in 2008 than in 1993, but also a larger share of the unemployed. Surprisingly, the proportion of adults over 60 participating in the labour market increased substantially between 1993 and 2008, while their unemployment rate increased by a smaller margin. This suggests that the majority of older adults who stayed in the labour force remained employed As this was the same period in which access to the old age pension reached its current levels, this labour market behaviour is presumably not driven by desperation, but is perhaps due to increased demand for experienced workers or pressure from household members wanting to retain a source of income (if wages among these adults are higher than the old age pension).

Participation rates decreased for the extreme ends of the education spectrum, for those with no or incomplete primary education, and for those with degrees. For all other categories, participation rates were higher in 2008 than in 1993. Unemployment rates were lower only for those with incomplete and complete primary education; for all other categories, unemployment rates increased between 1993 and 2008. Participation rates increased for all racial groups, while unemployment rates increased for all races except Africans, who experienced a slight decrease from 40% to 39%.

Table 4: Comparison of unemployment and participation rates in each year

Estimation sample	1993	2008	1993	2008
	Unemployment rate		Participation rates	
Overall	30.15	33.79	62.73	65.87
Male	26.28	24.85	73.44	70.43
Female	34.87	42.12	53.26	61.98
16-19	65.98	59.5	16.16	22.54
20-24	51.02	51.53	53.76	65.76
25 - 34	33.33	36.21	79.87	84.42
35-49	19.62	24.84	80.54	82.21
50-59	16.17	15.35	66.94	61.92
60+	15.07	19.25	29.8	37.46

Table 5: Comparison of unemployment and participation rates in each year

Estimation sample	1993	2008	1993	2008
	Unemployment rate		Participation rates	
By education level:				
None	29.24	31.72	61.69	51.82
Incomplete primary	39.42	34.96	63.16	61.85
Complete primary	40.25	37.43	61.73	62.93
Incomplete secondary	33.08	41.86	55.22	56.19
Complete secondary	24.76	32.82	74.06	79.18
Incomplete tertiary	4.6	25.89	58.36	90.28
Diploma	5.13	19.89	84.57	91.6
Degree	2.45	4.07	89.03	85.56
By race:				
African	40.17	38.93	59.3	63.11
Coloured	20.82	24.24	69.16	74.9
Asian/Indian	10.2	13.62	59.72	75.88
White	4.29	14.9	74.39	75.48

4.4 Earnings and inequality

Tables 6 and 7 show the distributions of labour earnings from different sources in 1993, while tables 8 and 9 give similar information for 2008. Tables 6 and 8 analyse annual earnings and tables 7 and 9 show earnings standardised by hours worked[45]. In all cases, the analysis is restricted to those giving non-zero incomes, and excludes those who claimed to participate in that category of employment but gave no earnings.

In 1993, we have three potential sources of earnings – a regular job, self-employment and casual work. The majority of earners participate in regular work, rather than being self-employed or working casually. Regular jobs produce the highest mean annual income, of R86540, followed by casual work (R18360) and self-employment (R1286). Average total earnings is R70287. Hourly wages appear to be a less reliable measure, probably due to misreporting of hours worked, particularly for those doing casual labour. According to these figures, mean hourly wages are for regular and self-employment work are both around R40, while those for casual work are R3442 – a highly unlikely figure. This results in total hourly income being rather overestimated as well.

In 2008, although earnings from secondary jobs, working in a family garden or kraal and helping someone else with their business are all reported, these groups are negligible, totaling only 113 individuals, compared to the numbers reporting incomes from a primary regular job, self-employment and casual employment. Earnings from a regular job is the largest category, followed by casual employment. A regular job is also the most profitable activity, with mean annual income of over R73000, compared to R58099 from self-employment and only R11219 from casual work. Some individuals report earnings from more than one category, and mean earnings from all sources of income are R64860.

In contrast, highest mean hourly earnings come from self-employment (R52 per hour), suggesting that the self-employed may work for fewer hours than the employed. Those with a regular job earn just short of R42 per hour, while those with casual work earn R34 per hour.

45 For the NIDS data, income categories were converted to point estimates by simple mid-point substitution. For regular earnings (primary and secondary), net incomes were converted to gross incomes using a ratio calculated for each income category and number of deductions. The point estimates for those who gave earnings categories were also converted to gross incomes in this way. The other earnings information seems more readily interpreted as gross income, reading the questionnaire, and is treated as such.

Hourly earnings for the two years are fairly similar, though it seems like the self-employed made substantial gains. This may reflect changes in sampling bias, however. In terms of annual incomes, total income seems to have decreased by over R5000, regular earnings by R13000, casual earnings by R7000, and self-employment has increased by almost a factor of five.

Table 6: Annual labour earnings in 1993

Annual labour earnings: 1993 (inflated to 2008 equivalents)				
Annual	Total	Regular	Self-employment	Casual
mean	70287.1	86540.9	1286.184	18360.05
N	6611	5174	910	655
p1	57.47127	689.6552	12.77139	689.6552
p5	280.9706	3965.517	31.92848	1896.552
p10	1149.425	6896.552	63.85696	2758.621
p25	8620.689	18917.24	130.2682	5344.828
p50	31347.96	43312.64	439.3359	11379.31
p75	71873.56	89655.17	1149.425	20689.66
p90	154827.6	174942.5	3134.797	34482.76
p95	227586.2	255747.1	4789.272	51724.14
p99	462344.8	517241.4	15673.98	137931
min	6.385696	34.48276	6.385696	68.96552
max	1.21E+07	1.21E+07	31347.96	551724.1

Table 7: Hourly labour earnings in 1993

Hourly labour earnings: 1993 (inflated to 2008 equivalents)				
Hourly	Total	Regular	Self-employment	Casual
mean	368.9233	39.9432	39.9011	3442.554
N	6539	5150	869	642
p1	0.290259	0.290259	0.222359	89.65517
p5	1.596424	1.662392	0.615764	280.1724
p10	2.873563	2.971526	1.077586	397.878
p25	8.210181	8.163532	3.412356	788.1774
p50	21.45497	19.45642	10.77586	1551.724
p75	58.14854	41.96056	32.4152	2801.724
p90	279.6826	82.50065	102.6273	5655.173
p95	1537.356	117.5549	191.5709	10344.83
p99	5548.901	232.062	361.8078	39655.17
min	0.013062	0.013062	0.085523	17.24138
max	275894.7	5108.557	1596.424	275862.1

Table 8: Annual labour earnings in 2008

Annual labour earnings: 2008	Total	Primary job	Secondary job	Self-employment	Casual	Family plot	Helping
mean	64860.07	73051.34	14895.72	58098.9	11219.26	7628.143	4366.464
N	3384	2539	31	424	466	33	49
p1	100	3000	2400	40	480	100	360
p5	1440	6000	2400	100	1200	500	480
p10	3600	8640	2880	160	1800	800	840
p25	9748.79	14400	4560	500	3360	1440	1200
p50	23520	30000	6000	3600	7200	3500	3000
p75	72000	96000	12000	24000	14400	6000	4800
p90	168000	176305.2	60000	264000	24000	24000	9600
p95	240000	228000	60000	480000	30000	36000	12000
p99	6.52E+05	687857.1	96000	480000	48000	60000	54000
min	3	480	2400	3	240	100	240
max	1680000	1680000	96000	600000	480000	60000	54000

Table 9: Hourly labour earnings in 2008

Hourly labour earnings: 2008	Total	Primary job	Secondary job	Self-employment	Casual	Family plot	Helping
mean	43.61337	41.82972	28.12676	52.43128	34.11946	16.68191	13.84025
N	2757	2129	27	342	287	21	40
p1	0.933336	1.333338	1.312504	0.466668	0.933336	0.25926	0.500002
p5	2.333341	3.00001	1.555561	0.933336	3.111121	0.77778	1.166671
p10	3.733346	4.083347	2.153853	1.348153	4.666682	1.333338	1.555561
p25	7.000023	7.000023	5.600019	4.666682	7.777804	3.733346	3.733346
p50	16.33339	16.59265	6.222243	14.00005	15.55561	5.833353	4.666682
p75	48.80437	49.5835	11.66671	46.66682	43.5557	10.50004	14.00005
p90	93.33364	90.41697	77.77804	207.4081	70.00024	46.66682	37.33346
p95	178.1376	145.8338	77.77804	233.3341	116.6671	88.66696	37.33346
p99	420.0014	420.0014	373.3346	606.6687	186.6673	88.66696	210.0007
min	0.103704	0.373335	1.312504	0.103704	0.848488	0.25926	0.466668
max	933.3364	816.6694	373.3346	933.3364	700.0023	88.66696	210.0007

Tables 10 and 11 present four inequality measures for each type of earnings in each year. The Gini coefficient measures the overall income distribution in each category, and higher values indicate higher inequality. The 90/10 ratio is the ratio of the earnings of the top 10% of earners in a category to those of the bottom 10%, and is thus more sensitive to changes at the ends of the distribution than is the Gini coefficient. The Atkinson class of measures can be specified to be sensitive to different sections of the income distribution by changing the inequality aversion parameter. Specifically, the closer to zero the inequality aversion parameter chosen is, the greater the weight given to the top end of the distribution, and the larger the parameter, the greater the weight given to the bottom end of the distribution. In this case, two Atkinson measures are reported, with the inequality aversion parameter equal to 1 and to 0.5.[46]

There is relatively little variation in the Gini coefficient for the annual earnings distributions between the two years, except for self-employment income. However, the 90/10 ratio and the Atkinson inequality measures show substantial changes between 1993 and 2008, indicating that the gap between the very rich and the very poor has changed, though the distribution in the middle may have stayed similar.

Total annual earnings appear to have become more equal between 1993 and 2008, using all three measures. Regular earnings also became more equal between 1993 and 2008. Earnings from self-employment activities became less equal, while the evidence on casual earnings is mixed: the Gini and Atkinson (0.5) decreased, while the 90/10 ratio and the Atkinson (1) increased. The Atkinson (0.5) measure is uniformly lower than Atkinson (1) measure, indicating that earnings at the bottom of the distributions are spread very thinly.

Table 10: Inequality in 1993 and 2008: annual earnings

	1993				2008			
	Total	Regular	Self-employment	Casual	Total	Regular	Self-employment	Casual
Gini	0.67	0.62	0.71	0.53	0.67	0.61	0.84	0.51
90/10 Ratio	134.7	25.37	49.09	12.5	46.67	20.41	1650	13.33
Atkinson(1)	0.71	0.56	0.67	0.42	0.65	0.52	0.93	0.93
Atkinson(0.5)	0.42	0.34	0.42	0.25	0.37	0.31	0.65	0.22

46 Deaton, A. (1997). "Welfare, poverty and distribution". Chapter 3 in The analysis of household surveys. Washington: World Bank Publications

Hourly earnings are perhaps less feasible to analyse, due to the problems discussed above. Indeed, the results of the various measures are far less consistent for hourly earnings than for annual earnings. Except for regular income, the Gini coefficients have all decreased, in some cases markedly. The 90/10 ratio has decreased for total and regular earnings, but increased for self-employment and casual earnings. The Atkinson (1) measure has uniformly decreased, as has the Atkinson (0.5) measure. In general, it seems that hourly earnings from the different have become less unequal, contrasting with the results from annual incomes, which suggest that all sources of earnings have become more unequal. Due in part to the problems with the hourly earnings measure, the discussion in Section 6 will focus on annual earnings.

Table 11: Inequality in 1993 and 2008: hourly earnings

	1993				2008			
	Total	Regular	Self-employment	Casual	Total	Regular	Self-employment	Casual
Gini	0.92	0.62	0.74	0.64	0.65	0.64	0.71	0.59
90/10 Ratio	97.33	27.76	95.24	14.21	25	22.14	153.85	15
Atkinson(1)	0.93	0.57	0.74	0.55	0.58	0.56	0.72	0.49
Atkinson(0.5)	0.78	0.35	0.47	0.36	0.35	0.34	0.43	0.29

These summary statistics show a country which is making progress on improving socioeconomic indicators – education has increased, average earnings have risen and inequality has (for the most part) fallen. At the same time, unemployment has risen. This suggests that only part of the population may be benefitting from these improvements. The following section presents the methodology that will be used to examine the effect of some of these changes – in education, in participation and in employment – on the earnings distribution.

5. METHODOLOGY

In traditional models of the labour market, participation and unemployment rates reflect the informed choices of individuals. Each person selects the labour market category that gives him or her the highest utility. In practice, particularly in a labour market such as South Africa, with extremely high numbers of discouraged job-seekers, making this assumption seems particularly unrealistic. Survey respondents often report being unable to obtain jobs despite actively searching. We suspect that many more might prefer to participate, but for unobservable reasons cannot (family dynamics, location, etc.). Thus, there are many individuals who cannot enter their preferred labour market category, for any number of reasons. Though economic models are never perfect descriptions of reality, given South Africa's labour market situation the assumption of complete markets seem especially unrealistic in this case.

That said, any attempt to deal directly with the non-clearing labour market and involuntary unemployment has to make its own assumptions about the nature of the selection into participation and employment. This paper makes the assumption that there is substantial randomness in the labour market, at each stage of selection. The theoretical model presented in Section 3 presented the idea that if this randomness is substantial, individuals may misallocate themselves into labour market categories. Individual A and individual B might be equally likely to be employed, based on their observable characteristics, but one may be employed and the other not due to the effects of incomplete information on their expected wage distributions. In the extreme case, individuals who are more likely to be employed, based on observed characteristics, may be unemployed or out the labour force due to unobservable effects – true randomness, or the frictions that hamper the effective functioning of South African labour markets, either of which could be driving the noise term disturbing the wage distribution. Incomplete information causes individuals to place themselves into sub-optimal categories due to their inability to make accurate assessments of their own expected wages or based on their assessment that employers, also facing incomplete information, will not offer wages high enough to overcome their search costs. To assess the potential impact of incomplete information, noise must be added to the labour market selection and wage offer decisions. Implementing this assumption might add substantial instability to the estimates of the earnings distributions, however. To address this concern, the counterfactual simulations are performed repeatedly, and confidence estimates of the mean inequality measures are reported.

The methodology of this paper takes its inspiration from Bourguignon, Ferreira and Leite's 2008 paper on the decomposition of income changes through the use of counterfactual distributions and Gonzalez-Rozada and Menendez's 2006 paper on Argentinian earnings distributions. The methodology owes more to the techniques used in the latter paper, and aims to separate the contributions of labour force participation, employment, education and returns to individual characteristics to the changes in income inequality observed between 1993 and 2008. Like any decomposition technique, this approach assumes that changes in labour force composition caused by use of counterfactual selection probabilities do not change returns to characteristics observed in reality, either at the selection or remuneration stage of the problem. This amounts to examining partial rather than general equilibrium effects, by assuming away the possibility that the composition of the labour force affects the returns experienced by it. For instance, decreasing the overall level of education in the 2008 labour force (by attributing to it the levels seen in 1993) might change the returns to education – most likely increasing them. This is a standard characteristic of decomposition analysis, and not unique to the method discussed here.

To implement the decomposition techniques necessary for an assessment of contributors to changes in inequality, the distribution of labour income at time t is specified as YL_t. This function has as arguments the participation rate (P_t) and unemployment (U_t) rate at time t, the formal education structure of the relevant population (E_t), and the returns to individual sociodemographic characteristics (R_t). The observed distribution is thus $YL_t(P_t, U_t, E_t, R_t)$ – actual labour income based on the true participation, unemployment and education rates and measured remuneration. Formally, sociodemographic characteristics other than education also appear as arguments in this function, but as the method leaves them unchanged, we will not include them in each reference to $YL_t()$. The heart of the methodology is to replace each of the time t arguments in turn with their corresponding time $t+1$ values, to generate counterfactual distributions. Each counterfactual distribution represents what the distribution of labour market income would have looked like had a given argument X been different. In each case, a particular argument X from the set Z is changed from its value at time t to its value at time $t+1$, and the resulting distribution of labour market earnings – $YL^*_t(Z_t, X_{t+1})$ – is created. Inequality measures for this counterfactual distribution can then be calculated, and these values can be compared to the inequality measures obtained for the measured earnings distribution. The differences between the two sets of measures can be attributed to the change in the argument X between time t and $t+1$.

In this way, it is possible to assess the role that changes in a particular aspect of the labour market may have played in inequality changes between the two time periods. When this is done cumulatively, we obtain a counterfactual distribution $YL^{**}_t(Z_{t+1}, X_{t+1})$. The differences between YL^{**}_t and YL_{t+1} can be attributed to residual effects, assuming that the samples drawn at time *t* and *t+1* are representative of the true populations at each respective time period. In this case, residual effects include changes in the sociodemographic characteristics of the population, as well as true unobservable changes. The exact procedures used for each step of the process are detailed in the following sections. To test the stability of these counterfactual distributions in the presence of partially random selections into employment, each distribution is simulated multiple times using Monte Carlo techniques.

Although this methodology was derived from that used in Gonzalez-Rozada and Menendez, there are three substantial departures from Gonzalez-Rozada and Menendez's approach. First, and most importantly, a slightly different strategy is used to model the stochastic component of labour market outcomes. The original authors used an innovative technique to combine the effects of labour market behaviour with the effects of the wage distribution by perturbing the process of selection into labour market categories, which is discussed in detail below. This paper modifies their approach to correct some important theoretical flaws. This issue is discussed in detail in Section 5.1.1.1, as it has important implications for the stability of the model.

Second, the income distributions are different. Gonzalez-Rozada and Menendez model household income per adult equivalent, which includes all sources of income aggregated at the household level and allocates total household income to members. However, their technique was applied only to earnings, and other income sources were added as exogenous lump sums. Non-labour income is a significant contributor to household income in South Africa, and is believed to have substantial impacts on household formation and possibly labour market behaviour.[47] While there is disagreement on the nature of these impacts, treating household formation as wholly exogenous to labour market outcomes is not a justifiable assumption in the South African context. This paper restricts its analysis to inequality in individual wage earnings, as these are the subject of the technique.

