DPRU
WORKING PAPERS

Household Incomes, Poverty and Inequality in a Multivariate Framework

Murray Leibbrandt and Ingrid Woolard

No 99/31 December 1999
ISBN: 0-7992-1974-6

Development Policy Research Unit
University of Cape Town

Abstract

The existing work on household poverty and inequality in South Africa has shown that poverty and inequality differ markedly by race, location, education, gender of the head, household demographics and household labour market participation. However, it is important to try and go further than this listing of key correlates and to give any indication of the relative importance of these dimensions. This paper uses a multivariate approach, based on a model of the determinants of household income, to provide a sense of the importance of the key correlates of household poverty and inequality while controlling for the impact of all other correlates.

The models confirm the ongoing importance of race as a fundamental factor structuring South African poverty and inequality even after the influence of all the other poverty and inequality correlates are accounted for. On the other hand, in the multivariate context, the relative income and poverty rankings across provinces appear to be quite different from the rankings derived from conventional provincial poverty and inequality decompositions.

The models also highlight the important role that is played by household members that are educated to at least the secondary school level in pushing households up the income distribution and above the poverty line. This coincides with the finding that the share of adults that are employed is a key driver of household inequality and poverty status. However, there are important differences in the role played by employment status in urban and rural areas. These differences reflect the different types of employment and levels of unemployment that are available in rural and urban labour markets.

Acknowledgements

This paper is part of a larger one-year project on Labour Markets, Poverty and Inequality in South Africa. We would like to thank the African Economic Research Consortium (AERC) and the South African Department of Labour for their generous financial assistance.

1. Introduction

In previous papers[1] we provided detailed descriptions of South African poverty and inequality and also used established poverty and inequality decomposition techniques to further the analysis. Wherever possible we tied our analysis to the role of the labour market. What remains is to provide a sense of the importance of the key correlates of poverty and inequality relative to one another. Is the provincial impact more important than the rural-urban divide in terms of location factors? Can one compare the impact of state welfare assistance relative to educational interventions and which education interventions seem to provide the best return? How large is the burden of unemployment on households and what contribution will employment creation make to household poverty and inequality?

All of these questions are important policy issues in South Africa and this paper provides a framework to address them. Such an exercise requires an integrated household earnings generation model that includes all of the key correlates and indicates the relative importance of each correlate. This necessitates a multivariate approach based on a model of the determinants of household income.

Such an approach is common in the labour economics field where an earnings function serves as the basis for much of the empirical work that is done on the relative importance of various factors influencing individual earnings and earnings inequality (Willis, 1987). However, we apply this approach to household incomes rather than individual earnings. There is far less precedent for such work (Glewwe, 1991 and Ravallion, 1996). The best-developed literature in this spirit uses binary dependent variable models to look at the factors determining whether households lie above or below a poverty line. These poverty regressions have been a standard part of any World Bank country poverty profile for the last ten years. However, such regressions only form part of what we need to do here. We are interested in four interrelated areas:

1. The determinants of household income
2. Whether these relationships are stable across deciles
3. The determinants of household poverty status (the poverty regression issue)
4. The contribution of explanatory factors to household income inequality.

2. Econometric Issues

2.1 Estimation Issues

The sequencing of these questions ties in well with previous methodological approaches used. We derived poverty and inequality indices and decompositions from a framework that started by focussing on the full distribution of household income either in the form of a cumulative distribution function (poverty) or a Lorenz curve (inequality). Here we start with household income before looking more closely at poverty and inequality. The estimation of the first three models requires the use of techniques that are well established in the literature and can therefore be briefly dealt with here. The fourth technique is new and will be discussed in more detail.

[1] See Leibbrandt, Bhorat and Woolard (1999) and Bhorat and Leibbrandt (1999).

We motivate the use of per capita income as the appropriate dependent variable in section 2.3 below. Having decided on this, we estimate the percentage contribution to per capita household income of our explanatory factors by regressing the log of household per capita income on these factors.[2] This is a household analogue to the literature on individual earnings functions. The estimates are presented in Table 1 below. Household incomes are definitely not normally distributed in South Africa but are closer to being log-normal. This provides one justification for the use of a logged form of the dependent variable (Willis, 1996). However, the ordinary least squares procedure gives heavy weighting to the mean values of the dependent and explanatory variables in estimating coefficients. Again, the fact that the distribution of income is generally skewed and that our particular interest is in understanding factors operating in the bottom of the distribution make this weighting problematic.