47 Posel, D. J. Fairburn and F. Lund. (2006). Labour migration and households: A reconsideration of the effects of the social pension on labour supply in South Africa. In *Economic Modelling*. Vol 23(5).

Third, this paper extends the analysis of the role of education on changes in inequality. Gonzalez-Rozada and Menendez assess the effect of education on the wage distribution, but do not analyse its role in selection into participation or employment. This paper considers the impact of education on inequality through its influence on employment as well as earnings.

5.1 Microsimulation approach

5.1.1 Microsimulations of labour market effects

The methodology must thus achieve two goals: counterfactual earnings distributions must be created by replacing arguments in the $YL_t(.)$ function with arguments from time $t+1$; and the non-clearing nature of the labour market must be accommodated.

To achieve these aims, a microsimulation approach is used, based on sequential random sampling. This is discussed in detail in the following sections, but a brief overview of the method is given here. For each individual in the sample, their 'score' for each particular category is calculated – the value of the calculated coefficients multiplied by their observed characteristics. For P_t and U_t, these categories refer to labour market status, and consist of out of the labour force, unemployed (using a broad definition), and employed. For E_t, the probability of falling into one of the eight possible education categories[48] is calculated. To allow for individuals ending up in categories that are not utility maximising, random numbers are drawn from a uniform distribution and these, along with the calculated scores, are used to construct probabilities of selection into any given category. Counterfactual distributions are created by shifting each argument from its t values to its $t+1$ values in turn, and assessing the cumulative effect on the earnings distribution.

We initially discuss only how participation, employment and returns are changed. In the second section, the education shifts are explained as well.

5.1.1.1 Participation effect

The logical starting point is to model the effect of changing the first argument, the participation rate. The procedure has been divided into its four component steps, to improve the clarity of the

48 Possible education levels: none; incomplete primary; complete primary; incomplete secondary; complete secondary; some tertiary; diploma; and degree.

explanation. The procedure is explained in detail for the participation effect, and in less detail for the following effects.

(a) Assessing the participation effect by simulating the counterfactual labour force

The aim is to draw a labour force of individuals from the time t sample that is representative of the time $t+1$ sample, to create our first counterfactual distribution $YL^1_t(P_{t+1}, U_t, E_t, R_t)$. The starting point is to obtain counterfactual probabilities of participation for the individuals from time t, as though they were operating in the time $t+1$ labour market. We use a binary logit model to estimate the coefficients associated with the participation decision at $t+1$, β_{t+1}. Employed and unemployed individuals are classed as participating, and individuals who are out the labour force are non-participants. The specification and results of this model are reported in Section 5.2. The counterfactual linear prediction index – the 'score' – of participation for the time t individuals is thus $S^*_p = \beta_{t+1}X_t$. We then calculate the number of people who would be participating in the labour force, if the participation rate observed in time $t+1$ applied to the number of people in the sample at time t, to obtain $N^*_p = N_t \text{ x } P_{t+1}$, where N_t is the sample at time t, P_{t+1} is the participation rate at time $t+1$, and N^*_p is the resulting counterfactual sample size.

We are now ready to implement the sequential random sampling stage of the model. Random numbers are drawn from a standard uniform distribution – U(0,1) – for each member of the time t sample. They are multiplied by the standard deviation of the counterfactual distribution of the participation scores, S^*_p, to give ξ_1, which is then added to each individual's participation score. The random numbers are rescaled in this way to ensure that the stochastic and deterministic components of the prediction have equal variance. This produces a new semi-random linear prediction index, or score, for each individual, which includes both the deterministic elements of observed characteristics and estimated returns to characteristics and the random component of ξ_1. This semi-random score is then used to compute a probability of participation:

$$P^*_p = \frac{\exp\left(\hat{\beta}_{p,t+1}X_t + \xi_1\right)}{\exp\left(\hat{\beta}_{p,t+1}X_t + \xi_1\right) + 1}$$

P^*_p is sorted in descending order and the N^*_p individuals with the highest P^*_p values are then selected as the counterfactual sample of the participating.

The probability of selection is dependent on the probability of participation, which depends on observed characteristics and the preferences they presumably represent. All else being equal, those individuals with the highest calculated counterfactual probabilities of participation will have the highest P^*_p. The random component of selection into participation enters through the random number ξ_1. This step is repeated with a new random number draw for each round of the estimation. This is an attempt to account for the non-clearing nature of the labour market. In any time period, certain individuals will participate and others will not. In most cases, those individuals selected to participate in the counterfactual sample will be those with the highest probability of participation. However, some individuals with high probabilities of participation will not be selected, and some individuals with low probabilities of participation will be selected. As a result of this process, our counterfactual sample will tend to include individuals whose observed choice differs from their assigned status. This element of randomness is necessary to simulate the fact that in reality, some individuals are unable to achieve their desired labour market status due to imperfections in the market. Given that involuntary unemployment is high, we know that some individuals are unable to obtain their most preferred labour market status or identify their true optimal behaviour. Assigning each individual to their observed choice category assumes away this mismatch, by assuming that all individuals would always end up in the labour market category which they are predicted to occupy.

Gonzalez-Rozada and Menendez use a similar method but their approach to modelling randomness has problematic features. They use sequential Poisson sampling to include the random component. Instead of adding a scaled ξ_1 term to the probability score to obtain P^*_p, they calculate it as $P^o_p = \beta_{t+1}X_t$.

The selection term is then found by calculating it as: $\varepsilon_p = \xi_p / P^o_p$. Individuals are ranked in descending order on ε_p, and the N^*_p individuals with the lowest ε_p values are selected as the counterfactual sample of the participating. A similar technique is used for each of the following steps in Gonzalez-Rozada and Menendez's paper.

This raises two problems. First, by the definition of probabilities, P^o_p may at times be very close to zero. This makes ε_p unstable, which is undesirable. With near-zero denominators, the estimates produced may not be consistent or asymptotically normal. If asymptotic results fail, the benefits of performing Monte Carlo simulations is largely lost, and any conclusions drawn from the method are questionable. Second, in certain scenarios, the combination of restricting the sample size with division by the probability will result in nonsensical results[49]. This technique amounts to a sampling

49 This concern was raised by Professor Wittenberg of the University of Cape Town, and the following analysis is his, though all errors should be attributed to the author.

proportional to size technique. Consider the extreme case in which the population is clearly divisible into *n* groups in two time periods. Assume that the probability of participation within each group remains the same across periods, but that the size of the least employable group increases between *t0* and *t1*. Thus the overall probability of participation will fall. If the technique explained in Gonzalez-Rozada and Menendez's paper is implemented, the change in the overall probability of participation will force fewer individuals from *all* categories to be selected, even though the change should occur entirely in the one category.

This is easiest to see through a numerical example. A population in two time periods is shown in the table below:

	T0		T1		
Type	Size	Probability of participation	Size	Probability of participation	Range of distribution
High ability	200	1	200	1	(0, 1)
Medium ability	300	0.5	300	0.5	(0, 2)
Low abilitity	500	0.1	2000	0.1	(0, 10)
Overall	400	0.4	550	0.22	

Using the original method, dividing ξ_p ~U(0,1) by P^o_p essentially creates a uniform distribution for each category, which will range from zero to $1/P^o_p$. This has the effect of allowing the counterfactual semi-random sample of participants to be drawn from all three categories. However, it also has the effect that a change in the overall participation rate is applied proportionately to all categories. This is undesirable, as logically, even with randomness, we would want the probability of participation for the high and medium ability groups to decrease only slightly, and the majority of the change in labour force status to be borne by the low ability group.

This problem occurs every time their procedure is applied (at the employment, wages and education stages). Thus, it was decided to allow the random error term to enter the process differently, in the procedure outlined above and which continues below.

(b) Simulating the counterfactual unemployed sample

Once the counterfactual sample of participating individuals is drawn, we continue the process of obtaining $YL^1(P_{t+1}, U_t, E_t, R_t)$. As suggested by the specification of YL^1, the only rate which is changed to its counterfactual value is participation. The distributions of unemployment, education and returns are left at their time *t* values. However, the process of sequential random sampling continues, as the imperfections in the labour market affect both the participation and employment selection stages.

We calculate N^*_u as the unemployment rate at time *t* applied to the number of participating people in our counterfactual sample drawn in part (a): $N^*_u = N^*_p \times U_t$.

We want to include an element of randomness in the selection of the employed sample, and thus we again use a sequential sampling technique. New random numbers are drawn from a uniform distribution for each member of the time *t* sample. However, in this step, the linear predictions are not counterfactual – they are estimated using time *t* characteristics and returns. This is done so that this first step in the analysis captures only the effects of counterfactual participation. In 5.1.1.2, counterfactual estimates for the probability of employment will be calculated as well.

To obtain these scores, a multinomial logit model is estimated to obtain the effects of various characteristics on working status (employed, unemployed and out the labour force).[50] These coefficients are used to calculate S_e and S_u – the predicted score for each individual for the employment and unemployment categories, respectively. For each individual and each category, a semi-random linear prediction is obtained by adding ξ_2 – a random number scaled by the standard deviation of the score – to the calculated score. The semi-random probability of unemployment for each individual is then calculated using the formula

$$P_u = \frac{\exp\left(\hat{\beta}_{u,t}X_t + \xi_2\right)}{1 + \exp\left(\hat{\beta}_{e,t}X_t + \xi_2\right) + \exp\left(\hat{\beta}_{u,t}X_t + \xi_2\right)}$$

The sample of participating workers is sorted by probability of unemployment from highest to lowest, and the N^*_u individuals with the highest P^*_p values are selected as the counterfactual sample of the unemployed. The remaining fraction of the sample $N^*_e = N^*_p - N^*_u$ constitutes the counterfactual employed.

50 Discussed in Section 5.2.1

(c) Assigning wages to the counterfactual employed sample

In the sample of N^*_e people, there are three potential wage situations: an individual may have been employed and have had earnings which are observed; an individual may have been employed and have had earnings which are not observed; and an individual may not have been employed in the original data. The first category of people can simply be assigned their observed wages. This group constitutes the majority of every employed sample. The latter two categories are treated in the same way, explained below.

In all cases, estimation is done using the natural log of wages, to approximate a normal distribution, and individuals' estimated log wages are then converted back to level wages for the calculation of the inequality statistics.

We need to impute earnings values for these individuals, so that we can include them in the counterfactual earnings distribution. To do this, we use a wage prediction equation that compensates for selection bias and uses the observed wage distribution to predict wages for those without observed wages:

$$\hat{W}_{k,t} = \hat{\beta}_{w,t} Z_{k,t} - \hat{\rho}_{w,t} G_{w,t} + \omega_{k,w,t}$$

where

$$G_{w,t} = \frac{\phi(J(\ddot{\beta}_{e,t} X_t))}{F(\hat{\beta}_{e,t} X_t)}$$

and

$$(\omega_{k,w,t})_{k=1}^{n} = \Omega_{w,t} \sim N(0, \sigma^2_{w,t})$$

Wages for an employed individual k at time t are thus equal to the returns to characteristics β_t, applied to their characteristics $Z_{t,k}$, compensating for the bias introduced by the correlation between the errors for selection and earnings with the term G and with individual error $\omega_{t,k}$. G – the inverse Mills ratio – is obtained from a binary logit regression of employment status conditional on participation for all participating individuals in the observed sample[51]. Using this procedure to correct for selection is not ideal in the absence of an instrument, but is unavoidable due to the lack of candidate instruments in South African data[52]. β_t is obtained from the wage prediction equation

51 Heckman, J.J. (1979). Sample selection bias as a specification error. In *Econometrica*, 47(1).

52 Fortin, N., T. Lemieux and S. Firpo. (2010). Decomposition Methods in Economics. In *Handbook of Labour*

for those for whom wages are observed.[53] The error term $\omega_{t,k}$ is drawn for each individual for whom we need to estimate wages from a standard normal distribution with variance equal to the empirical variance of the errors from the wage prediction equation.

(d) Evaluation of the participation effect

Now every individual in the sample for the counterfactual participation effect has a wage, and we can analyse the distribution of counterfactual labour earnings. Various measures of inequality are then reported. The comparison between $YL^1(P_{t+1}, U_t, E_t, R_t)$ – the counterfactual distribution – and $YL_t(P_t, U_t, E_t, R_t)$ – the observed distribution – allows us to examine how earnings would have been distributed had the individuals in time *t* made participation decisions according to the same rubric as did individuals in time *t+1*.

Because we are adding a component of randomness to the creation of the counterfactuals, there is potential instability in our estimates. To ensure that our analysis is valid, the above procedure is repeated 1000 times, and we examine the means and 95% confidence intervals of the inequality measures.

5.1.1.2 Unemployment effect

The procedure to account for counterfactual unemployment follows much the same steps as those used to examine counterfactual participation. Remember that the counterfactual arguments are examined cumulatively, so we now aim to calculate the earnings distribution $YL^2(P_{t+1}, U_{t+1}, E_t, R_t)$ for the counterfactual employment effect.

The difference in the procedure arises after step (a). Instead of using the probability of being unemployed P_u, we calculate the counterfactual probability of unemployment P^*_u. This is obtained by using the coefficients generated by a multinomial logit of working status in time *t+1* to calculate counterfactual probability scores for each individual and each category.

As laid out in the discussion of the participation effect, semi-random working status scores are created by adding a scaled random variable from the standard uniform distribution – ξ_3 – to the calculated score. These are then used to calculate the counterfactual probability P^*_u.

Economics Volume 4. Eds Orley Ashenfelter and David Card. Amsterdam: Elsevier / North Holland.

53 Discussed in Section 5.2.1

$$P_u^* = \frac{\exp\left(\hat{\beta}_{u,t+1}X_t + \xi_3\right)}{1 + \exp\left(\hat{\beta}_{e,t}X_{t+1} + \xi_3\right) + \exp\left(\hat{\beta}_{u,t+1}X_t + \xi_3\right)}$$

In addition to the time t individual choosing to participate as though they were making the decision in time $t+1$, they become employed as though the selection process was occurring in time $t+1$.

The counterfactual number of the unemployed is calculated as $N^*_u = N^*_p \times U_{t+1}$, where N^*_p is the size of the counterfactual participation sample and U_{t+1} is the unemployment rate at time $t+1$. P^*_u is sorted in descending order, and the N^*_u individuals with the lowest P^*_u values are selected into unemployment. The remainder of the participating sample is the group of employed individuals.

Wages are assigned as in the discussion of the participation effect, and the counterfactual earnings distribution for the employment effect, $YL^2(P_{t+1}, U_{t+1}, E_t, R_t)$, can be examined. As randomness again plays a role in the creation of this distribution, we again examine means and confidence intervals of the inequality measures, rather than simply considering one distribution. Differences between $YL^2(P_{t+1}, U_{t+1}, E_t, R_t)$ and $YL_t(P_t, U_t, E_t, R_t)$ show the effects of time t individuals making their participation and employment decisions under time $t+1$ conditions. Similarly, differences between $YL^2(P_{t+1}, U_{t+1}, E_t, R_t)$ and $YL^1(P_{t+1}, U_t, E_t, R_t)$ show the effect of counterfactual employment alone.

To make these distributions properly comparable, step (a) is not repeated for the second stage of the analysis. Thus, the participating sample is the same in the first and the second stage. The sample will change in each of the 1000 rounds of implementation, due to the randomness introduced by the influence of the ξ_3 term, but the sample of participants within each round is the same for the two stages.

5.1.1.3 Wage effect

The final step of the initial section is to examine the effect of counterfactual wage returns to characteristics on the earnings distribution. To do this, we need to obtain $YL^3(P_{t+1}, U_{t+1}, E_t, R_{t+1})$. The aim is again to produce a distribution that is comparable to $YL^2(P_{t+1}, U_{t+1}, E_t, R_t)$ and $YL^1(P_{t+1}, U_t, E_t, R_t)$, so the same approach is taken: the sample of employed individuals remains the same within each round, but changes between each round of analysis due to changes in the ξ_1, ξ_2 and ξ_3 terms.

The difference arises in the allocation of wages. We again have three categories of employed individuals – those who were employed and whose earnings were observed, those who were employed and whose earnings are not observed, and those who were not employed in the original data. However, in this stage of the analysis, we calculate counterfactual wages for all three groups. Wages for each individual k are equal to the returns to characteristics in time $t+1$, β_{t+1}, applied to their characteristics in time t, $Z_{t,k}$. Selection bias is compensated for with an error selection term G, obtained from the employment selection equation for time $t+1$. The individual error term $\omega_{t+1,k}$ is drawn from a normal distribution with variance equal to the empirical variance of the errors from the wage prediction equation in time $t+1$.

$$\widehat{W}^{*}_{k,t+1} = \hat{\beta}_{w,t+1}Z_{k,t} - \hat{\rho}_{w,t+1}G_{w,t+1} + \omega_{k,w,t+1}$$

where

$$G_{w,t} = \frac{\phi(J(\ddot{\beta}_{e,t}X_t))}{F(\hat{\beta}_{e,t}X_t)}$$

and

$$(\omega_{k,w,t})_{k=1}^{n} = \Omega_{w,t} \sim N(0,\sigma^2_{w,t})$$

Inequality measures are calculated for this earnings distribution, and the procedure is repeated 1000 times to assess stability. The means and confidence intervals of the measures from $YL^3(P_{t+1}, U_{t+1}, E_t, R_{t+1})$ are compared to those obtained from $YL^2(P_{t+1}, U_{t+1}, E_t, R_t)$, $YL^1(P_{t+1}, U_t, E_t, R_t)$ and $YL_t(P_t, U_t, E_t, R_t)$.