Quantile regressions provide estimates that answer question number two and, in doing so, provide a check on the ordinary least squares estimates. Quantile regressions estimate a conditional quantile. That is, given a set of explanatory factors and a position in the error distribution, what is the predicted income? Thus, median regression, the most common quantile regression, gives the best estimate of the relation between x and y for households at the median of the conditional error distribution. The 10% quantile regression gives the best estimate of the relation between x and y for households at the tenth percentile of the conditional error distribution, and so on (Rousseeuw and Leroy, 1987 and StataCorp, 1997).[3]

The third question focuses more explicitly on the contribution of our explanatory factors to allocating households above and below the poverty line. This is the standard poverty regression issue. We estimate a series of probit models here.[4] The coefficients from these models are difficult to interpret and we therefore always report a set of marginal effects estimates for each coefficient. These marginal effects are estimated holding all other variables at their mean value.

Technically speaking, question four is the most challenging of the four tasks. There is some international work in this area that has made use of sets of surveys conducted over time.[5] These data have enabled researchers to throw light on factors driving household income inequality by focussing on the *changes* to static decomposition results over time. Unfortunately we do not have a set of reliable surveys over time in South Africa and we will stick to the use of the 1995 October Household Survey and its accompanying Income and Expenditure Survey in this section. Fortunately, there have been two major advances in recent years. At the moment these are only reflected in unpublished work (Fields (1998) and Bourguignon et al (1998)). The two approaches are both much more promising than any preceding methods. In this study we will focus on the Fields approach.

Fields frames his work in terms of two questions: the levels question and the differences question. The levels question seeks a precise method of attributing shares of income inequality to the chosen set of explanatory factors. The differences question seeks to pin down the contribution of each explanatory factor to *changes* in inequality between groups.

In this present context the levels question estimates the contribution of a range of explanatory factors to the inequality of household per capita income in models covering all South African households (Table 5), White households and African households (Table 6) and African urban and rural

[2] In order to be consistent with our earlier work, the estimates that we report in the main text are based on sample data weighted up to the national population by using the appropriate frequency weights. However, the reported levels of statistical significance are based on the unweighted sample data.
[3] The quantile regression coefficients are actually fitted by iterative programming. The statistical package Stata does not allow quantile regressions to use frequency weights to boost the sample observations to population levels as there will be too many observations to converge to a solution. The estimates in Table 2 are therefore unweighted.
[4] Ravallion (1996) provides a thorough and jaundiced review of such probit-based poverty regressions. We will use the probit approach as a complement to the ordinary least squares and quantile approaches.
[5] See the articles in Fiszbein and Psacharopoulos (1995) for a good example.

households (Table 7). A summary presentation of inequality shares in all households is given in Table 4. The differences question then goes on to examine the role of these explanatory factors in explaining the differences in the income inequality patterns *between* White and African households (Table 6) and African urban and rural households (Table 7).

In addressing the levels question, we start with the standard ordinary least squares model of household income generation that we estimated in answering the first question. Fields shows that such a model can be used to carry out an exact decomposition of the contribution of all the variables in the model to the *variance* of log per capita income. In our model, Yit is household per capita income. We use the same set of explanatory factors, x1...xj, as we have in answering the first three questions. Using ordinary least squares regression we then estimate the coefficients, aj. The value of these coefficients reflect the percentage contribution that each factor makes to household per capita income. Clearly this still focuses on the determinants of income and not income inequality. However, the heart of the Fields technique is to prove that an inequality share for each of the factors can be derived from the following formula:

$$s_j = \text{cov}\,[a_j\, Z_j, \ln Y] \,/\, \sigma^2(\ln Y) = \frac{a_j * s\,(Z_j) * cor[Z_j, \ln Y]}{s\,(\ln Y)}$$

Strictly interpreted this provides us with the share of factor Z_j in explaining inequality as measured by the log variance. The elements of this formula are intuitive showing that a factor may play a large role in explaining income inequality if:

- It has a large aj; i.e., it is an important factor in explaining earnings;
- It has a large standard deviation, $\sigma(Z_j)$; i.e., it is a variable that is highly unequal itself; or,
- It is highly correlated with the log of income, *cor*$[Z_j, \ln Y]$.

The presence of the standard deviation of lnY, σ(lnY), in the denominator ensures that all of these effects are interpreted relative to the magnitude of the inequality in lnY.

From Table 47 it can be seen that, in some cases, the contribution of individual variables to inequality is represented whereas, in other cases, the contribution of a block of variables to inequality is represented. Block contributions are simply derived by aggregating individual contributions.

The role of the residual requires some discussion. A strength of this regression-based methodology is the fact that the regression model generates a residual which is treated as one of the factors contributing to inequality in lnY. In telling us what portion of the inequality in lnY is explained by the residual we are implicitly being told what portion of inequality is left unexplained by our explanatory factors.

Finally, the log variance is a recognised inequality measure but it is not one that enjoys routine usage. This is not a cause for concern though as Fields shows that the estimated shares that are derived using the log variance are those that would be derived for a broad class of the most popular income distribution measures. Thus, the decomposition is very robust.

The differences question then goes on to examine the role of these factors in explaining the differences in the income inequality patterns between two groups. Unfortunately, Fields shows that the differences question cannot be addressed in such a way that the answer is independent of the choice of inequality measure. For any chosen inequality measure I(.), the contribution of the j'th factor (including the residual) to the change in a particular inequality measure between country/group/time 1 and country/group/time 2 is given by:

$$\pi_j(I(.)) = [s_{j,2} * I(.)_2 - s_{j,1} * I(.)_1] / [I(.)_2 - I(.)_1]$$

It is an empirical question whether the choice of inequality measure makes a large difference or a small one in any particular context. Therefore, we use two inequality measures; the Gini coefficient and the log variance in our decomposition work.

2.2 Choice of variables

The usefulness of the answers that we attain from any modelling is dependent on the formulation of a suitable household income generation equation. There are two aspects to suitability here. The first is that the right hand side variables provide the links that we need between households and the labour market. The second is econometric. It is difficult enough to formulate a suitable specification for individual earnings and near impossible to do so at the level of the household (Glewwe (1991)). We make no pretence at deriving a structural model based on a careful analysis of household welfare and decision-making. Rather, we choose a variable set that is consistent with the South African inequality and poverty situation that we have sketched in previous studies.[6] Then, we work hard to ensure econometric adequacy for our estimates (Ravallion, (1996)). Although our particular focus is on the role of the labour market, many variables make a contribution through the labour market or interact with the labour market variables. Therefore, the scope cannot be defined too narrowly. The estimations below all use the following variable set:

- Household Head

  South Africa has a history of migrant labour and divided families. This legacy is still very much with us and the female-headed households in rural areas are often regarded as the most vulnerable of all households. We therefore specify a dummy variable set that covers female and male, resident and absent possibilities. The household with a resident male head is the default.

- Household Composition

  In line with international findings, poorer households are generally larger than better-off households are. In addition, such households usually have more children. As mentioned in the discussion of the household head, household composition was highly disrupted as households attempted to adjust to apartheid policies. Therefore, it is not adequate to merely flag household size (and household size squared) in the equations. We specify a set of variables capturing the numbers of children less than 7, children 8-15, females 16-59, males 16-59 and adults older than 60. These variables can be expressed as numbers or shares and we tested and used both alternatives.

- Locational and Regional Effects

  Earlier tables clearly showed that the incidence of poverty is far higher in South Africa's rural areas and particularly in the previously African areas. This is captured through a rural/urban dummy variable in which urban takes on the value of 1. In addition, the best set of proxies for regional economies within South Africa are the ten provinces. We therefore include a full set of provincial dummy variables with the Western Cape being the default.

[6] The variable set is identical to that used in the poverty profile presented in earlier papers. See Leibbrandt, Bhorat and Woolard (1999) and Bhorat and Leibbrandt (1999).

- Race

  Previous sections of this study have repeatedly flagged the importance of race as a dominant and lingering marker of both inequality and poverty. There are four race dummies with African being the default.

- Labour Market Factors

  The dominant theme of the decomposition analyses in our previous work was the role of employment and unemployment in inequality and poverty. In this paper we therefore include variables capturing the extent of successful integration into wage or self-employment and the extent of the unemployment burden. To allow for the impact of migrant labour, we include a variable capturing the number of remitters providing remittance transfers into the household.

  These labour market variables are not dummy variables. Rather two types of variables are constructed. First, the number of working, unemployed or absent migrant adults is used. Second, these numbers are converted into shares of the economically active adults in each household and these shares are used. The shares of these three variables do not have to add up to 100 percent. As a rule remitters are not counted as formal members of households. In addition, adult household members that are not participating in the labour market are neither employed or unemployed. Thus, only in the case of households with no migrants and full labour market participation by adults will the shares of employed and unemployed sum to 100 percent.