5.1.2 Microsimulations including education effects

Gonzalez-Rozada and Menendez present education as an additional argument in the earnings distribution function, and examine the impact of counterfactual education on earnings through education's effect on wages. Given the importance of education in the South African context, it was decided that this paper will extend their analysis by examining counterfactual education's impact on the earnings distribution through its effect on participation and employment, as well as its impact on wages.

Three different approaches to this problem were considered. The first mimics the procedure used thus far, by including a random component to the counterfactual probability of an individual being in any particular education category. This approach is hard to justify theoretically, as the South African education market is extremely stratified, with educational achievement strongly related to parental achievement. This approach, and the corresponding results, are presented in Appendices B and C.

The second and third approach assign individuals counterfactual education levels without the use of randomness, and differ only in how the individuals are assigned to categories. The following section explains the two basic methodologies of creating counterfactual education distributions first, and then goes on to detail how counterfactual education's impact on the three other arguments is estimated.

(a) Simulating the counterfactual education distribution

The basic methodology applied to creating counterfactual education probabilities is the same for both methods. The difference depends on the decision rule used to assign individuals to education categories.

Educational attainment is split into eight mutually exclusive categories: none; incomplete primary; complete primary; incomplete secondary; complete secondary; some tertiary; diploma; and degree. All individuals with one or more degree are included in the degree category, as dividing this group into more specific sets resulted in extremely small cell sizes. To obtain the predicted education level of each individual, we estimate an ordered logit model. This produces individual *Z*-scores equal to $(\beta_{t+1}X_{t+1})$ and general cut-off points – the *Z*-values which indicate a shift from one level of education to the next. For our time *t* individuals, counterfactual *Z*-scores are calculated as $(\beta_{t+1}X_t)$. The counterfactual probability that an individual *k* belongs to any particular education category *j*, for *j* between 1 and 6, is then:

$$P^*_{j,k} = \frac{\exp\left(cutoff_{j+1} - \ddot{\beta}_{t+1}X_t\right)}{1 + \exp\left(cutoff_{j+1} - \hat{\beta}_{t+1}X_t\right)} - \frac{\exp\left(cutoff_j - \ddot{\beta}_{t+1}X_t\right)}{1 + \exp\left(cutoff_j - \hat{\beta}_{t+1}X_t\right)}$$

For *j*=0, it is:

$$P^*_{0,k} = \frac{\exp\left(cutoff_1 - \hat{\beta}_{t+1}X_t\right)}{1 + \exp\left(cutoff_1 - \hat{\beta}_{t+1}X_t\right)}$$

and for *j*=7, it is:

$$P^*_{7,k} = 1 - \frac{\exp\left(cutoff_7 - \hat{\beta}_{t+1}X_t\right)}{1 + \exp\left(cutoff_7 - \hat{\beta}_{t+1}X_t\right)}$$

The counterfactual number of individuals in each education category is calculated as $N^*_j = N^*_s \times E_{j,t+1}$, where N^*_s is the size of the relevant counterfactual sample (whichever one was calculated immediately prior to this step) and $E_{j,t+1}$ is the percentage of the 'matching' population group at time $t+1$ which falls into education category j. The 'matching' population group is the section of the population which meets the same criteria as the counterfactual sample currently in use. For instance, Gonzalez-Rozada and Menendez use this technique at the last stage before wages are calculated. In that case, the counterfactual sample currently in use would be the counterfactual employment sample. The 'matching' population group would be the portion of the population which is employed in time $t+1$, and thus the proportions in each education category, $E_{j,t+1}$, are calculated from the group of individuals employed in $t+1$.

With this information, we can draw the N^*_j individuals with the highest values of $P^*_{j,k}$ for each category j. At this point, we encounter a problem. It is entirely possible that one individual may fall into two (or more) education categories based on this method, if by chance they fall among the group with highest values of $P^*_{j,k}$ for two or more categories j. It is necessary to decide which category the person will be assigned to, in this instance. The first method is to allow individuals to be allocated to the category for which they have the highest $P^*_{j,k}$, regardless of N^*_j. The second method is to give priority to higher categories of education – if an individual would be assigned both to the degree and incomplete tertiary categories, he or she will be assigned to the degree category. Thus in practice, the individuals assigned to category j are the N^*_j individuals with the highest $P^*_{j,k}$ values who have not yet been assigned to an education category. This approach preserves category size, potentially at the expense of allocating individuals to their highest-probability outcome. We use the latter method in generating the primary results, as preservation of category size is important to ensure that summary statistics such as the standard deviation are comparable between the observed and counterfactual distribution. Results from the former method are reported in Appendix C. The following steps were applied to the resulting education distribution from each method to obtain inequality measures.

(b) Applying the counterfactual education distribution to the earnings function's arguments

Changing the education distribution itself will have no effect on the distribution of earnings, as this would not change the observed wages of individuals who are employed. The effect of education on earnings must be assessed by looking at its effect on expected wages and selection into employment.

The direct effect of education on wages, leaving all else constant, would result in an earnings distribution $YL^4(P_t,\ U_t,\ E_{t+1},\ R_t)$. Workers in this distribution are remunerated according to their counterfactual education level, but still at the wage coefficients of time *t*. $YL^5(P_t,\ U_t,\ E_{t+1},\ R^*_{t+1})$ is the earnings distribution obtained from remunerating individuals at time *t+1* returns to other characteristics as well.

However, to assess education properly, it is also necessary to consider its impact on earnings through participation and employment – its indirect effects. In each case, education's effect is assessed by using counterfactual education to calculate an individual's probability of participation or employment. The counterfactual education levels then feed through the rest of the steps, as employment and returns are altered to their *t+1* level.

To assess the effect of education on participation, education is the very first argument to be altered. Once individuals have been assigned counterfactual education levels, the procedure for obtaining the counterfactual earnings distribution for the participation effect proceeds in exactly the same way as was explained in Section 5.1.1.3. The only differences are that through the entire procedure, we use individuals' counterfactual education levels, not their observed education levels, and that one additional earnings distribution is reported is reported. This additional earnings distribution is the equivalent of $YL^4(P_{t+1},\ U_{t+1},\ E_{t+1},\ R_t)$: the earnings distribution obtained once education, participation and employment have been altered to their counterfactual distributions, but before returns have been changed to their *t+1* level. Thus from the participation effect we obtain four earnings distributions: $YL^6(P^*_{t+1},\ U_t,\ E_{t+1},\ R_t)$, $YL^7(P^*_{t+1},\ U^*_{t+1},\ E_{t+1},\ R_t)$, $YL^8_w(P^*_{t+1},\ U^*_{t+1},\ E_{t+1},\ R_t)$ and $YL^9(P^*_{t+1},\ U^*_{t+1},\ E_{t+1},\ R^*_{t+1})$. We refer to arguments as X^*_{t+1} once they have been changed to reflect their own counterfactual level using counterfactual education.

To assess the impact of education on employment, we follow the same procedure. In this case, however, the counterfactual participation sample is drawn using observed education levels. Individuals are assigned their counterfactual education levels before the counterfactual employment sample is drawn, and these counterfactual education levels are used in the calculation of the semi-random probability of employment. This gives rise to the earnings distribution $YL^{10}(P_{t+1},\ U^*_{t+1},\ E_{t+1},\ R_t)$ - earnings for individuals selected into participation on the basis of their time *t* education levels, but selected into employment based on their counterfactual education levels. The next step is to pay these individuals according to their counterfactual education levels, but at time t returns. This produces the distribution $YL^{11}{}_{w}(P_{t+1}, U^*_{t+1}, E_{t+1}, R_t)$. Finally,returns to characteristics (including counterfactual education) are altered to their time $t+1$ levels, and the full effect of counterfactual education is obtained from $YL^{12}(P_{t+1}, U^*_{t+1}, E_{t+1}, R^*_{t+1})$.

Inequality measures are calculated for each of these earnings distribution, and the procedure is repeated 1000 times to assess stability. 95% confidence intervals are reported for each counterfactual earnings distribution. The resulting measures can be compared to those from the other distributions to assess the impact of education has earnings inequality – through its effect on participation and employment, jointly and separately – between time t and time $t+1$.

5.2 Probability calculations

5.2.1 Estimation of individual working status

As is discussed in the methodology section, the simulations rely on the calculation of counterfactual probabilities. To calculate these counterfactual probabilities, it is necessary to estimate equations for labour force status and education level.

Labour force status was estimated both as a binary logit regression ('participating/not participating') and as a multinomial logistic regression, with categories 'unemployed', 'employed' and 'out the labour force'). Separate equations were estimated for 1993 and 2008, for each approach. These regressions were also used to construct a sample selection correction term for the wage equations, as outlined in Section 5.1.1.1.

In both cases, the same group of independent variables was used. These were a quadratic term in age, observed education level, gender, marital status and a set of household relationship variables. A variable to capture joint effects of being female and married was included as well. Education levels are defined to be: none; incomplete primary; complete primary; incomplete secondary; complete secondary; some tertiary; diploma; degree. The set of household relationship variables was included to attempt to control for effects that were dependent on the dynamics of the household. It includes a dummy variable for whether the individual was the household head, whether the individual has a biological child present in the household, whether the individual is the mother of a biological child in the household, and a variable for the working status of other adults in the household. This last variable was coded as 'employed' if any other adult in the household, other than an employee or a lodger, was employed, and 'unemployed' or 'out the labour force' only if no other adult household member was working.

The final three variables, as well as the interaction between gender and marital status, are included because it is hypothesised that these issues may affect an individual's ability or desire to participate in the labour force and obtain employment[54]. Marriage might act as a motivator in the labour market, or result in one or other partner withdrawing. Traditionally, this would be the female partner – thus the interaction term between gender and marital status is included. The presence of a child

54 Casale, D. and D. Posel. (2002). The continued feminisation of the labour force in South Africa: an analysis of recent data and trends. In *The South African Journal of Economics*. 70(1).

might similarly act either as a motivator, or as an incentive to withdraw due to increased demand for within-household labour. Again, this effect may vary by gender. The non-nuclear structure of many South African households motivates the restriction of this variable to biological children – the presence of a grandchild or a niece may well result in different and (hypothesised) attenuated effects compared to that of a child.

The potential for households to include extended families also motivates the definition of the final variable. In many labour market papers, this variable would be spouse's working status. However, South Africans have comparatively low rates of living with marriage partners, which was reflected in the data[55]. This creates a problem: it is likely that the effect of the labour market status of another adult on an individual will vary depending on the relationship between the two, but accounting for this without making major assumptions about the nature of this dependence results in estimation difficulties. Thus, it was decided to treat all adults within a household, provided they are not lodgers or employees, as equal. These categories are ruled out on the grounds that such individuals are highly unlikely to engage in income or labour sharing with the other residents.

In addition, both equations are estimated using a cluster-robust variance estimator that allows for arbitrary correlation in the residuals across residents within each survey cluster. This option was taken instead of including controls for race, location and type of household (rural, urban, metropolitan). The latter two variables are obviously non-varying within cluster. Race was found to be substantially non-varying between clusters – 73% of clusters consist of only one racial group, and 88% of clusters at least 90% of their population drawn from one racial group. Of the clusters that are not racially homogenous, over 80% have at least 75% dominance by one racial group. In light of these results, it was decided to use cluster effects, rather than include race explicitly, as this has the advantage of also controlling for many other unobservable characteristics that individuals within a cluster tend to share – epitomised by labour market characteristics such as information networks, transport, availability of jobs or type of employers.

Finally, all the estimation equations were estimated using the applicable post-sampling weights.

55 Casale, D. and D. Posel. (2002). The continued feminisation of the labour force in South Africa: an analysis of recent data and trends. In *The South African Journal of Economics*. 70(1).

Table 12: Results of the labour force participation prediction equations for both years.

		1993			2008		
		Coefficient	Std Error		Coefficient	Std Error	
Age		0.422	0.014	**	0.515	0.013	**
Age^2		-0.005	0.000	**	-0.007	0.000	**
Education level	None	-1.498	0.220	**	-1.114	0.251	**
Incomplete Primary		-1.151	0.217	**	-0.890	0.250	**
Complete Primary		-1.093	0.231	**	-0.848	0.256	**
Incomplete Secondary		-0.994	0.212	**	-1.143	0.247	**
Complete Secondary		-0.024	0.220		-0.415	0.251	
Some Tertiary		0.234	0.325		-1.508	0.342	**
Diploma		0.551	0.252	*	-0.149	0.265	
Degree					Reference category		
Female		-0.291	0.069	**	-0.271	0.059	**
Married		0.592	0.114	**	0.452	0.097	**
Widowed/Divorced		0.714	0.228	**	0.318	0.303	
Unmarried					Reference category		
Not female					Reference category		
Female*Married		-0.846	0.132	**	-0.859	0.112	
Female*Widowed/Divorced		-0.838	0.246	**	-0.776	0.320	*
Female*Unmarried					Reference category		
Household head		Collinear			0.971	0.096	**
Biological child present		0.260	0.120	*	0.283	0.101	**
Mother of biological child		0.008	0.136		-0.551	0.117	**
Another employed adult		0.596	0.080	**	0.266	0.062	**
Another unemployed adult		0.564	0.081	**	0.401	0.076	**
All other adults out the labour force					Reference category		
Constant		-6.018	0.322	**	-7.296	0.336	**

** Significant at 1%; * Significant at 5%

Note: Robust standard errors are computed assuming observations are independent only between clusters.

Table 13: Results of labour force status prediction equations for both years (probabilities relative to being employed).

	Unemployed						Out the labour force					
	1993			2008			1993			2008		
	β	Std Er		β	Std Er		β	Std Er		β	Std Er	
Age	-0.062	0.016	**	-0.040	0.019	*	-0.553	0.015	**	-0.458	0.016	**
Age^2	0.001	0.000	*	0.000	0.000		0.007	0.000	**	0.006	0.000	**
Education level												
None	2.513	0.363	**	2.373	0.406	**	1.530	0.263	**	1.917	0.235	**
Incomplete Primary	2.638	0.364	**	2.446	0.387	**	1.402	0.259	**	1.603	0.224	**
Complete Primary	2.510	0.374	**	2.190	0.388	**	1.303	0.264	**	1.453	0.243	**
Incomplete Secondary	2.153	0.360	**	2.298	0.379	**	1.431	0.251	**	1.407	0.217	**
Complete Secondary	1.699	0.359	**	2.105	0.378	**	0.506	0.256	*	0.348	0.226	
Some Tertiary	-0.114	0.648		1.836	0.447	**	1.283	0.356	**	0.023	0.334	
Diploma	0.243	0.417		1.471	0.384	**	0.047	0.271		-0.427	0.257	
Degree	Reference category						Reference category					
Female	0.138	0.079		0.607	0.090	**	0.338	0.071	**	0.597	0.081	**
Married	-0.260	0.102	*	-0.622	0.132	**	-0.569	0.107	**	-0.716	0.121	**
Widowed/Divorced	0.148	0.327		-0.393	0.334		-0.390	0.332		-0.787	0.252	**
Unmarried	Reference category						Reference category					
Not female	Reference category						Reference category					
Female*Married	-0.482	0.126	**	0.583	0.145	**	0.643	0.120	**	0.928	0.142	**
Female*Widowed/Divorced	-0.370	0.362		-0.247	0.357		0.638	0.349		0.690	0.265	**
Female*Unmarried	Reference category						Reference category					
Household head	-1.589	0.083	**	(omitted)			-1.401	0.100	**	(omitted)		
Biological child present	-0.261	0.095	**	-0.163	0.144		-0.367	0.105	**	-0.246	0.123	*
Mother of bio. child	0.171	0.131		0.031	0.157		0.600	0.129	**	-0.100	0.144	
Another employed adult	-0.365	0.077	**	-0.414	0.073	**	-0.429	0.076	**	-0.762	0.091	**
Another unemployed adult	0.444	0.083	**	0.303	0.088	**	-0.178	0.076	*	-0.420	0.083	**
All other adults out the labour force	Reference category						Reference category					
Constant	-0.567	0.455		-1.345	0.497	**	8.606	0.373	**	6.989	0.352	**

** Significant at 1%; * Significant at 5%

Note: Robust standard errors are computed assuming observations are independent only between clusters.

The results of the regressions are shown in Table 12. In both years, the probability of labour force participation initially increases with age, before decreasing. The turning points for the two years are different – 39 in 1993, and 38 in 2008 – but not very different. In 2008, individuals with any educational qualification other than a degree were less likely to participate than those with degrees, though the difference was insignificant for a diploma or complete secondary education, while in 1993 having a diploma or incomplete tertiary increased the chance of participation relative to degree-holders. These effects were not significant (neither was the decrease associated with merely having complete tertiary), and may be attributable to the small sample sizes in the higher education categories. In both years, women were less likely to participate than men, with married or widowed women even less likely to participate than their unmarried equivalents. In general, being married (currently or previously) increased the probability of participation. In both years, the presence of a biological child increased the probability of participation for parents of either sex. In 1993, being the mother of a resident child had no statistically significant effect, while in 2008 it had a significant negative effect on participation. Thus, in 2008, the net effect of having a resident child on a woman's participation is indeterminate. In both years, the presence of another adult who was in the labour force increased the probability of participation, though the magnitude of this effect decreased from 1993 to 2008.

These results are mirrored in the multinomial logit equations, presented in Table 13. The only changes are that in 1993, having a degree or a diploma makes no difference between employment and unemployment, and that being the household head is associated with increased probability of employment.

5.2.2 Estimation of wage equations

The simulation methodology used in this paper potentially will allocate individuals who were not observed to be working to the employed sample in some rounds of the procedure. In order to be able to calculate inequality measures for the selected sample in each round, these individuals must be assigned an income. To do this, we need to estimate wages, corrected for selection bias. The selection bias term is obtained from the regressions discussed in the above section. Other variables used are a quadratic expression in age, gender and education level. This equation is estimated for each type of income – regular, casual, self-employment and total – and for each year, with differing selection terms for the two years. The regressions were estimated using a cluster-robust variance estimator that allows for arbitrary correlation in the residents within each survey cluster.