- Education Levels

  Education is key on both the individual and household levels. In 1995 and even today the provision of education is overwhelmingly (especially for African and Coloured groups) the responsibility of the state. We capture the influence of education through a set of variables covering adult household members with no education, primary, some secondary, completed secondary and any form of tertiary education. The completed secondary variable is important because secondary education ends with a standardised, national matriculation examination.

  As with the labour market variables, we reflect these educational variables either in terms of numbers of adults or in terms of shares of adults. In interpreting these education effects the derivations are important. In the case of the number of adult household members with no education, all household members with some education are represented as a zero. For 75 percent of all households (99 percent of White households and 66 percent of African households) a zero (or a zero share) is recorded for this variable. The tertiary education variable has a similar pattern. In this case, 88 percent of African households have no adults with tertiary education and, therefore, they record a zero share for this variable. The respective figure is 63 percent for White households.

- Social Welfare

  South Africa has an array of child maintenance grants, unemployment insurance schemes and universal state-funded old age pensions. As of 1995 and even today, the extent and coverage of child support schemes and unemployment insurance has been patchy. Pensions are by far the dominant form of social transfer in South Africa. We therefore include a variable capturing the number or old age pensioners in the household or the share of pensioners as a percentage of adults.

This variable list is not exhaustive. There are two major omissions. First, a potentially important labour market effect that is not explicitly captured in the models is the type of employment. International literature sometimes attributes sectoral and occupational variables to households (Huppi and Ravallion, 1991). This allocation of individual labour market characteristics to households is usually based on the labour market participation of the head of the household or the major earner in the household. Given that the survival strategies of South African households generally involve participation in a diverse array of activities, it is difficult to justify this practice here. Rather, one of our specifications estimates separate equations for rural and urban households. *A priori* the major reason why employment, unemployment and education coefficients would differ across these two estimations is because the labour markets differ by sector and occupation in urban and rural areas.

Second, aside from human capital, there is not a block of variables reflecting assets and wealth. The 1995 October Household Survery (OHS) and Income and Expenditure Survey (IES) data do not contain very rich information on assets and are particularly weak on the agricultural assets that are usually fully specified in developing countries. One variable that is contained in the data is the valuation of the place of residence. When this variable is included in the models it makes a very small contribution and has no impact on the values of the other coefficients. However, it needs to be acknowledged that the inclusion of this variable is only a limited exploration of possible interactions between assets and income generation.

2.3 Specification Issues

In estimating our models two specific econometric (and conceptual) issues arise. First, there are a range of possible interactions between household size and household composition and the other right-hand side variables. Second, there are other endogeneity issues that require attention.

We confront the first problem in a number of ways. First, we use per capita income as the left-hand side variable in preference to total household income. We could have used income per adult equivalent instead of per capita income. However, we do not want to include the influence of household composition on the left-hand side variable because we use a full set of household composition factors on the right hand side of all models. Finally, as observed when we defined our variables, we specify all models using numbers of household members as well as shares of the household.

Estimates are very sensitive to these choices between various household size and composition blocks and between the use of numbers versus shares in defining education, labour market and welfare variables. The specification that is most successful in untangling the relationships between household composition, education, pensions and the labour market is one that retains a full household composition block as numbers and then uses shares for education and labour market and pensions blocks. We report these results in the discussion below.[7]

The second major econometric issue involves endogeneity on the right-hand side of the equation. Aside from race none of the explanatory variables are truly independent. South Africa's history is such that race is certainly partly responsible for the movement in nearly all of the other right-hand side variables. For example, in simple regressions of race on the education and labour market variables, the race dummies are always significant.

We acknowledge this problem by estimating our models for all households and then separately for African and White households. The estimations by race are interesting in their own right as they provide useful information on the within-race determinants of income, poverty and inequality. Inspection of Table 1 reveals that the estimated coefficients for the "All Households" regression lie close to the African estimates and within the range implied by appropriate weighting of the separate African

[7] All other estimations are available from the author.

and White estimates. While this is not a rigorous control for the influence of race, dramatic changes that took the estimated coefficients outside of this range would certainly have implied a major endogeneity problem with race that is not adequately dealt with by the inclusion of race dummies.

Besides the racial factor, there are other endogeneity issues that require attention. An important labour market possibility is the fact that the labour market and education blocks may operate differently in urban and rural areas if urban and rural labour markets are very different. It is true that the estimated coefficients for all households and for African households change appreciably if the models are estimated without the urban-rural dummy variable. Thus, we always include this dummy variable or estimate separate equations for urban and rural areas. In order to ensure that this spatial effect is not wrapped up with the racial effects, we limit these rural-urban estimations to African households. Thus, in all models, the flow is from national households to African households and then to urban or rural African households.

The final endogeneity issue that we address is the influence of education on the labour market variables. It is not easy to think of an explicit control at the household level. The usual labour market procedure would be to handle the indirect impact of education on occupational attainment (for example) through a multinomial logit estimation of education on occupations. However, the labour market variables are not categorical, as the relevant variables are shares of adult household members that are employed or unemployed or remitters. A roughly analogous procedure to the multinomial logit is to regress all of the educational variables on each of the three labour market variables. This was done and, while some of the educational coefficients were significant, the R-squared coefficients for these models were very low indeed. An additional piece of evidence in support of this is derived by inspection of the last two columns of Table 1. These show that the impact of education on African household per capita income is very sensitive to separate rural/urban divisions but that the labour market variables retain their consistency despite this.

In sum then, in this sub-section we have made the case for a fairly simple, linear specification of our chosen variables as the basis for all of our modelling. We now proceed to answer the four questions that we tabled at the beginning of this section with the help of four models that all use this specification.

3. Estimation, Results and Discussion:

The important determinants of household income, poverty and inequality.

Before we move to a variable-by-variable discussion, there are a few general points to be made about the four models. Looking across Tables 1 and 2 it is noticeable that the median-based quantile estimates, based on the median of the error distribution, are generally quite close to the mean-dominated estimates derived by OLS. However, this is not true of the bottom-decile quantile case. The coefficients for this regression are usually lower than at the median or at top decile. In African households, it is only the share of remitters and old age pensioners that offer an exception. Lower "returns" to factors at the bottom of the error distribution hints at the fact that factors play a larger role where income is more widely dispersed. The factors therefore appear to be positively correlated with household income inequality. We will have more to say about such contributions to inequality below.

The first thing to note about the poverty results of Table 3 is that the White model does not work well at all. This is a reflection of the fact that there are not enough poor white households in many of the categories to estimate the coefficients. On the other hand, the African model shows that, generally, the factors that are a positive influence on incomes are also positive influences on the probability of not being poor. Some factors show themselves to be more important in the poverty regression than in the full income models. Old age pensions in African households are an example of this.

For both African and White inequality models, there are substantial amounts of residual (unexplained) inequality (44% and 63% respectively). As in the case of the poverty regressions, for White households in particular, we are left with the strong impression that we have not come to grips with the key factors driving inequality. It might well be that a focus on wealth and asset variables would be necessary to explain white inequality. While this is speculative there is no denying the fact that the within-race equations leave far more residual inequality than the "all households" model (30% residual inequality) that explicitly deals with race through the racial dummy variables. This is also true of the African urban and African rural equations in which a large amount of the inequality (42.1% and 55.7% respectively) is left unexplained. Thus, we seem to have a better model of all household income inequality in South African than within-race group inequality or African urban-rural inequality.

The four issues that we are examining are closely linked and best dealt with in one coherent discussion. In order to do this we need to look across all four models (Tables 1-7), variable by variable, as this allows us to tell an integrated story around the results.

In the "all households" estimations of the **head of household** block, the female coefficients have the anticipated signs and values. Relative to having a male head that resides at home, average household income (per capita) is 27% lower if there is a female head at home and 17% lower if there is a female head working away from home. Households with an absent male head earn 14% more on average. The poverty probabilities are consistent with this in that households with resident female heads increase the probability of being poor by 7.8% relative to resident male heads.

The above picture remains consistent across races and deciles although the respective coefficient magnitudes vary. It is noteworthy that the disadvantage associated with a resident female head relative to a resident male head is particularly acute in African rural households and in White households. The former finding is expected, given the conventional picture of disrupted African rural households. The latter is more surprising. The quantile results are even more surprising in that, for both race groups, they show higher female disadvantage when the top decile of the error distribution is given explicit attention.