For earnings from a regular job, education level is significant with increasing effect in both years, and being female has a negative effect. In 1993, earnings are weakly increasing with age, but there is no relationship in 2008. Casual labour earnings show few significant associations in either year.

For self-employment income in 1993, gender and education level are significant. Increased education increases income, and being female decreases it. In 2008, gender has a significant negative effect, while education level is jointly significant. In both years, income is insignificantly increasing with age.

Total labour earnings shows much the same patterns as regular labour earnings – not surprisingly, since as we saw in Section 4.3 total earnings are driven by regular labour earnings. In both years education and gender are significant. Increased education is associated with increased earnings, as is being male. Age is significant in 1993 but not in 2008.[56]

56 These patterns are the same for both annual and hourly earnings. Only annual earnings are shown in the in-text tables.

Table 14: Results of the wage prediction equations for all earnings distributions in both years.

Distribution:	Total				Regular			
	1993		2008		1993		2008	
Age	0.05	**	0.02		0.05	**	0.00	
Age^2	0.00	**	0.00		0.00	*	0.00	
None	-2.17	**	-2.43	**	-2.17	**	-2.42	**
Incomplete Primary	-2.14	**	-2.30	**	-2.15	**	-2.17	**
Complete Primary	-1.78	**	-2.06	**	-1.84	**	-2.00	**
Incomp. Secondary	-1.37	**	-1.86	**	-1.36	**	-1.70	**
Complete Secondary	-0.74	**	-1.22	**	-0.77	**	-1.02	**
Some Tertiary	-0.18		-1.39	**	(omitted)		-1.35	**
Diploma	-0.37	**	-0.75	**	-0.45	*	-0.60	**
Degree	(omitted)		(omitted)		-0.09		(omitted)	
Female	-0.26	**	-0.41	**	-0.25	**	-0.40	**
Selection term	-2.06	**	-2.51	**	-1.37	**	-1.91	**
Constant	11.9	**	12.29	**	11.79	**	12.34	**

Distribution:	Casual				Self-employment			
	1993		2008		1993		2008	
Age	0.08	*	0.04		0.02		0.11	
Age^2	0.00	*	0.00		0.00		0.00	
None	-0.83	*	-0.73		-2.24	**	-1.10	*
Incomplete Primary	-0.86	*	-0.61		-2.06	**	-0.69	
Complete Primary	-0.56		-0.46		-1.71	**	-0.23	
Incomp. Secondary	-0.45		-0.24		-1.50	**	0.02	
Complete Secondary	-0.22		0.09		-1.04	*	0.34	
Some Tertiary	(omitted)		0.52		-0.56		(omitted)	
Diploma	-0.03		(omitted)		-0.46		0.90	
Degree	0.99	*	(omitted)		2.20	**	0.99	*
Female	-0.14		-0.30	**	-0.69	**	-0.59	*
Selection term	-0.57		-0.46		-1.76	*	-2.14	
Constant	8.71	**	8.76	**	11.69	**	6.64	*
Coefficients shown. ** – significant at 1%; * – significant at 5%								

5.2.3 Estimation of education levels

To perform the final steps in our analysis, it is necessary to have the probability that each individual falls into each education category, as defined above. The variables used to predict educational attainment are a quadratic term in age, gender, race, location, occupation and parental education.

Expected education level increases with age, but with concavity, in both years. The turning point is

25 in 1993 and 35 in 2008. This is promising, as we would expect more (and older) people to have more education in 2008, as access to education increased over time. Black and coloured individuals have less education than their white and Indian counterparts, in both years, and women have lower education levels than do men, though the gap decreased between 1993 and 2008. Location – both provincial and rural or urban – had significant effects on expected level of education, as did the occupational sector of the individual. Higher levels of parental education were associated with higher levels of individual education.

Table 15a: Cuts – the values at which individuals' predicted education levels switch from one to another

	1993	2008
From none to incomplete primary	-4.38803	-5.0756
From incomplete to complete primary	-2.66364	-3.48925
From complete primary to incomplete secondary	-2.11905	-2.92979
From incomplete to complete secondary	0.120864	-0.78551
From complete secondary to some tertiary	2.462739	0.93808
From some tertiary to diploma	2.672217	1.190944
From diploma to degree	5.498392	3.069369

Table 15b: Results of education level prediction equations for both years

	1993			2008		
	Coefficient	Std Error		Coefficient	Std Error	
Age	0.075	0.044		0.070	0.024	**
Age^2	-0.001	0.001	*	-0.001	0.000	**
African	-0.798	0.260	**	-0.534	0.203	**
Coloured	-0.424	0.240		-0.775	0.210	**
Indian	0.099	0.312		0.035	0.358	
White	(omitted)			(omitted)		
Female	-0.209	0.108		-0.081	0.082	
Metro	-0.550	0.211	**	0.390	0.162	*
Urban	-0.227	0.162		0.947	0.148	**
Rural	(omitted)			(omitted)		
Western Cape	-0.091	0.205		-0.753	0.308	*
Eastern Cape	-0.821	0.580		-0.749	0.286	**
Northern Cape	0.131	0.243		-0.981	0.342	**
Free State	0.010	0.218		-0.830	0.341	*
KZN	-0.076	0.257		-0.901	0.256	**
Northwest	0.540	0.333		-0.794	0.314	*
Gauteng	0.389	0.285		-0.903	0.305	**
Mpumalanga	0.223	0.405		-0.628	0.268	*
Limpopo	(omitted)			(omitted)		
Parental education level						
None	-2.437	0.511	**	-3.006	0.380	**
Incomplete Primary	-1.708	0.507	**	-2.090	0.378	**
Complete Primary	-1.535	0.524	**	-1.786	0.410	**
Incomplete Secondary	-1.495	0.512	**	-1.704	0.373	**
Complete Secondary	-0.730	0.494		-1.517	0.378	**
Some Tertiary	-1.752	0.926		-1.572	0.517	**
Diploma	0.078	0.501		-1.120	0.436	*
Degree	(omitted)			(omitted)		
Sector dummies			**			**
** Significant at 1%; * Significant at 5%						
Note: Robust standard errors are computed assuming observations are independent between clusters.						

6. Results

The simulations discussed in the methodology give rise to several sets of results. The primary results are presented and discussed here, with occasional reference to results given in Appendix C. These results are inequality measures for counterfactual earnings distributions between the observed 1993 and 2008 wage distributions. They must be interpreted in the context of assessing the contributions of labour market composition, wage effects and human capital – education – to changes in inequality. The methodology uses a step-procedure, in which arguments are altered recursively to provide comparisons of various sources of potential change. The starting distribution is in 1993, and as arguments in the selection and earnings functions are changed, the counterfactual distributions move towards the observed distribution of 2008. The key interpretation of these results is whether each change increases or decreases inequality. The counterfactual distributions assume that there are no general equilibrium differences in the wage structure between 1993 and 2008. If such changes have occurred, such as slightly lower returns to education due to higher education levels, then the counterfactual distributions will understate the role of changes in returns to characteristics and overstate the role of changes in characteristics. However, this does not detract from the value of the exercise, which is to test the direction and magnitude of the effect of changes in labour force participation, employment, wages and education on earnings inequality.

There are two important things to bear in mind when interpreting the resulting measures. First, as discussed in Section 5.1.1.1, the major effect of adding a degree of randomness to employment should be seen among individuals who are at the margin of participating in the labour market or being employed. Someone with a high probability of employment will almost certainly not be reassigned to unemployment, and someone who is very unlikely to participate in the labour market should not be reassigned to employment. Although this may occasionally happen, the advantage of performing multiple Monte Carlo simulations is that very unusual occurrences should have little effect on the average. Thus, the effect of changes in participation and employment on earnings inequality should operate through action at the bottom of the wage distribution. Second, the summary statistics reported for the counterfactual distributions provide only limited information on where in the distribution (upper tail, centre, etc.) changes are occurring.

All methods were applied to both annual and hourly earnings, as each unit gives slightly different information about the earnings distribution. Hourly earnings are more appropriate if the focus is returns to characteristics, while annual earnings are more useful if the focus is on welfare. Typically, hours worked will vary quite highly between different individuals, and between different groups. Thus, the distribution of wage per hour may differ substantially from the distribution of total earnings. Wage per hour, as well as hours worked, may give a better idea of the impact of characteristics on labour market outcomes, as these measures provide insight into what a unit of labour from a particular individual is worth to employers, and how many units an employer wants to use. Wage per hour may also better reflect the welfare of workers employed on an hourly basis – casual or part-time workers[57].

Total earnings (whether annual, monthly or weekly) can obscure this issue by conflating wage per hour, hours worked, and non-salary benefits. However, it will give a better measure of welfare, which depends more on income than on hourly wage. The longer the reporting period used, the smoother the measure should be, as it will incorporate things such as benefits in kind or bonuses that are not visible on a monthly basis. Both hourly and annual earnings were created from the data available, which gave respondents the opportunity to provide weekly, monthly or annual earnings as well as average hours worked per week. As the focus of this paper is on earnings inequality with an eye to welfare, the primary results discussed in the paper are those drawn from annual earnings. Although, as noted above, hourly earnings may be the better measure for some types of workers, the way in which earnings were reported in the data means that annualizing earnings should take into account the number of hours worked. The results of hourly earnings are shown in Appendix C.

As is discussed in Section 4.3, four different earnings distributions were created – total earnings, regular income, casual earnings and self-employment income. However, the primary focus in the results section will be on regular and total earnings, with some discussion of casual and self-

57 Card, D. (1999). The causal effect of education on earnings. In *Handbook of Labour Economics Volume 3.* Eds Orley Ashenfelter and David Card. Amsterdam: Elsevier.

employment earnings where needed. Self-employment activities and casual employment make up a relatively small portion of primary (or even secondary) labour activities in South Africa. In 1993, only 14% of earners reported income from self-employment and another 10% reported casual job earnings. In 2008 only 13% of respondents reported any earnings from casual labour, and another 12% reported self-employment earnings. In addition to making up a relatively small segment of the labour market, self-employment earnings are not 'well-behaved' – the distribution is decidedly bimodal in 2008, which suggests that treating all individuals with self-employment earnings as part of the same category is incorrect. Where relevant, these will be discussed, but they will not be systematically presented and are available in Appendix C.

6.1 Inequality analysis with observed education levels

The first table presents inequality measures for the earnings distributions created by the simple Monte Carlo simulations, which correspond to distributions $YL^1(P_{t+1}, U_t, E_t, R_t)$, $YL^2(P_{t+1}, U_{t+1}, E_t, R_t)$ and $YL^3(P_{t+1}, U_{t+1}, E_t, R_{t+1})$ in the methodology section. The first and final columns give the observed values for the inequality measures in 1993 and 2008, respectively. The intervening columns show the average values for the inequality measures for the three rounds of simulation, with standard errors below each mean. The first of these columns presents the effect of adding randomness only to the participation decision. The second column shows the measures for the earnings distribution obtained by adding randomness to both the participation and employment decision, and the third column shows the measures for the distribution obtained by adding randomness to both participation and employment, and remunerating the counterfactual employed sample by 2008 returns to characteristics.

The key advantage of this methodology is that it allows the effect of changes in labour market status to be combined with the effect of changing returns to labour market characteristics in an assessment of inequality. Much has been written on each side of this argument, but this is one of the few approaches that allows the joint effect of changes in participation and employment rates to be combined with changes in earnings. As both of these issues will influence how the distribution of earnings changes, and specifically how inequality changes, this is an important advance.

Table 16: Inequality measures for the counterfactual distributions discussed in 6.1

Standard errors shown in italics.

		Participation	**Employment**	**Price**	
	1993	Total earnings			2008
90/10 Ratio	38.33	35.83	36.26	40.06	46.67
		0.30	*0.65*	*0.76*	
Gini	0.65	0.66	0.66	0.71	0.67
		0.00	*0.00*	*0.00*	
Atkinson(1/2)	0.38	0.38	0.38	0.43	0.37
		0.00	*0.00*	*0.00*	
Atkinson (1)	0.61	0.61	0.61	0.67	0.64
		0.00	*0.00*	*0.00*	
		Regular job earnings			
90/10 Ratio	25.00	26.32	26.87	15.70	19.59
		0.27	*0.42*	*0.21*	
Gini	0.62	0.63	0.63	0.57	0.61
		0.00	*0.00*	*0.00*	
Atkinson(.5)	0.34	0.34	0.34	0.27	0.31
		0.00	*0.00*	*0.00*	
Atkinson (1)	0.56	0.56	0.56	0.45	0.52
		0.00	*0.00*	*0.00*	
		Casual job earnings			
90/10 Ratio	12.50	14.87	15.49	14.48	13.33
		0.10	*0.17*	*0.19*	
Gini	0.53	0.56	0.57	0.55	0.51
		0.00	*0.00*	*0.00*	
Atkinson(.5)	0.25	0.26	0.27	0.25	0.22
		0.00	*0.00*	*0.00*	
Atkinson (1)	0.42	0.44	0.45	0.43	0.39
		0.00	*0.00*	*0.00*	
		Self-employment earnings			
90/10 Ratio	58.00	49.46	45.64	200.63	1500.00
		0.51	*0.56*	*6.05*	
Gini	0.75	0.72	0.70	0.88	0.84
		0.00	*0.00*	*0.00*	
Atkinson(.5)	0.48	0.45	0.42	0.69	0.65
		0.00	*0.00*	*0.01*	
Atkinson (1)	0.72	0.69	0.67	0.90	0.93
		0.00	*0.00*	*0.00*	

The results from the simple Monte Carlo simulations (shown in Table 16) suggest that the effects of changes to status and returns move in opposite directions for all four earnings distributions, but that these changes are not the same for all measures.

The semi-random counterfactual total earnings distributions for the participation and the employment effects exhibit lower inequality than that observed in 1993 or 2008 according to the 90/10 ratio. The mean values for the other three measures are in all cases lower than those seen in 2008, but are either the same or higher (for the Gini) than those seen in 1993. The price effect, which examines counterfactual earnings for the counterfactual employed sample, is in all cases positive – it shows markedly increased inequality. This suggests that for total earnings, changes in the composition of the labour force had unclear effects on inequality, but changes in remuneration unequivocally resulted in increased earnings inequality.

The fact that the Gini measure and the 90/10 ratio show differing directions of change for inequality for the first two columns may reflect the different weights these measures give different sections of the income distribution. The 90/10 ratio, a simple ratio – an unweighted comparison of earners at the top and bottom 10% of the distribution – shows that changes in the composition of the labour force resulted in decreased earnings inequality, while the Gini – which gives most weight to the middle of the distribution – shows increased inequality. This might reflect the withdrawal of some low-income earners from the earning pool or lower earnings at the top end, either of which would decrease the 90/10 ratio, in combination with an increased spread of 'middle class' earnings.

These results stand in stark contrast to those for the regular earnings distribution, in which inequality decreased by all measures between 1993 and 2008, and for which the price effect was uniformly negative. Changes in the composition of the labour force, according to the simulation results, had negligible or positive effects on inequality, while changes in remuneration uniformly and significantly decreased earnings inequality.

The earnings distribution for casual labour exhibits similar patterns to that of regular earnings, with the proviso that the price effect is far smaller, though still significantly negative.

The difference in results between total and regular earnings thus seems to be driven by self-employment earnings. Despite its small fraction of the labour force, the self-employment sector appears to contribute highly to measurements of inequality. However, as discussed above, this result is somewhat unreliable. The distribution of self-employment income does not look particularly normal, suggesting that the variable creation process was flawed or that the distribution is truly bimodal. The latter option is plausible, as the category of the self-employed includes both street vendors and small or medium-sized business owners. If this is the case, the distribution is unsuitable for modelling using a normal distribution, which was used to create representative wages. As self-employment makes up such a small proportion of the sample and is potentially problematic to analyse due to this bimodality, it will not be the focus of discussion.

However, this result is relevant more generally for the study of labour earnings in South Africa. Most analyses (see Section 2) use total earnings. When these are not disaggregated by source, it is possible that authors are missing important differences in the behaviour of different spheres of the labour market. Total earnings' dynamics are considerably different from regular employment earnings, and are dominated disproportionately by the behavior of the earnings of the self-employed. This is despite the fact that the self-employed make up a very small fraction of the labour force. These facts may be the result of the sample used in this paper, and an isolated incidence rather than a trend, but they suggest that caution is warranted. Conclusions drawn from analysis of total earnings may thus draw unwarranted conclusions about the experiences of the majority of South Africans.

For instance, the results of the decomposition of changes in total earnings inequality into labour force composition and wage effects supports the conclusions drawn by Leibbrandt, Lam and Garlick (2010), among others. According to these results, changes in the composition of the labour force would have decreased inequality had the behaviour of remuneration been unchanged, but that changes in wages, without changes in composition, would have caused increased inequality. This is consistent with a situation in which lower-wage individuals would not have been employed based on their counterfactual probability of participation and employment, which would lower the variance of the earnings distribution, but that despite this, wages at the bottom of the distribution declined. This could be driven by marked decreases in the returns to labour market characteristics between 1993 and 2008, both in terms of employment and wages. If characteristics that were marginally valuable in 1993 lost some of their value by 2008, this would explain this movement. One of these characteristics might have been education. The returns to having 'marginal' education levels – some or complete secondary schooling – declined substantially relative to higher levels of education, according to the literature. This could be due to a number of effects: general equilibrium analysis would suggest that if the availability of workers with complete secondary schooling increases, then returns to this characteristic should decrease. Concerns with the 'information quality' of educational qualifications have also been raised, which would be in line with the search model proposed in Section 3 – if a signal that a worker can offer to the market contains less information about his ability, his wage will experience higher variance, leading to higher inequality. This applies even to workers in self-employment, in which outside assessments of ability can be very important to earnings.