The head of household block of variables makes a very small contribution to overall inequality (1.6%). It is more important in explaining African inequality (2.6%) and especially African rural inequality (3.1%).

It is dismal to note the negative sign of the **household composition** estimates in Table 1 and the positive signs in Table 2. These imply that each household member is a net burden on per capita household income and increases the probability of being poor. This is robust across deciles and across racial and urban-rural estimations. Surprisingly, adults that are 60 or older provide the single exception to this trend. In particular it can be seen that this positive coefficient is significant in the bottom-decile quantile equation and in the African rural OLS equation. However, in most models the coefficient is negative but not significant. Finally, this coefficient is negative and significant in White households.

As a whole, household composition factors account for an important share of inequality in the "all household" model as well as in all of the African models. Tables 5-7 show that the major contributors responsible for this are the two factors covering children (Kid7, Kid15). The decompositions show that these high inequality contributions arise from the high negative income coefficient associated with these two factors and from the strong negative correlation between these factors and per capita income.

These findings are not that startling because, in static estimations it is almost by definition that children make negative contributions to per capita income. More worrying is the fact that economically active females and males (F16_59 and M16_59) also make negative income contributions on average. These negative coefficients are even larger in the bottom-decile quantile regression. The female economically active variable also makes a sizeable contribution to inequality in African and African urban models.

The conventional wisdom in South Africa has it that Gauteng and Western Cape are the two best-off provinces in South Africa. Western Cape is the omitted dummy in the **provincial** dummy variable block and our models therefore allow for an assessment of this claim in the multivariate context. For example, it is interesting to note that Free State is revealed to be the worst off province across all models and that the Northern Province appears to be one of the better-off provinces. Both of these results are strongly contra the conventional provincial poverty rankings.

More generally the results reveal a fairly complex situation that differs strongly along rural and urban dimensions. For “all households” and for African households, the results suggest that Gauteng and also KwaZulu-Natal and Northern Province have higher mean and median incomes than Western Cape after controlling for all other factors. They also have relatively lower average probabilities of being poor. The African urban-rural results show that this aggregate outcome is the result of two contrasting processes. In urban areas the general trend tabled above is strongly observed. However, in rural areas all provinces are strongly disadvantaged relative to Western Cape both in terms of mean income and in terms of the probability of being poor. This rural result is due to the fact that the Western Cape did not absorb any of the predominantly rural and very poor homeland areas in 1994 whereas many other provinces did.

Table 4 shows that the aggregate provincial contribution to "all household" inequality is 3%. This is lower than expected. However, the provincial block is competing with the urban-rural dummy variable in this model as both are components of the contribution of spatial factors. The inequality contribution rises to a high of close to 5% for urban African inequality. In this case, the provincial contribution is picking up the fact that, for some provinces urban implies large metropolitan cities whereas in other provinces it implies very much smaller secondary cities.

The estimation of **urban-rural** differences in household per capita incomes reveals smaller than expected coefficients in all the OLS and quantile models. The models even suggest that, holding all other variables constant, mean and median household incomes are higher in rural White households than urban White households with similar characteristics. Moreover, the contribution to inequality in "all households" and in African households is just short of 5% in both cases. This is certainly a lesser share than expected. However, the estimated poverty marginal effects indicate large and significant increases in the probability of being poor associated with rural households, particularly for African households. In any event our provincial discussion above has flagged the fact that, the separate urban and rural equations for African households allow for a much fuller assessment of the influence of urban-rural dimension on all explanatory factors. This is clearly a more important dimension than is indicated by estimating a urban-rural dummy variable in Tables 1-6.

The analysis of the contribution of the explanatory factors to differences in urban and rural inequality provides a useful tool for direct comparison of the urban and rural equations. The final two columns of Table 7 present the results of the differences decomposition using two inequality measures: the Gini coefficient and the log variance. The urban and rural Gini coefficients are almost exactly the same (0.52 and 0.53 respectively). Therefore, there is very little difference to explain by a large number of factors. This is what lies behind the unstable results for the Gini coefficient decomposition in Table 7. In the log variance case, African urban inequality (0.96) is about 20% higher than African rural (0.80). This is a substantial difference and it is hardly surprising that the log variance is far more successful in decomposing the full extent of this difference (100%) in a stable way. Table 7 reveals that 42% of the wider urban distribution can be attributed to *Shwork*, 27% to the share of adults with matric, 17% to the provincial block and 15% to the share of unemployed. Education and labour market factors are therefore seen to play the largest role in driving the differences between the African urban and rural equations and, more specifically, in explaining the greater urban inequality.