However, the results from the analysis of regular earnings alone tell a very different story. The overall implication of these results is that, all else being equal, had only the composition of the labour force changed between 1993 and 2008, inequality in earnings from regular jobs would have increased a little, while if only remuneration patterns had changed, earnings inequality would have decreased markedly. For those employed in regular jobs, the spread of wages decreased, but more individuals at the lower end of the wage distribution became employed. The net effect was a decrease in inequality. Potential explanations for this are somewhat more structural. The effects of counterfactual participation and employment are small compared to those caused by price effects or those in total earnings. This suggests that the dominant theme in the experience of regular earnings was a decrease in earnings inequality. This could be driven by changes in the legal environment governing regular employment, such as the implementation of minimum wage laws which would raise wages at the lower end of the distribution, or a general contraction of the real wage distribution, perhaps attributable to inflationary pressures.

Either way, an important lesson is that analysing the total earnings distribution may obscure differences in the behaviour of different sectors of the labour market. If different types of workers are employed in these sections, as is plausible for self-employment versus regular work[58], conflating wage distributions could result in incorrect welfare analysis.

6.2 Effects of education on inequality

6.2.1 On wages alone:

The second part of the analysis of counterfactual earnings distributions was to assess the impact of changes in the education distribution. The first step in this process is to examine the results of changing the education levels of the individuals selected as the employed sample only prior to the calculation of their wages. Thus, they are selected into participation and employment based on their observed education level, and are assigned a counterfactual education level after these steps occur. This is followed by a calculation of their wages based on this counterfactual educational attainment at 1993 returns, and another calculation of wages using counterfactual education level and 2008 returns. These results will be summarised here and interpreted in Section 6.2.2

The results for total earnings are similar to those found in the discussion of the simple Monte Carlo simulations, which is unsurprising, as only the price effect should change. The earnings distribution using counterfactual education is more equal according to the 90/10 ratio, but less equal by all three other measures. The shift in the inequality measures in the counterfactual earnings distribution (using 2008 wage returns to characteristics) is even greater here, and uniformly positive. In every instance, inequality is higher according to these measures, and higher than the inequality actually observed in 2008 according to the Gini and the two Atkinson measures.

These results are very different from those for the regular earnings distribution, as was the case in the previous section. The impact of semi-random participation and employment is increased or

58 Remember that the distribution of self-employment income is bi-modal. This is attributed to the self-employed consisting of two distinct groups – high earners such as the owners of small or medium business or independent professionals and low earners such as street vendors or micro firms. Regular earners may be another distinct group, with a wage distribution roughly between the first two.

unchanging inequality. Once counterfactual education levels are used, the new wage distribution is more unequal, by some measures, but inequality decreases when 2008 returns are used.

For regular and casual earnings, changing the education distribution increases inequality or leaves it statistically unchanged, dependent on the measure of inequality used. When counterfactual education levels and the 2008 returns to characteristics are used, all the measures indicate a decrease in inequality. By these results, changes in the distribution of education between 1993 and 2008 increased inequality, while changes in remuneration decreased inequality. For total earnings, the effects were the opposite: changes in the distribution of education slightly increased inequality, but the major increase in inequality was driven by changes in remuneration. This result is in line with the studies of the South African labour market discussed in Section 2. The differences between these results are again driven by the peculiar behaviour of self-employment earnings.

6.2.2 On employment and participation

The second step of the analysis of the effect of education on inequality is to examine its impact on the probabilities of participation and employment in the labour market. These results go some way to explaining the slightly unexpected conclusions of the effect of counterfactual education on wages alone. The majority of the discussion of these will be postponed until the next section, to provide a clearer explanation.

When counterfactual education levels are used to obtain individuals' probabilities of employment once they are in the labour market, inequality for all earnings distributions and by all measures increases (see table 18). The effect of counterfactual education on wages for the sample selected into employment based on counterfactual education is to decrease inequality or leave it unchanged, for total and regular earnings.

In line with previous results, once these individuals are remunerated at 2008 levels, inequality in the total earnings distribution increases, and decreases for regular and casual earnings.

Table 17: Inequality measures for the counterfactual distributions discussed in 6.2.1

Standard errors shown in italics.

	Participation	**Employment**	**Wage**	**Price**
	Total earnings			
90/10 Ratio	35.83	35.26	34.00	45.15
	0.31	*0.47*	*0.00*	*0.00*
Gini	0.66	0.66	0.69	0.74
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1/2)	0.38	0.38	0.40	0.47
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1)	0.61	0.61	0.63	0.70
	0.00	*0.00*	*0.00*	*0.00*
	Regular job earnings			
90/10 Ratio	26.31	25.54	25.65	17.84
	0.28	*0.30*	*0.00*	*0.00*
Gini	0.63	0.63	0.65	0.60
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1/2)	0.34	0.34	0.35	0.29
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1)	0.56	0.56	0.57	0.49
	0.00	*0.00*	*0.00*	*0.00*
	Casual job earnings			
90/10 Ratio	14.87	14.46	15.46	14.98
	0.10	*0.10*	*0.00*	*0.00*
Gini	0.56	0.55	0.58	0.55
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1/2)	0.26	0.25	0.27	0.25
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1)	0.44	0.43	0.46	0.44
	0.00	*0.00*	*0.00*	*0.00*
	Self-employment earnings			
90/10 Ratio	49.45	46.63	50.28	266.59
	0.51	*0.50*	*0.00*	*0.00*
Gini	0.72	0.72	0.74	0.90
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1/2)	0.45	0.44	0.47	0.75
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1)	0.69	0.68	0.71	0.93
	0.00	*0.00*	*0.00*	*0.00*

Table 18: Inequality measures for the counterfactual distributions discussed in 6.2.2, arising from counterfactual employment probabilities incorporating counterfactual education.

Standard errors shown in italics.

	Participation	**Employment**	**Wage**	**Price**
	Total earnings			
90/10 Ratio	35.81	40.79	33.45	42.51
	0.31	*0.47*	*0.49*	*0.38*
Gini	0.66	0.68	0.68	0.73
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1/2)	0.38	0.40	0.40	0.45
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1)	0.61	0.64	0.63	0.69
	0.00	*0.00*	*0.00*	*0.00*
	Regular job earnings			
90/10 Ratio	26.31	28.54	24.57	16.38
	0.26	*0.22*	*0.28*	*0.11*
Gini	0.63	0.64	0.64	0.57
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1/2)	0.34	0.35	0.34	0.27
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1)	0.56	0.58	0.56	0.46
	0.00	*0.00*	*0.00*	*0.00*
	Casual job earnings			
90/10 Ratio	14.86	15.05	15.60	14.61
	0.10	*0.12*	*0.11*	*0.11*
Gini	0.56	0.56	0.58	0.55
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1/2)	0.26	0.26	0.27	0.25
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1)	0.44	0.45	0.46	0.43
	0.00	*0.00*	*0.00*	*0.00*
	Self-employment earnings			
90/10 Ratio	49.42	50.57	47.41	242.34
	0.51	*0.50*	*0.45*	*3.83*
Gini	0.72	0.73	0.73	0.90
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1/2)	0.45	0.45	0.45	0.74
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1)	0.69	0.70	0.69	0.92
	0.00	*0.00*	*0.00*	*0.00*

Table 19: Inequality measures for the counterfactual distributions discussed in 6.2.2, arising from counterfactual participation probabilities incorporating counterfactual education.

Standard errors shown in italics.

	Participation	**Employment**	**Wage**	**Price**
	Total earnings			
90/10 Ratio	37.73	47.13	37.49	55.80
	0.52	*0.52*	*0.67*	*0.63*
Gini	0.66	0.70	0.72	0.76
	0.01	*0.00*	*0.00*	*0.00*
Atkinson (1/2)	0.38	0.42	0.44	0.49
	0.01	*0.00*	*0.00*	*0.00*
Atkinson (1)	0.62	0.67	0.67	0.73
	0.01	*0.00*	*0.00*	*0.00*
	Regular job earnings			
90/10 Ratio	27.46	31.60	28.00	22.36
	0.28	*0.33*	*0.35*	*0.12*
Gini	0.63	0.66	0.67	0.62
	0.01	*0.00*	*0.00*	*0.00*
Atkinson (1/2)	0.34	0.38	0.37	0.31
	0.01	*0.00*	*0.00*	*0.00*
Atkinson (1)	0.56	0.60	0.59	0.53
	0.01	*0.00*	*0.00*	*0.00*
	Casual job earnings			
90/10 Ratio	15.29	15.80	17.07	15.94
	0.10	*0.11*	*0.14*	*0.14*
Gini	0.56	0.57	0.60	0.56
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1/2)	0.26	0.27	0.30	0.26
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1)	0.45	0.46	0.49	0.45
	0.00	*0.00*	*0.00*	*0.00*
	Self-employment earnings			
90/10 Ratio	52.18	54.93	61.06	339.57
	0.63	*0.72*	*0.77*	*5.02*
Gini	0.73	0.74	0.75	0.91
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1/2)	0.45	0.47	0.49	0.77
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1)	0.70	0.71	0.73	0.94
	0.00	*0.00*	*0.00*	*0.00*

In table 19, the results for using counterfactual education to predict participation in the labour market, employment for labour market participants, and wages at both 1993 and 2008 remuneration levels are given.

Apart from the 90/10 ratio for total earnings, all the inequality measures indicate that the effect of counterfactual education on inequality via participation is positive or neutral. The inequality measures increase or remain constant between those observed in 1993 and those calculated using counterfactual education and participation, so changes in the returns to characteristics in terms of probability of labour market participation did contribute to increased inequality, but there is very little change between the measures obtained in the simple Monte Carlo results and those acquired using counterfactual education. What changes there are, are in the Atkinson(1/2) measure, which has a very high inequality aversion parameter. This suggests that changes in the education distribution did cause individual participation probabilities to change in such a way to increase inequality, but that this was a relatively minor increase, as the Atkinson(1) measure did not pick up the shift.

When the analysis is continued on the counterfactual probability of employment, based on counterfactual education and the counterfactual sample of labour market participants, inequality increases by all measures and for all earnings distributions.

When the wage distribution is changed to reflect counterfactual education, the change is less clear. The 90/10 ratios for total and regular earnings decrease, but all three other measures for total earnings increase or remain constant. The Gini for regular earnings increases, while the two Atkinson measures decrease. By all measures, inequality in casual earnings increases, as it does for self-employment earnings too.

The results for the regular earnings distribution are interesting. This suggests that the earnings distribution saw increased variance around the centre of the distribution, causing the Gini coefficient to increase, but that the overall levels of inequality did not increase, possibly because the tails were less spread out.

The differences between total and regular earnings were driven by casual and self-employment earnings, which behave similarly for this step – inequality increases by all measures.

Finally, the price effect – the effect of changing the earnings distribution to reflect remuneration in 2008 – is positive for total earnings, negative for regular and casual earnings, and positive for self-employment earnings. Thus, for the total earnings distribution, changes in education between 1993 and 2008 would have resulted in decreased inequality, had returns to characteristics in terms of earnings remained the same. However, they did not – as would be predicted by basic supply theory, even if the nature of the characteristics did not change[59] – and this change drove the increase in inequality observed between the two years. For regular earnings, the effects were different. Changes in education increased inequality through their effect on the probability of employment, and decreased it through their impact on the wage distribution, whether 1993 or 2008 levels of remuneration are considered.

6.2.3 Interpretation of the results from the counterfactual education scenarios

An important starting point for analysis of these scenarios is to reiterate the proviso of the previous section. The behaviour of total and regular earnings are substantially different, and the primary driver of this difference is self-employment income, not aggregation issues.

The three key results of Sections 6.2.1 and 6.2.2 are that using counterfactual education levels to predict employment increases total or regular earnings inequality, that using them to predict participation levels has little to no effect, and that counterfactual education's effect on inequality via wages tends to be negative for regular earnings, but positive for total earnings.

The first interesting result is that obtained from comparing the result of counterfactual education on the 1993 and 2008 wage distribution. Inequality increases when 1993 returns to characteristics are used, but decreases when 2008 returns to characteristics are used. This is

59 For instance, the results discussed in the literature review suggest that the meaning attached to a complete matric in the labour market may have changed between 1993 and 2008.

almost certainly capturing effects that are made clearer when changes induced in employment and participation by counterfactual education are considered.

The changes in total earnings induced by giving individuals their counterfactual education levels are almost uniformly in the direction of increased inequality. Increased levels of education (though a particular individual's own education level might fall in this exercise) resulted in higher inequality in the earnings distribution using 1993 returns to characteristics and even higher inequality using 2008 characteristics. Changing employment probabilities based on counterfactual education also increased inequality using either set of returns. Changing participation probabilities did not have much of a direct effect on inequality, but pushed the inequality measures for the employment effect even higher.

For total earnings, changes in the education distribution appear to have resulted in substantially higher inequality. This is not necessarily a bad result, as this may be driven by changes in employment. Counterfactual education's effect on the wage distribution measured in isolation can be interpreted as the result of the increasing overall education in the population, which was shown in Section 4. Individuals with complete secondary education or higher experience substantially higher wages, so if more of the population is assigned to these categories, earnings inequality will increase. This effect is exaggerated by the use of 2008 returns to characteristics because by 2008 the returns to higher education levels relative to lower levels had become even more extreme[60]. This would present itself in the simulation results as increased inequality. These changes need not be interpreted as the result of lower wages.

The effect of counterfactual education, operating through individuals' participation and employment probabilities, can also be interpreted as the result of higher education levels in the population. Individuals with more education are more employable. Giving marginal-employment individuals their counterfactual education level should have increased their chances of employment. Combined with higher unemployment rates in 2008, this could give the random component of the model more scope for operation. However, the standard errors of the estimates are fairly low; it appears that the inequality measures themselves were not subject to much

60 This was found by several papers discussed in Section 2.

fluctuation, even if there was more churning in the selection of the employed sample. The net result of counterfactual education would have been to shift some of these marginal workers into employment, which would explain the increased inequality in the earnings distribution – they would tend to stretch out the lower tail, as (in the first part of the analysis) they would receive low wages due to their observed education level. Even when their wages are assigned using counterfactual education, they would still tend to receive lower wages due to the influence of their other characteristics, which are also likely to drive their wages down. For the reasons discussed in the above paragraph, using 2008 returns to characteristics will only exacerbate this trend, explaining the increased levels of inequality in the counterfactual distribution.

Regular earnings tell a more nuanced story. Remunerating individuals at their counterfactual education level, but at 1993 wages, leads to increased inequality. When 2008 wages are used, inequality decreases. Using counterfactual education to assign individuals to participation and employment results in higher inequality, but if they are then paid at their new education level, whether at 1993 or 2008 wage rates, inequality decreases. The first effect likely is operating through a similar mechanism to the one discussed for total earnings, in which lower-wage individuals enter the distribution, resulting in a higher spread. The second effect suggests that either individuals at the bottom of the wage distribution were being paid more, based on their characteristics, or individuals in the top and middle were being paid less. This is in line with the fact that observed inequality in regular earnings has fallen between 1993 and 2008.

Based on what is known about the South African economy, this may be surprising. In the face of rising unemployment and high earnings premia on educational attainment, falling inequality in earning is not the expected result. However, there are several plausible scenarios that could be driving this result.

First, this could reflect a change in demand for labour. Possibly individuals who would receive a very low wage, if they were employed, are no longer being employed. There is a large body of literature on skills-biased labour demand in South Africa[61], which concludes that the country has

61 Bhorat, H. (2004). Labour market challenges in the post-apartheid South Africa. In *The South African Journal of Economics* 72(5).
Daniels, R.C. (2007). Skills shortages in South Africa: A literature review. Development Policy Research Unit Working Paper No. 07/121.

been experiencing a shift in labour demand towards more skilled workers and that the worst of the unemployment burden is borne by low or unskilled labour. It is plausible to conclude that including these individuals in the earnings distribution, by changing their employment probabilities, would result in increased earnings inequality, while changing their characteristics would decrease it, as they then benefit from the increased premium on educational attainment.

An alternative explanation is that this is the result of changes in the legislative environment. This period saw the expansion of minimum wage laws and required standards in work environments, which are thought to have substantially raised the non-wage costs of labour[62]. This would make marginally employable individuals less attractive to employers, as the cost to hiring them has increased with no associated increase in productivity. This would result in the same observed changes as the story told in the above paragraph: fewer low-wage individuals hired, resulting in a decrease in the variance of earnings distribution and a concomitant decrease in inequality.

A third interpretation of the results uses the concept of efficiency wages, and fits well with the search model proposed in Section 3 and modelled in Section 5. Incomplete information affects all parties in the labour market: as much as workers will find it difficult to obtain a job offering a wage that fits their skills, employers will find it difficult to identify workers who are suitable for their jobs. Thus, employers may be less willing to take a risk on a worker who may or may not be suitable for the job they need filled, particularly given the non-wage costs associated with hiring and firing mentioned above. This would result in less employment in the lower end of the wage distribution. At the same time, employers may be willing to pay an efficiency wage to their current workers to avoid having to deal with the costs associated with replacing them[63]. This results in decreased inequality as the wage distribution shifts up and the lower tail is thinned or removed.

Oosthuizen, M. (2006). The post-apartheid labour market: 1995 – 2004. Development Policy Research Unit Working Paper No. 06/103.
Pauw, K., M. Oosthuizen. and C. Van der Westhuizen, C. (2006). Graduate unemployment in the face of skills shortages: A labour market paradox. Development Policy Research Unit Working Paper No. 06/114.
Rodrik, D. (2006). Understanding South Africa's Economic Puzzles. CID Working Paper No. 130.

62 Pauw, K., M. Oosthuizen. and C. Van der Westhuizen, C. (2006). Graduate unemployment in the face of skills shortages: A labour market paradox. Development Policy Research Unit Working Paper No. 06/114.

63 Mortensen, D. and C. Pissarides. (1999). New developments in models of search in the labor market. In *Handbook of Labour Economics Volume 3.* Eds Orley Ashenfelter and David Card. Amsterdam: Elsevier

6.3 Sensitivity of results to data decisions

6.3.1 Comparison between hourly and annual earnings

The major difference between using hourly or annual earnings, in terms of results, is that inequality decreases for total and self-employment earnings, by most measures, when hourly earnings are used instead of annual earnings. Self-employment earnings are likely to be driving this change. This suggests that the spread of hours worked changed more than did the spread of earnings per hour, for the self-employed.

The results for total and regular earnings based on hourly measures are in line with those found for other studies: if labour market participants were remunerated at 2008 returns, even with counterfactual education levels, there would be substantially higher inequality than was actually observed. Thus, changes in the characteristics of participants, including fundamental changes in the distribution of education, allowed inequality to be lower than it would otherwise have been. This effect seems to operate primarily through changes in the probability of participation, which would have increased inequality[64].

6.3.2 Comparison between the three methodologies used to obtain counterfactual education

There are two points of interest in a comparison of the results obtained through the three different approaches to creating counterfactual education. First, the two approaches discussed in the methodology produce substantially the same results. These two approaches both use non-random techniques to create counterfactual education, and differ only in whether they give priority to the preservation of category size or to the probability associated with particular education categories. The main set of results uses the former approach, while the results associated with the latter are available in Appendix C. That the results are stable between the two

64 This result is entirely consistent with those found in early decomposition studies, done on labour data from the United States of America by Juhn, Murphy and Pierce (1993).

approaches suggests that which is chosen makes little difference, at least in the South African context. This will not necessarily be the same across all labour markets, however, so any subsequent use of this methodology should test both approaches.

Second, the difference produced by using the third of the educational assignment methods is fairly stable across the simulated earnings distributions: using the method which incorporates some randomness into the allocation of educational levels produces results that show lower inequality than those obtained without randomness. This make a certain amount of sense: educational attainment in South Africa is not random, even in the sense of depending on unobservables. It is highly determined by race, location and parental educational attainment and, by implication, depends less on traits such as ability or ambition than it might in other countries. Thus, including a random term in the allocation of education is less theoretically sound than it might be in other countries, and produces lower inequality estimates by compensating for some of the reinforcing effects of race on education and earnings. By making educational attainment slightly more random, the effect of unobservables, which would normally reinforce the impact of education on wages, as well as determining education, is diluted.

7. CONCLUSION

The results of this paper are interesting for several reasons: they present the findings of a relatively new methodology applied to South African data; they find inconsistencies in the behaviour of different types of income; and they provide insight into the mechanisms influencing change in the inequality of South African earnings distributions, and specifically how changes in education affected inequality over a 15 year period.

The methodology proposed by Gonzalez-Rozada and Menendez is designed to make use of a fact that is typically more of a nuisance in economic analysis – very high unemployment levels. As explained in the search model presented in Section 3, this situation is hypothesised to be both the cause and partially the result of information incompleteness in the labour force. Section 5 explains how this problem is used to allow a combined analysis of changes in labour force outcomes and the wage distribution. The original methodology was highly inventive, but subject to theoretical problems that could cause substantial difficulties in implementation and correct interpretation. A correction is proposed and implemented as an alternative in this paper. Thus, the specific methodology used in this paper is new, to the best knowledge of the author, and certainly has never been applied to the examination of South African labour markets. Its major advantage is that it allows a unified analysis of the two major channels through which the labour market affects inequality – selection into employment and determination of earnings – which must usually be treated separately.

The results of this unified analysis produce an important caveat, even before they are substantially interpreted. The behaviour of the distributions of earnings from all labour market sources and of earnings from regular employment are very different. In many cases, they react to changes in completely opposite ways. This appears to be due to the disproportionate effect of the earnings of the self-employed on total earnings. Although this group of workers makes up less than 15% of earners (14% in 1993 and 12% in 2008), the distribution of earnings is extremely unequal, and thus contributes disproportionately to estimation of overall inequality. This is hypothesised to be due to the bi-modal nature of the distribution of self-employment earnings, which is attributed to the self-employed consisting of two distinct groups with very different earnings expectations. The conclusion drawn from this is that future analyses of the South African labour market should be cautious about *which* earnings distribution they are examining – if total earnings are used, the results may not reflect the reality of the majority of South African earners. This result may be an artifact of the data used in this paper and it certainly should be tested in other data sets before being discarded or accepted.

Analysis of the results of the methodology must thus tell very different stories for the behaviour of earnings from all labour market sources and of earnings from regular employment. The baseline result for the former – total earnings – is that changes in the composition of the labour force between 1993 and 2008 would have resulted in increased inequality, but that changes in the remuneration of the labour force would have caused decreased inequality. These results are in line with those found by other researchers, as discussed in Section 2. This suggests that the recommendations of the previous paragraph are particularly relevant. The observed change was a slight decrease in earnings inequality, so in reality the wage effect seems to have dominated. However, when changes in the education distribution are assessed, and analysis is done to see how these changes affected inequality – through composition or remuneration – it is found that improvements in the education levels of South Africa's labour force would have caused inequality to increase both through changes in the composition of the workforce and through changes in how this workforce is paid for its characteristics (and thus its productivity). By its nature, decomposition analysis using counterfactual earnings distributions cannot control for general equilibrium effects. Thus, the observed decrease in inequality of total earnings can be reconciled with the results of the simulated earnings distribution through reference to changes in the value of workers' characteristics during this period – education in particular, but also characteristics such as race, age and location.

Changes in inequality in the distribution of regular earnings have operated very differently. In general, changes in the composition of the labour force would have resulted in increased inequality, while changes in remuneration associated with workers' characteristics would have caused decreased inequality. The observed change in inequality was negative – inequality in regular earnings fell between 1993 and 2008 – so in reality the latter change dominated. Changes in the distribution of education seem to have reinforced these effects – in general, a labour force with counterfactual education levels experienced increased inequality as a result of employment decisions, but decreased inequality as a result of changes in remuneration. For both regular and total earnings, education had little direct effect on inequality via participation, but it exacerbated the results at the employment stage, suggesting that entry into the labour force and into employment have reinforcing effects on inequality.

These results could be due to several different changes. In general, the conclusion for regular earnings seems to be that there were changes in the labour market that would have pulled more low-wage earners into employment, but that, had they been employed, they would have been paid more. There are several economic interpretations that can be placed on these dynamics, which were discussed in Section 6.2.3. They provide some support for the search model of unemployment proposed in Section 3, but can also be used to support at least two other stories of labour market dynamics, so this is not conclusive evidence one way or the other.

This paper can however draw some firm conclusions. The interaction between regular and total earnings should be examined carefully by future researchers, as this dynamic might not operate as expected. Education has played a significant role in the changes seen in the South African earnings distribution between 1993 and 2008, though what exactly this role was depends on which earnings distribution is considered. Its effect on labour market outcomes was to increase inequality in the counterfactual distributions of both total and regular earnings. However, counterfactual education increased inequality through wage effects for the counterfactual distribution of total earnings, and decreased it for the counterfactual regular earnings distribution.

APPENDIX A: References

1) Altman, M. and D. Lee. (2004). Meeting equity targets: Are there enough graduates?. Employment & Economic Policy Research Programme, Human Sciences Research Council.

2) Altonji, J. and R. Blank. (1999). Race and gender in the labor market. In *Handbook of Labour Economics Volume 3.* Eds Orley Ashenfelter and David Card. Amsterdam: Elsevier.

3) Anderson, K., A. Case and D. Lam. (2001). Causes and Consequences of Schooling Outcomes in South Africa: Evidence from Survey Data. In *Social Dynamics* 27(1).

4) Argent, J. (2009). Household Income: Report on NIDS Wave 1. Technical paper No. 3. July 2009. Available online at: http://www.nids.uct.ac.za/home/index.php?/Nids-Documentation/technical-papers.html

5) Ashenfelter, O., C. Harmon and H. Oosterbeek. (1999). A review of estimates of the schooling/earnings relationship, with tests for publication bias. In *Labour Economics*. 6(4).

6) Ashenfelter, O. and C. Rouse. (1998). Income, Schooling, And Ability: Evidence From A New Sample Of Identical Twins. In *The Quarterly Journal of Economics.* 113(1) .

7) Berndt, E.R. (1991). *The Practice of Econometrics: Classic and Contemporary.* Massachusetts: Addison-Wesley Publishing Company.

8) Bhorat, H. (2000). The impact of trade and structural changes on sectoral employment in South Africa. In *Development Southern Africa* 17(3).

9) Bhorat, H. (2004). Labour market challenges in the post-apartheid South Africa. In *The South African Journal of Economics* 72(5).

10) Birdsall, N. and J. Londono. (1997) Asset Inequality Matters: An Assessment of the World Bank's Approach to Poverty Reduction. In *The American Economic Review* 87(2).

11) Bourguignon, F., F.H.G. Ferreira and P.G. Leite. (2008). Beyond Oaxaca-Blinder: Accounting for differences in household income distributions. In *Journal of Economic Inequality* Vol. 6.

12) Bourguignon, F. (2004). The poverty-growth-inequality triangle. Indian Council for Research on International Economic Relations, New Delhi Working Papers 125, Indian Council for Research on International Economic Relations, New Delhi, India.

13) Branson, N., M. Leibbrandt and T. Zuze. (2009). The demand for teriary education in South Africa. Final report to the Centre for Higher Education Transformation, September 2009.

14) Card, D. (2001). Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems. In Econometrica 69.

15) Card, D. (1999). The causal effect of education on earnings. In *Handbook of Labour Economics Volume 3.* Eds Orley Ashenfelter and David Card. Amsterdam: Elsevier.

16) Casale, D. and D. Posel. (2002). The continued feminisation of the labour force in South Africa: an analysis of recent data and trends. In *The South African Journal of Economics*. 70(1).

17) Case, A. and M. Yogo. (1999). Does School Quality Matter? Returns to education and the characteristics of schools in South Africa. NBER Working Paper 7399, October 1999.

18) Chamberlain, D. and S. van der Berg. (2002). Earnings Functions, Labour Market Discrimination and Quality of Education in South Africa. Stellenbosch Economic Working Papers 2/2002.

19) Daniels, R.C. (2007). Skills shortages in South Africa: A literature review. Development Policy Research Unit Working Paper No. 07/121.

20) Deaton, A. (1997). "Welfare, poverty and distribution". Chapter 3 in The analysis of household surveys. Washington: World Bank Publications

21) Dias, R. and Posel, D. (2007). Unemployment, education and skills constraints in post-apartheid South Africa. Development Policy Research Unit Working Paper No. 07/120.

22) Finn, A., M. Leibbrandt and I. Woolard. (2009). Income & Expenditure Inequality: Analysis of the NIDS Wave 1 Dataset. Discussion Paper No. 5. July 2009. Available online at http://www.nids.uct.ac.za/home/index.php?/Nids-Documentation/discussion-papers.html

23) Fortin, N., T. Lemieux and S. Firpo. (2010). Decomposition Methods in Economics. In *Handbook of Labour Economics Volume 4.* Eds Orley Ashenfelter and David Card. Amsterdam: Elsevier / North Holland.

24) Goldin, C. and L. Katz. (2008). *The Race between Education and Technology.* Cambridge, MA: Harvard University Press.

25) Gonzalez-Rozada, M. and A. Menendez. (2006). Why have urban poverty and income inequality increased so much? Argentina, 1991 – 2001. In *Economic Development and Cultural Change*. 55(1).

26) Gould, E., O. Moav and B. Weinberg. (2001). Precautionary Demand for Education, Inequality, and Technological Progress. In *Journal of Economic Growth*, 6(4).

27) Girdwood, S. and M. Leibbrandt. (2009). Intergenerational mobility: Analysis of the NIDS Wave 1 Dataset. Discussion Paper No. 15. July 2009. Available online at: http://www.nids.uct.ac.za/home/index.php?/Nids-Documentation/discussion-papers.html

28) Heckman, J., L. Lochner and C. Taber. (1998). Explaining Rising Wage Inequality: Explorations with a Dynamic General Equilibrium Model of Labor Earnings with Heterogeneous Agents. In *Review of Economic Dynamics*, 1, 1998.

29) Heckman, J.J. (1979). Sample selection bias as a specification error. In *Econometrica*, 47(1).

30) Hoogeveen, J. and B. Ozler. (2005). Not separate, not equal: Poverty and inequality in post-Apartheid South Africa. William Davidson Institute Working Paper No. 739. University of Michigan Business School.

31) Juhn, C., K.M. Murphy, and B. Pierce. (1993). Wage inequality and the rise in returns to skill. In *The Journal of Political Economy*. 101(3).

32) Keswell, M. and L. Poswell. (2004). Returns to education in South Africa: A retrospective sensitivity analysis of the available evidence. In *The South African Journal of Economics*. 72(4).

33) Kingdon, G. and J. Knight (2001). What have we learnt about unemployment from microdatasets in South Africa? In *Social Dynamics*, 21(1).

34) Klasen, S. (2002). Low Schooling for Girls, Slower Growth for All? Cross-Country Evidence on the Effect of Gender Inequality in Education on Economic Development. *The World Bank Economic Review*, 16(3).

35) Lam, D., C. Ardington and M. Leibbrandt. (2008). Schooling as a lottery: Racial differences in school advancement in Urban South Africa. University of Michigan Population Studies Center Research Report No. 08-632.

36) Lam, D., M. Leibbrandt and J. Garlick. (2010). Investing in human capital to reduce inequality. Discussion paper prepared for Centre for Development Enterprise roundtable discussion, January 2010. Available from the authors on request.

37) Leibbrandt, M., I. Woolard and L. De Villiers. (2009). Methodology: Report on NIDS Wave 1. Technicial Paper No. 1. July 2009. Available online at: http://www.nids.uct.ac.za/home/index.php?/Nids-Documentation/technical-papers.html

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

39) Leibbrandt, M., I. Woolard, H. McEwen and C. Koep. (2010). Inequality outcomes in South Africa. University of Cape Town. Southern Africa Labour and Development Research Unit.

40) Leibbrandt, M., J. Levinsohn and J. McCrary. (2010). Incomes in South Africa after the Fall of Apartheid. *Journal of Globalization and Development* Vol. 1(1) Article 2.

41) McCall, J. J. (1970). Economics of Information and Job Search. In *The Quarterly Journal of Economics.* 84(1)

42) Moll, P. (1998). Primary Schooling, Cognitive Skills and Wages in South Africa. In *Economica*, 65(258).

43) Mortensen, D. and C. Pissarides. (1999). New developments in models of search in the labor market. In *Handbook of Labour Economics Volume 3*. Eds Orley Ashenfelter and David Card. Amsterdam: Elsevier

44) Mwabu, G. and P. Schultz. (1996). Education Returns Across Quantiles of the Wage Function: Alternative Explanations for Returns to Education by Race in South Africa. In *The American Economic Review*, 86(2).

45) Mwabu, G. and T.P. Schultz. (2000). Wage premiums for education and location of South African workers, by gender and race. In *Economic Development and Cultural Change,* 48(2).

46) Oosthuizen, M. (2006). The post-apartheid labour market: 1995 – 2004. Development Policy Research Unit Working Paper No. 06/103.

47) Pauw, K. and L. Edwards. (2005). Evaluating the General Equilibrium Effects of a Wage Subsidy Scheme for South Africa. In *Economic Research Southern Africa*, Working Paper No. 21.

48) Pauw, K., M. Oosthuizen. and C. Van der Westhuizen, C. (2006). Graduate unemployment in the face of skills shortages: A labour market paradox. Development Policy Research Unit Working Paper No. 06/114.

49) Posel, D. J. Fairburn and F. Lund. (2006). Labour migration and households: A reconsideration of the effects of the social pension on labour supply in South Africa. In *Economic Modelling*. Vol 23(5).

50) Ranchhod, V. (2009). Labour Market: Analysis of the NIDS Wave 1 Dataset. Discussion Paper No. 12. July 2009. Available online at: http://www.nids.uct.ac.za/home/index.php?/Nids-Documentation/discussion-papers.html

51) Ranchhod, V. (2006). The Effect of the South African Old Age Pension on Labour Supply of the Elderly. In *The South African Journal of Economics.* Vol. 74(4).

52) Rodrik, D. (2006). Understanding South Africa's Economic Puzzles. CID Working Paper No. 130.

53) Schultz, T.P., and G. Mwabu. (1998). Labour unions and the distribution of wages and employment in South Africa. In *Industrial and Labour Relations,* 51(4).

54) Seekings, J. (2007). Poverty and inequality after Apartheid. Centre for Social Science Research. CSSR Working Paper No. 200.

55) South Africa Labour Development Research Unit. (1994). South Africans Rich and Poor: Baseline Household Statistics. University of Cape Town, South Africa. Available online at: http://www.datafirst.uct.ac.za/catalogue3/index.php/ddibrowser/5

56) Stigler, G.J. (1961). The Economics of Information. In *Journal of Political Economy*. 69(3)

57) Stigler, G.J. (1962). Information in the Labour Market. In *Journal of Political Economy*. 70(5)

58) Woolard, I. and S. Klasen. (2004). Determinants of Income Mobility and Household Poverty Dynamics in South Africa. IZA Discussion Paper 1030.

59) Van der Berg, S. (2008). How effective are poor schools? Poverty and educational outcomes in South Africa. Centre for European Governance and Economic Development Research Discussion Paper 69.

60) Van der Berg, S., R. Burger, M. Louw, R. Burger and D. Yu. (2006) Trends in poverty and inequality since the political transition. Development Policy Research Unit Working Paper 104

61) Van der Berg, S. and R. Burger. (2003). Education and socio-economic differentials: A study of school performance in the Western Cape. Development Policy Research Unit Working Paper No. 03/73.