As with the urban-rural situation, the influence of **race** is captured both as a dummy variable set and in equations that are separately specified by race. However, unlike the urban-rural case, the racial dummies are strikingly large in their own right. Relative to African incomes, there are large premiums associated with Coloured, Asian and especially White incomes. The quantile regressions suggest that the 84% mean difference between African and White households with the same characteristics may underestimate the difference as both the bottom decile and top decile estimates are higher. Table 3 reveals that racial differences in the probabilities of being poor are also very large.

In addition, from Table 4 it can be seen that the most important of the block contributions to inequality is the one due to race. In the "all households" models, it accounts for 17.3% of total inequality. Even if the inequality contributions of individual factors are aggregated into group shares, the contribution of race remains the largest of any of the variable groups. In the multivariate context, these findings are particularly startling as this racial contribution does not include racial biases in education or the labour market. As such, it is a lower-bound inequality estimate that starkly confirms the continuing importance of race in South Africa. When race is included as a single explanatory factor in this model it accounts for 39% of the total inequality. This upper-bound estimate is very much in line with the between race contributions that we derived using Theil and Atkinson measures earlier in this report.

The final two columns of Table 6 present the results for the decomposition of the difference between African and White inequality using two inequality measures: the Gini coefficient and the log variance. Both measures suggest that the inequality within White households is lower than the inequality within African households. The difference that needs to be explained is 4% in terms of the Gini coefficient and 39% in terms of the log variance. Thus, as with the urban-rural case earlier, the two measures seek to explain markedly different inequality differences.

Given this situation, it is hardly surprising that the log variance is more successful in decomposing the full extent of the difference (100%) in a stable way. Both measures suggest that the old age pension and residual factor go against the trend and contribute to a situation in which White households are more uneven than African, holding everything else constant. All the other factors work in the direction of the measured total difference in that they explain a move to a wider African distribution. For both the Gini coefficient and the log variance *Shwork*, Urban, Kid7 and Kid15 and Shsec are seen to be the major factors responsible for the greater African inequality.

Shwork is by far the largest contributor. In this case, the standard deviations of the African and White *shwork* variables are very similar, this difference is largely attributable to differences in the income coefficients and in the correlations between *Shwork* and log per capita income. Indeed, the *Shwork* income coefficient (76.6%) and correlation coefficient (0.518) are both the largest of any variables in the African model.

The second largest contributor to the difference is the urban-rural factor. This is due to the fact that both the income coefficient and the correlation coefficient are positive in the African model and negative in the White model. The shift from urban to rural widens the distribution of income in both models but it corresponds to a shift down the distribution in the African case and a slight shift up the distribution in the White case. Thus, the difference is quite marked.

The impacts of the two children variables (Kid7 and Kid15) and the Shmatric education factor are very similar in size and in underlying explanation. In all cases, the standard deviations and the income correlations of the factors are much larger in African households. White households of all income levels rarely have large numbers of children and almost all adults have some secondary education. Thus, these three factors are not major drivers of White inequality. In African households there is a far bigger range of numbers of children in the household and shares of adults with secondary education. In addition, households with large numbers of children tend to be found in the lower half of the African distribution and households with higher shares of adults with secondary schooling tend to be found in the upper half of the African distribution. These three factors are therefore much more important contributors to African household inequality than to White inequality.

The education block follows the race block in Tables 1-5. However, in order to make sense of the education results it is necessary to talk about the labour market. Thus, we discuss the labour market results before we discuss the education results.

Indeed, the contribution of an increased **share of working adults** in the household is the highest of any single continuous variable in all income, poverty and inequality models. In the "all households" model, each additional worker makes a large contribution to household per capita income (68%), to the probability of avoiding poverty (28%) and to inequality (10%). The contributions are even higher for African households at 77%, 44% and 14.7% respectively. In addition, the high income benefits are robust across quantiles. While income and poverty contributions stay high when urban and rural African households are examined separately, there is an interesting reversal in the estimated coefficients. For household per capita income urban benefits are greater than rural at 83% and 74% respectively. For the probability of poverty avoidance, urban benefits are less than rural at 33% and 46% respectively. A plausible explanation of this reversal would be the fact that better remunerated employment is available in urban areas compared to rural areas thus raising the income benefits of increases in the share of employed adults to urban households above those to rural households. However, given the scarcity of employment in rural areas relative to urban areas and the absence of viable alternative activities for rural households, increased access to any employment makes a larger contribution to lifting a household out of poverty in rural areas than in urban.