62) World Development Report 2006, *Equity and Development*, 2007.

63) World Development Report 2007, *Development and the Next Generation*, 2008.

APPENDIX B: *Alternative assignment methodology for counterfactual education level: adding randomness to the distribution of education.*

The basic methodology applied to creating a counterfactual education distribution is similar to the methodology used for counterfactual participation and employment. We will again be estimating counterfactual probabilities, and using sequential Poisson sampling to assess the role of randomness.

Educational attainment is split into eight mutually exclusive categories: none; incomplete primary; complete primary; incomplete secondary; complete secondary; some tertiary; diploma; and degree. All individuals with one or more degree are included in the degree category, as dividing this group into more specific sets resulted in extremely small cell sizes. To obtain the predicted education level of each individual, we estimate an ordered logit model. This produces individual Z-scores equal to $\beta_{t+1}X_{t+1}$ and general cut-off points – the Z-values which indicate a shift from one level of education to the next. For our time t individuals, counterfactual Z-scores are calculated as $\beta_{t+1}X_t$.

We return to our uniform distribution $U(0, 1)$ to draw eight random numbers $\xi_{j,k}$ for each individual k. For each individual, eight selection scores are then calculated, such that we have $S_{j,k} = Z^*_{j,k} + \xi_{j,k}$. The counterfactual probability than an individual k belongs to any particular education category j is then, for j=1,...,6, $P^*_{j,k}$ is

$$P^*_{j,k} = \frac{\exp\left(cutoff_{j+1} - \hat{\beta}_{t+1}X_t + \xi_{j+1,k}\right)}{1 + \exp\left(cutoff_{j+1} - \hat{\beta}_{t+1}X_t + \xi_{j+1,k}\right)} - \frac{\exp\left(cutoff_{j} - \hat{\beta}_{t+1}X_t + \xi_{j,k}\right)}{1 + \exp\left(cutoff_{j} - \hat{\beta}_{t+1}X_t + \xi_{j,k}\right)}$$

and for j=0

$$P^*_{0,k} = \frac{\exp\left(cutoff_{1} - \hat{\beta}_{t+1}X_t + \xi_{1,k}\right)}{1 + \exp\left(cutoff_{1} - \hat{\beta}_{t+1}X_t + \xi_{1,k}\right)}$$

and j=7

$$P^*_{7,k} = 1 - \frac{\exp\left(cutoff_7 - \hat{\beta}_{t+1}X_t + \xi_{7,k}\right)}{1 + \exp\left(cutoff_7 - \hat{\beta}_{t+1}X_t + \xi_{7,k}\right)}$$

The counterfactual number of individuals in each education category is calculated as $N^*_j = N^*_s$ x $E_{j,t+1}$, where N^*_s is the size of the relevant counterfactual sample (whichever one was calculated immediately prior to this step) and $E_{j,t+1}$ is the percentage of the 'matching' population group at time $t+1$ which falls into education category j. The 'matching' population group is the section of the population which meets the same criteria as the counterfactual sample currently in use. For instance, Gonzalez-Rozada and Menendez use this technique at the last stage before wages are calculated. In that case, the counterfactual sample currently in use would be the counterfactual employment sample. The 'matching' population group would be the portion of the population which is employed in time $t+1$, and thus the proportions in each education category, $E_{j,t+1}$, would be calculated from the group of individuals employed in $t+1$.

With this information, we can draw the N^*_j individuals with the highest values of $P^*_{j,k}$ for each category j. At this point, we encounter a problem. It is entirely possible that one individual may fall into two (or more) education categories based on this method, if by chance they fall among the group with highest $P^*_{j,k}$ scores for two or more categories j. It is necessary to decide which category the person will be assigned to, in this instance. A number of options were tried and the problem seems fairly minor – differing approaches make little difference to the final outcomes. The method selected is to give priority to higher categories of education – if an individual would be assigned both to the degree and incomplete tertiary categories, he or she will be assigned to the degree category. Thus in practice, the individuals assigned to category j are the N^*_j individuals with the highest $P^*_{j,k}$ values who have not yet been assigned to an education category. This approach preserves category size, potentially at the expense of allocating individuals to their highest-probability outcome.

We now have a counterfactual education distribution which respects both the distribution of education in time $t+1$ and individuals' counterfactual probabilities of education, in line with Gonzalez-Rozada and Menendez's methodology. We can now use this education distribution to change the earnings distribution, through wages, participation or employment.

APPENDIX C: Extended results

C1: Results from the primary methodology used to construct counterfactual education, for hourly earnings.

Table A1: Baseline results (without changes in education)

		Participation	Employment	Wage	Price	
	1993	Total earnings				2008
90/10 Ratio	34.87	34.33	37.31	25.38		25.00
		0.29	*0.62*	*0.48*	*0.00*	
Gini	0.66	0.66	0.66	0.64		0.65
		0.00	*0.00*	*0.00*	*0.00*	
Atkinson(.5)	0.38	0.37	0.38	0.34		0.35
		0.00	*0.00*	*0.00*	*0.00*	
Atkinson (1)	0.61	0.61	0.61	0.56		0.58
		0.00	*0.00*	*0.00*	*0.00*	
		Regular job earnings				
90/10 Ratio	27.78	28.23	30.64	20.75		21.80
		0.28	*0.55*	*0.34*	*0.00*	
Gini	0.62	0.63	0.64	0.61		0.64
		0.00	*0.00*	*0.00*	*0.00*	
Atkinson(.5)	0.35	0.35	0.35	0.31		0.34
		0.00	*0.00*	*0.00*	*0.00*	
Atkinson (1)	0.57	0.57	0.58	0.52		0.56
		0.00	*0.00*	*0.00*	*0.00*	
		Casual job earnings				
90/10 Ratio	14.08	17.89	20.77	17.04		15.00
		0.15	*0.33*	*0.31*	*0.00*	
Gini	0.64	0.61	0.63	0.56		0.59
		0.00	*0.00*	*0.00*	*0.00*	
Atkinson(.5)	0.37	0.31	0.33	0.26		0.29
		0.00	*0.00*	*0.00*	*0.00*	
Atkinson (1)	0.55	0.51	0.54	0.46		0.49
		0.00	*0.00*	*0.00*	*0.00*	
		Self-employment earnings				
90/10 Ratio	71.43	72.33	72.96	42.23		177.78
		0.82	*1.66*	*1.01*	*0.00*	
Gini	0.73	0.77	0.77	0.70		0.71
		0.00	*0.00*	*0.00*	*0.00*	

Atkinson(.5)	0.45	0.52	0.52	0.41		0.44
		0.00	*0.00*	*0.00*	*0.00*	
Atkinson (1)	0.72	0.76	0.76	0.66		0.72
		0.00	*0.00*	*0.00*	*0.00*	

<u>Table B1: Allowing counterfactual education to affect remuneration only</u>

	Participation	Employment	Wage	Price
	Total earnings			
90/10 Ratio	34.34	34.84	32.16	27.94
	0.29	*0.36*	*0.00*	*0.00*
Gini	0.66	0.66	0.67	0.66
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1/2)	0.37	0.38	0.38	0.37
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1)	0.61	0.61	0.61	0.59
	0.00	*0.00*	*0.00*	*0.00*
	Regular job earnings			
90/10 Ratio	28.23	28.21	26.60	22.57
	0.28	*0.21*	*0.00*	*0.00*
Gini	0.63	0.64	0.64	0.63
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1/2)	0.35	0.35	0.35	0.33
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1)	0.57	0.57	0.57	0.54
	0.00	*0.00*	*0.00*	*0.00*
	Casual job earnings			
90/10 Ratio	17.89	17.58	18.15	18.46
	0.15	*0.14*	*0.00*	*0.00*
Gini	0.61	0.60	0.61	0.58
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1/2)	0.31	0.31	0.31	0.28
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1)	0.51	0.50	0.51	0.48
	0.00	*0.00*	*0.00*	*0.00*
	Self-employment earnings			
90/10 Ratio	72.34	69.63	72.24	41.80
	0.82	*0.80*	*0.00*	*0.00*
Gini	0.77	0.77	0.78	0.69
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1/2)	0.52	0.52	0.53	0.40
	0.00	*0.00*	*0.00*	*0.00*

Atkinson (1)	0.76	0.76	0.77	0.65
	0.00	*0.00*	*0.00*	*0.00*

Table C1: Allowing counterfactual education to affect employment and remuneration.

	Participation	Employment	Wage	Price
	Total earnings			
90/10 Ratio	34.34	38.19	31.77	26.33
	0.30	*0.33*	*0.43*	*0.23*
Gini	0.66	0.68	0.67	0.64
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1/2)	0.37	0.40	0.38	0.34
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1)	0.61	0.64	0.61	0.57
	0.00	*0.00*	*0.00*	*0.00*
	Regular job earnings			
90/10 Ratio	28.23	30.78	26.31	21.26
	0.27	*0.33*	*0.30*	*0.14*
Gini	0.63	0.65	0.64	0.61
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1/2)	0.35	0.36	0.34	0.31
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1)	0.57	0.59	0.56	0.51
	0.00	*0.00*	*0.00*	*0.00*
	Casual job earnings			
90/10 Ratio	17.89	18.22	18.71	17.65
	0.15	*0.12*	*0.14*	*0.14*
Gini	0.61	0.61	0.62	0.57
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1/2)	0.31	0.31	0.32	0.27
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1)	0.51	0.51	0.52	0.46
	0.00	*0.00*	*0.00*	*0.00*
	Self-employment earnings			
90/10 Ratio	72.36	70.36	71.27	41.07
	0.83	*0.85*	*0.79*	*0.44*
Gini	0.77	0.78	0.78	0.69
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1/2)	0.52	0.53	0.52	0.40
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1)	0.76	0.77	0.76	0.64
	0.00	*0.00*	*0.00*	*0.00*

Table D1: Allowing counterfactual education to affect participation, employment and wages.

	Participation	Employment	Wage	Price
	Total earnings			
90/10 Ratio	36.79	44.78	34.70	35.52
	0.28	*0.46*	*0.57*	*0.30*
Gini	0.66	0.70	0.70	0.68
	0.01	*0.00*	*0.00*	*0.00*
Atkinson (1/2)	0.38	0.42	0.41	0.39
	0.01	*0.00*	*0.00*	*0.00*
Atkinson (1)	0.61	0.66	0.64	0.63
	0.01	*0.00*	*0.00*	*0.00*
	Regular job earnings			
90/10 Ratio	29.64	35.54	29.06	29.38
	0.22	*0.42*	*0.29*	*0.26*
Gini	0.63	0.67	0.67	0.65
	0.01	*0.00*	*0.00*	*0.00*
Atkinson (1/2)	0.35	0.38	0.37	0.35
	0.01	*0.00*	*0.00*	*0.00*
Atkinson (1)	0.57	0.61	0.60	0.58
	0.01	*0.00*	*0.00*	*0.00*
	Casual job earnings			
90/10 Ratio	18.31	18.70	20.36	20.27
	0.16	*0.15*	*0.19*	*0.18*
Gini	0.61	0.62	0.64	0.59
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1/2)	0.32	0.32	0.34	0.29
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1)	0.52	0.52	0.54	0.50
	0.00	*0.00*	*0.00*	*0.00*
	Self-employment earnings			
90/10 Ratio	74.96	78.36	87.36	41.98
	0.84	*0.85*	*1.15*	*0.42*
Gini	0.78	0.79	0.80	0.69
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1/2)	0.53	0.55	0.55	0.40
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1)	0.77	0.79	0.79	0.65
	0.00	*0.00*	*0.00*	*0.00*

C2: Results from the secondary methodology used to construct counterfactual education probabilities, as discussed in 5.1.2

For annual earnings:

Table A2: Allowing counterfactual education to affect remuneration only.

	Participation	Employment	Wage	Price
	Total earnings			
90/10 Ratio	35.84	35.24	32.75	47.65
	0.30	*0.48*	*0.00*	*0.00*
Gini	0.66	0.66	0.68	0.74
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1/2)	0.38	0.38	0.39	0.47
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1)	0.61	0.61	0.62	0.71
	0.00	*0.00*	*0.00*	*0.00*
	Regular job earnings			
90/10 Ratio	26.31	25.54	23.45	17.66
	0.26	*0.30*	*0.00*	*0.00*
Gini	0.63	0.63	0.63	0.59
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1/2)	0.34	0.34	0.33	0.28
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1)	0.56	0.56	0.54	0.48
	0.00	*0.00*	*0.00*	*0.00*
	Casual job earnings			
90/10 Ratio	14.86	14.46	16.32	14.55
	0.10	*0.10*	*0.00*	*0.00*
Gini	0.56	0.55	0.59	0.55
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1/2)	0.26	0.25	0.28	0.25
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1)	0.44	0.43	0.48	0.43
	0.00	*0.00*	*0.00*	*0.00*
	Self-employment earnings			
90/10 Ratio	49.43	46.60	50.18	275.55
	0.50	*0.51*	*0.00*	*0.00*
Gini	0.72	0.72	0.73	0.90
	0.00	*0.00*	*0.00*	*0.00*

Atkinson (1/2)	0.45	0.44	0.46	0.75
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1)	0.69	0.68	0.70	0.93
	0.00	*0.00*	*0.00*	*0.00*

Table B2: Allowing counterfactual education to affect employment and remuneration.

	Participation	Employment	Wage	Price
	Total earnings			
90/10 Ratio	35.86	40.40	31.46	44.23
	0.30	*0.45*	*0.41*	*0.45*
Gini	0.66	0.68	0.67	0.72
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1/2)	0.38	0.40	0.38	0.45
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1)	0.61	0.64	0.61	0.68
	0.00	*0.00*	*0.00*	*0.00*
	Regular job earnings			
90/10 Ratio	26.33	28.47	22.32	15.28
	0.26	*0.21*	*0.23*	*0.10*
Gini	0.63	0.64	0.62	0.56
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1/2)	0.34	0.35	0.32	0.26
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1)	0.56	0.58	0.53	0.44
	0.00	*0.00*	*0.00*	*0.00*
	Casual job earnings			
90/10 Ratio	14.87	15.07	16.78	14.19
	0.10	*0.12*	*0.09*	*0.10*
Gini	0.56	0.56	0.58	0.54
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1/2)	0.26	0.26	0.28	0.24
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1)	0.44	0.45	0.47	0.42
	0.00	*0.00*	*0.00*	*0.00*
	Self-employment earnings			
90/10 Ratio	49.44	50.67	46.27	236.60
	0.50	*0.48*	*0.51*	*3.16*
Gini	0.72	0.73	0.72	0.89
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1/2)	0.45	0.45	0.44	0.73
	0.00	*0.00*	*0.00*	*0.00*

Atkinson (1)	0.69	0.70	0.68	0.92
	0.00	*0.00*	*0.00*	*0.00*

Table C3: Allowing counterfactual education to affect participation, employment and wages.

	Participation	Employment	Wage	Price
	Total earnings			
90/10 Ratio	36.55	48.30	36.04	62.33
	0.37	*0.43*	*0.48*	*0.72*
Gini	0.66	0.70	0.71	0.76
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1/2)	0.38	0.42	0.42	0.49
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1)	0.62	0.67	0.66	0.74
	0.00	*0.00*	*0.00*	*0.00*
	Regular job earnings			
90/10 Ratio	27.29	32.07	25.78	23.24
	0.22	*0.36*	*0.31*	*0.16*
Gini	0.63	0.66	0.65	0.61
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1/2)	0.34	0.37	0.35	0.31
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1)	0.57	0.60	0.57	0.52
	0.00	*0.00*	*0.00*	*0.00*
	Casual job earnings			
90/10 Ratio	15.17	15.85	19.08	15.68
	0.11	*0.10*	*0.18*	*0.12*
Gini	0.56	0.57	0.61	0.56
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1/2)	0.26	0.27	0.31	0.26
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1)	0.45	0.46	0.51	0.44
	0.00	*0.00*	*0.00*	*0.00*
	Self-employment earnings			
90/10 Ratio	51.68	55.35	61.63	365.73
	0.54	*0.66*	*0.77*	*6.05*
Gini	0.73	0.74	0.75	0.91
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1/2)	0.45	0.47	0.48	0.76
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1)	0.70	0.71	0.73	0.94
	0.00	*0.00*	*0.00*	*0.00*

For hourly earnings:

Table D2: Allowing counterfactual education to affect remuneration only.

	Participation	Employment	Wage	Price
	Total earnings			
90/10 Ratio	34.34	34.83	31.53	29.14
	0.31	*0.35*	*0.00*	*0.00*
Gini	0.66	0.66	0.67	0.66
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1/2)	0.37	0.38	0.37	0.37
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1)	0.61	0.61	0.60	0.59
	0.00	*0.00*	*0.00*	*0.00*
	Regular job earnings			
90/10 Ratio	28.23	28.20	25.06	23.06
	0.28	*0.21*	*0.00*	*0.00*
Gini	0.63	0.64	0.63	0.62
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1/2)	0.35	0.35	0.33	0.32
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1)	0.57	0.57	0.55	0.54
	0.00	*0.00*	*0.00*	*0.00*
	Casual job earnings			
90/10 Ratio	17.89	17.57	19.52	18.15
	0.15	*0.14*	*0.00*	*0.00*
Gini	0.61	0.60	0.63	0.58
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1/2)	0.31	0.31	0.33	0.27
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1)	0.51	0.50	0.53	0.47
	0.00	*0.00*	*0.00*	*0.00*
	Self-employment earnings			
90/10 Ratio	72.34	69.63	72.67	41.85
	0.80	*0.78*	*0.00*	*0.00*
Gini	0.77	0.77	0.78	0.69
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1/2)	0.52	0.52	0.54	0.40
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1)	0.76	0.76	0.77	0.65
	0.00	*0.00*	*0.00*	*0.00*

Table E2: Allowing counterfactual education to affect employment and remuneration.

	Participation	Employment	Wage	Price
	Total earnings			
90/10 Ratio	34.35	37.97	31.39	25.37
	0.31	*0.29*	*0.40*	*0.23*
Gini	0.66	0.68	0.66	0.64
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1/2)	0.37	0.40	0.37	0.34
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1)	0.61	0.63	0.60	0.56
	0.00	*0.00*	*0.00*	*0.00*
	Regular job earnings			
90/10 Ratio	28.24	30.65	24.62	19.78
	0.28	*0.30*	*0.24*	*0.17*
Gini	0.63	0.65	0.63	0.60
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1/2)	0.35	0.36	0.32	0.29
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1)	0.57	0.59	0.54	0.50
	0.00	*0.00*	*0.00*	*0.00*
	Casual job earnings			
90/10 Ratio	17.90	18.32	20.56	16.62
	0.15	*0.10*	*0.14*	*0.12*
Gini	0.61	0.61	0.63	0.56
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1/2)	0.31	0.31	0.33	0.26
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1)	0.51	0.51	0.54	0.45
	0.00	*0.00*	*0.00*	*0.00*
	Self-employment earnings			
90/10 Ratio	72.38	70.66	69.21	41.72
	0.80	*0.85*	*0.68*	*0.41*
Gini	0.77	0.78	0.78	0.69
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1/2)	0.52	0.53	0.52	0.40
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1)	0.76	0.77	0.76	0.65
	0.00	*0.00*	*0.00*	*0.00*

Table F2: Allowing counterfactual education to affect participation, employment and wages.

	Participation	Employment	Wage	Price
	Total earnings			
90/10 Ratio	36.26	45.70	34.47	38.06
	0.40	*0.43*	*0.54*	*0.31*
Gini	0.66	0.70	0.69	0.68
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1/2)	0.38	0.42	0.40	0.39
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1)	0.62	0.66	0.63	0.63
	0.00	*0.00*	*0.00*	*0.00*
	Regular job earnings			
90/10 Ratio	29.26	35.94	28.19	30.75
	0.27	*0.39*	*0.37*	*0.25*
Gini	0.64	0.67	0.66	0.65
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1/2)	0.35	0.38	0.36	0.35
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1)	0.58	0.61	0.58	0.58
	0.00	*0.00*	*0.00*	*0.00*
	Casual job earnings			
90/10 Ratio	18.35	18.68	23.66	19.77
	0.18	*0.15*	*0.18*	*0.13*
Gini	0.61	0.62	0.65	0.59
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1/2)	0.31	0.32	0.36	0.29
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1)	0.51	0.52	0.57	0.49
	0.00	*0.00*	*0.00*	*0.00*
	Self-employment earnings			
90/10 Ratio	75.36	78.17	89.35	43.18
	0.95	*0.86*	*0.92*	*0.47*
Gini	0.78	0.79	0.80	0.70
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1/2)	0.53	0.55	0.56	0.41
	0.00	*0.00*	*0.00*	*0.00*
Atkinson (1)	0.77	0.79	0.80	0.65
	0.00	*0.00*	*0.00*	*0.00*

C3: Results from use of semi-random counterfactual education, as discussed in Appendix B

For annual earnings:

Table A3: Allowing counterfactual education to affect remuneration only

		Participation	Employment	Wage	Price	
	1993	Total earnings				2008
90/10 Ratio	38.33	35.82	35.26	34.88	45.15	46.67
		0.32	*0.48*	*1.42*	*1.23*	
Gini	0.65	0.66	0.66	0.68	0.73	0.67
		0.00	*0.00*	*0.01*	*0.01*	
Atkinson (1/2)	0.38	0.38	0.38	0.39	0.45	0.37
		0.00	*0.00*	*0.01*	*0.01*	
Atkinson (1)	0.61	0.61	0.61	0.62	0.69	0.64
		0.00	*0.00*	*0.01*	*0.01*	
		Regular job earnings				
90/10 Ratio	25.00	26.31	25.54	26.49	18.19	19.59
		0.29	*0.31*	*0.89*	*0.42*	
Gini	0.62	0.63	0.63	0.64	0.58	0.61
		0.00	*0.00*	*0.01*	*0.00*	
Atkinson (1/2)	0.34	0.34	0.34	0.34	0.28	0.31
		0.00	*0.00*	*0.01*	*0.00*	
Atkinson (1)	0.56	0.56	0.56	0.56	0.47	0.52
		0.00	*0.00*	*0.01*	*0.00*	
		Casual job earnings				
90/10 Ratio	12.50	14.86	14.45	15.85	14.95	13.33
		0.10	*0.10*	*0.30*	*0.29*	
Gini	0.53	0.56	0.55	0.57	0.55	0.51
		0.00	*0.00*	*0.00*	*0.00*	
Atkinson (1/2)	0.25	0.26	0.25	0.27	0.25	0.22
		0.00	*0.00*	*0.00*	*0.00*	
Atkinson (1)	0.42	0.44	0.43	0.46	0.43	0.39
		0.00	*0.00*	*0.01*	*0.00*	
		Self-employment earnings				
90/10 Ratio	58.00	49.43	46.59	50.31	246.72	1500.00
		0.52	*0.52*	*1.49*	*8.72*	
Gini	0.75	0.72	0.72	0.72	0.89	0.84
		0.00	*0.00*	*0.01*	*0.01*	

Atkinson (1/2)	0.48	0.45	0.44	0.45	0.73	0.65
		0.00	*0.00*	*0.01*	*0.02*	
Atkinson (1)	0.72	0.69	0.68	0.69	0.92	0.93
		0.00	*0.00*	*0.01*	*0.01*	

Table B3: Allowing counterfactual education to affect employment and remuneration.

		Participation	Employment	Wage	Price	
	1993	Total earnings				2008
90/10 Ratio	38.33	35.85	36.10	33.28	42.94	46.67
		0.29	*0.59*	*0.57*	*0.96*	
Gini	0.65	0.66	0.67	0.66	0.72	0.67
		0.00	*0.00*	*0.00*	*0.00*	
Atkinson (1/2)	0.38	0.38	0.39	0.37	0.44	0.37
		0.00	*0.00*	*0.00*	*0.01*	
Atkinson (1)	0.61	0.61	0.62	0.60	0.68	0.64
		0.00	*0.00*	*0.00*	*0.01*	
		Regular job earnings				
90/10 Ratio	25.00	26.33	26.73	24.55	16.72	19.59
		0.26	*0.23*	*0.45*	*0.32*	
Gini	0.62	0.63	0.63	0.62	0.57	0.61
		0.00	*0.00*	*0.00*	*0.00*	
Atkinson (1/2)	0.34	0.34	0.35	0.33	0.26	0.31
		0.00	*0.00*	*0.00*	*0.00*	
Atkinson (1)	0.56	0.56	0.57	0.55	0.45	0.52
		0.00	*0.00*	*0.00*	*0.00*	
		Casual job earnings				
90/10 Ratio	12.50	14.86	14.90	15.99	14.79	13.33
		0.10	*0.16*	*0.27*	*0.25*	
Gini	0.53	0.56	0.56	0.57	0.55	0.51
		0.00	*0.00*	*0.00*	*0.00*	
Atkinson (1/2)	0.25	0.26	0.26	0.27	0.25	0.22
		0.00	*0.00*	*0.00*	*0.00*	
Atkinson (1)	0.42	0.44	0.44	0.46	0.43	0.39
		0.00	*0.00*	*0.00*	*0.00*	
		Self-employment earnings				
90/10 Ratio	58.00	49.44	48.42	46.66	232.22	1500.00
		0.49	*0.71*	*0.98*	*7.13*	
Gini	0.75	0.72	0.72	0.72	0.89	0.84
		0.00	*0.00*	*0.00*	*0.01*	
Atkinson (1/2)	0.48	0.45	0.44	0.44	0.73	0.65
		0.00	*0.00*	*0.01*	*0.01*	

Atkinson (1)	0.72	0.69	0.69	0.68	0.92	0.93
		0.00	*0.00*	*0.00*	*0.00*	

Table C3: Allowing counterfactual education to affect participation, employment and wages.

		Participation	Employment	Wage	Price	
	1993	Total earnings				2008
90/10 Ratio	38.33	37.83	47.50	36.05	53.21	46.67
		0.64	*0.61*	*0.99*	*1.15*	
Gini	0.65	0.67	0.70	0.70	0.75	0.67
		0.00	*0.00*	*0.01*	*0.00*	
Atkinson (1/2)	0.38	0.39	0.43	0.42	0.48	0.37
		0.01	*0.00*	*0.01*	*0.01*	
Atkinson (1)	0.61	0.63	0.67	0.65	0.72	0.64
		0.00	*0.00*	*0.01*	*0.01*	
		Regular job earnings				
90/10 Ratio	25.00	27.31	31.94	26.68	21.37	19.59
		0.39	*0.38*	*0.64*	*0.33*	
Gini	0.62	0.64	0.66	0.65	0.61	0.61
		0.00	*0.00*	*0.00*	*0.00*	
Atkinson (1/2)	0.34	0.35	0.38	0.36	0.31	0.31
		0.01	*0.00*	*0.01*	*0.00*	
Atkinson (1)	0.56	0.57	0.60	0.58	0.52	0.52
		0.00	*0.00*	*0.01*	*0.00*	
		Casual job earnings				
90/10 Ratio	12.50	15.20	15.77	16.66	15.65	13.33
		0.16	*0.12*	*0.23*	*0.21*	
Gini	0.53	0.56	0.57	0.59	0.56	0.51
		0.00	*0.00*	*0.00*	*0.00*	
Atkinson (1/2)	0.25	0.26	0.27	0.29	0.26	0.22
		0.00	*0.00*	*0.00*	*0.00*	
Atkinson (1)	0.42	0.45	0.46	0.48	0.44	0.39
		0.00	*0.00*	*0.00*	*0.00*	
		Self-employment earnings				
90/10 Ratio	58.00	51.35	54.79	56.61	318.58	1500.00
		0.76	*0.75*	*1.31*	*8.62*	
Gini	0.75	0.73	0.74	0.74	0.91	0.84
		0.00	*0.00*	*0.00*	*0.01*	
Atkinson (1/2)	0.48	0.45	0.47	0.47	0.76	0.65
		0.00	*0.00*	*0.01*	*0.01*	
Atkinson (1)	0.72	0.70	0.71	0.72	0.93	0.93
		0.00	*0.00*	*0.00*	*0.00*	

For hourly earnings

Table D3: Allowing counterfactual education to affect remuneration only

		Participation	Employment	Wage	Price	
	1993	Total earnings				2008
90/10 Ratio	34.87	34.33	34.83	33.62	27.66	25.00
		0.31	*0.35*	*1.15*	*0.73*	
Gini	0.66	0.66	0.66	0.67	0.64	0.65
		0.00	*0.00*	*0.01*	*0.00*	
Atkinson (1/2)	0.38	0.37	0.38	0.38	0.34	0.35
		0.00	*0.00*	*0.01*	*0.01*	
Atkinson (1)	0.61	0.61	0.61	0.61	0.57	0.58
		0.00	*0.00*	*0.01*	*0.01*	
		Regular job earnings				
90/10 Ratio	27.78	28.23	28.20	28.01	23.09	21.80
		0.28	*0.21*	*0.86*	*0.51*	
Gini	0.62	0.63	0.64	0.65	0.61	0.64
		0.00	*0.00*	*0.01*	*0.00*	
Atkinson (1/2)	0.35	0.35	0.35	0.35	0.31	0.34
		0.00	*0.00*	*0.01*	*0.00*	
Atkinson (1)	0.57	0.57	0.57	0.58	0.52	0.56
		0.00	*0.00*	*0.01*	*0.01*	
		Casual job earnings				
90/10 Ratio	14.08	17.89	17.58	19.62	17.77	15.00
		0.15	*0.14*	*0.40*	*0.40*	
Gini	0.64	0.61	0.60	0.62	0.57	0.59
		0.00	*0.00*	*0.00*	*0.00*	
Atkinson (1/2)	0.37	0.31	0.31	0.32	0.26	0.29
		0.00	*0.00*	*0.01*	*0.00*	
Atkinson (1)	0.55	0.51	0.50	0.53	0.46	0.49
		0.00	*0.00*	*0.01*	*0.00*	
		Self-employment earnings				
90/10 Ratio	71.43	72.35	69.63	72.88	42.52	177.78
		0.86	*0.80*	*2.37*	*1.06*	
Gini	0.73	0.77	0.77	0.78	0.69	0.71
		0.00	*0.00*	*0.01*	*0.00*	
Atkinson (1/2)	0.45	0.52	0.52	0.53	0.40	0.44
		0.00	*0.00*	*0.01*	*0.01*	
Atkinson (1)	0.72	0.76	0.76	0.77	0.65	0.72
		0.00	*0.00*	*0.01*	*0.01*	

Table E3: Allowing counterfactual education to affect employment and remuneration.

		Participation	Employment	Wage	Price	
	1993	Total earnings				2008
90/10 Ratio	34.87	34.33	35.84	32.51	25.82	25.00
		0.29	*0.37*	*0.54*	*0.59*	
Gini	0.66	0.66	0.67	0.65	0.63	0.65
		0.00	*0.00*	*0.00*	*0.00*	
Atkinson (1/2)	0.38	0.37	0.39	0.37	0.33	0.35
		0.00	*0.00*	*0.00*	*0.00*	
Atkinson (1)	0.61	0.61	0.62	0.60	0.55	0.58
		0.00	*0.00*	*0.00*	*0.00*	
		Regular job earnings				
90/10 Ratio	27.78	28.23	29.38	26.34	21.62	21.80
		0.28	*0.32*	*0.55*	*0.47*	
Gini	0.62	0.63	0.64	0.63	0.60	0.64
		0.00	*0.00*	*0.00*	*0.00*	
Atkinson (1/2)	0.35	0.35	0.36	0.34	0.29	0.34
		0.00	*0.00*	*0.00*	*0.00*	
Atkinson (1)	0.57	0.57	0.58	0.56	0.50	0.56
		0.00	*0.00*	*0.00*	*0.00*	
		Casual job earnings				
90/10 Ratio	14.08	17.89	17.99	20.25	17.21	15.00
		0.15	*0.19*	*0.38*	*0.30*	
Gini	0.64	0.61	0.61	0.63	0.56	0.59
		0.00	*0.00*	*0.00*	*0.00*	
Atkinson (1/2)	0.37	0.31	0.31	0.33	0.26	0.29
		0.00	*0.00*	*0.00*	*0.00*	
Atkinson (1)	0.55	0.51	0.51	0.54	0.46	0.49
		0.00	*0.00*	*0.00*	*0.00*	
		Self-employment earnings				
90/10 Ratio	71.43	72.33	71.26	69.43	42.41	177.78
		0.81	*1.21*	*1.67*	*0.75*	
Gini	0.73	0.77	0.78	0.77	0.69	0.71
		0.00	*0.00*	*0.01*	*0.00*	
Atkinson (1/2)	0.45	0.52	0.53	0.52	0.40	0.44
		0.00	*0.00*	*0.01*	*0.00*	
Atkinson (1)	0.72	0.76	0.76	0.76	0.65	0.72
		0.00	*0.00*	*0.01*	*0.00*	

Table F3: Allowing counterfactual education to affect participation, employment and wages.

		Participation	Employment	Wage	Price	
	1993	Total earnings				2008
90/10 Ratio	34.87	36.66	45.00	34.89	33.62	25.00
		0.40	*0.61*	*0.78*	*0.57*	
Gini	0.66	0.67	0.70	0.69	0.67	0.65
		0.00	*0.00*	*0.01*	*0.00*	
Atkinson (1/2)	0.38	0.39	0.43	0.41	0.38	0.35
		0.01	*0.00*	*0.01*	*0.01*	
Atkinson (1)	0.61	0.62	0.67	0.64	0.61	0.58
		0.00	*0.00*	*0.01*	*0.00*	
		Regular job earnings				
90/10 Ratio	27.78	29.67	35.60	28.62	27.90	21.80
		0.29	*0.37*	*0.71*	*0.47*	
Gini	0.62	0.64	0.67	0.67	0.64	0.64
		0.00	*0.00*	*0.00*	*0.00*	
Atkinson (1/2)	0.35	0.36	0.39	0.37	0.34	0.34
		0.01	*0.00*	*0.01*	*0.00*	
Atkinson (1)	0.57	0.58	0.62	0.60	0.57	0.56
		0.00	*0.00*	*0.01*	*0.00*	
		Casual job earnings				
90/10 Ratio	14.08	18.16	18.31	19.72	19.83	15.00
		0.19	*0.17*	*0.34*	*0.26*	
Gini	0.64	0.61	0.62	0.63	0.59	0.59
		0.00	*0.00*	*0.00*	*0.00*	
Atkinson (1/2)	0.37	0.31	0.32	0.34	0.28	0.29
		0.00	*0.00*	*0.00*	*0.00*	
Atkinson (1)	0.55	0.51	0.52	0.54	0.49	0.49
		0.00	*0.00*	*0.00*	*0.00*	
		Self-employment earnings				
90/10 Ratio	71.43	73.85	77.03	79.24	42.50	177.78
		1.17	*1.03*	*1.98*	*0.70*	
Gini	0.73	0.78	0.79	0.80	0.69	0.71
		0.00	*0.00*	*0.01*	*0.00*	
Atkinson (1/2)	0.45	0.53	0.55	0.55	0.40	0.44
		0.00	*0.00*	*0.01*	*0.00*	
Atkinson (1)	0.72	0.77	0.79	0.79	0.65	0.72
		0.00	*0.00*	*0.01*	*0.00*	