We have already given extensive discussion to the dominant role of the share of working adults in explaining the differences in inequality between African and White households and between African urban and rural households. To some extent, this has pre-empted a discussion of the direct contribution of *shwork* to inequality. Across all inequality models, this contribution is attributable to:

- The size of the income coefficient. A unit increase in the share of working adults raises per capita household income by 76%.
- The large standard deviations for this variable. There are large differences across households in the shares of economically active adults that are employed.
- The high correlation of this variable with lnY

Unemployment makes a large negative contribution to income and poverty. However, the magnitude of this influence is never more than a half that of the comparable employment coefficient in all models. Thus, there is not an opposite-but-equal symmetry between the impacts of unemployment and employment. In a trivial sense this is to be expected because the income contribution made by working members to their households depends on the quality of employment whereas the direct income contribution of the unemployed is always zero. There is another plausible explanation of this result. If unemployed household members have weaker labour market characteristics than those members of the household that are already employed then the lost potential earnings of the unemployed would be lower than the actual earnings of the employed.

It is important to note that this analysis of the relative contributions of the employed and the unemployed does not imply that the costs to households of unemployment are lower than expected. In an absolute sense, a rising share of unemployed members takes a heavy income toll on households. Moreover, the quantile regression results in Table 2 show that this toll is higher when the estimate is anchored around the bottom decile than when it is anchored around the median or the top-decile. This is particularly true for African households. Thus, the absolute cost of the unemployment is higher for those at the bottom of the distribution.

From Table 4 it can be seen that the contribution of unemployment to inequality is low in all models. Particularly unexpected is the fact that the contribution to African rural inequality (1.6%) is less than the contribution to African urban inequality (3.7%). This is a reflection of the fact that the negative income coefficient is less in rural areas (-0.23% compared to -0.38) and that the negative income correlation is also weaker in rural areas (-0.243 compared to -0.356). Both of these findings require careful interpretation. The lower income coefficient for unemployment is most likely a reflection of the poorer earnings possibilities in rural areas. The lower (negative) correlation coefficient reveals that a household with a high share of unemployed adults is likely to be closer to the bottom of the urban distribution than the rural. Thus, rather than signalling the unimportance of the unemployment problem in rural areas, this is a reflection of the endemic nature of the unemployment problem in rural areas. The unemployed are found in all rural households including those in the middle of the income distribution and, therefore, unemployment is not strongly correlated with those households at the bottom end of the distribution.

The impact of **migrant remittances** on income, poverty and inequality is small. Rather surprisingly, this is the case even for African rural households. Thus, this is the one labour market factor that does not throw up any interesting results in the multivariate context.

The **education** variables show very strong 'returns' (in terms of income and poverty avoidance) to households in which a large share of adult members have secondary education and higher. The significance of completed secondary education (matric) as distinct from some secondary education is also clear. There are so few adults in white households with no education or only primary education that these coefficients are always statistically insignificant when estimated with unweighted sample data (quantiles and Appendix II). We will therefore ignore these coefficients. Focussing only on the secondary, matric and tertiary levels, it can still be seen that there are important differences across races in terms of the household returns to education. African returns are higher at all levels and across all quantiles. This is even more marked in the poverty regressions. Table 3 shows household education levels are not an important factor in avoiding poverty for White households. In contrast to this, African households get very large poverty avoidance returns from increasing the shares of adults with higher levels of education.

The income contribution of secondary-schooled adult household members is about 32% for African households in both urban and rural areas. However, there are interesting urban-rural differences at the matric and tertiary levels. For urban areas, matric generates the highest return (51% compared to 45%). In rural areas this is reversed (46% and 66%). All of these returns are high. However, it would seem that adults with completed secondary education have good opportunities for income-generation in urban areas whereas the best rural opportunities require tertiary education. This is plausible. In the 1980s and early 1990s, the best rural income-earning opportunities for Africans have generally involved skilled employment in the public sector (Leibbrandt and Woolard, 1995).

When the focus shifts to the poverty regressions, rural returns are markedly higher than urban for secondary schooling (a 21% versus a 12% decrease in the poverty probability) and matric (a 32% versus a 17% decrease in the poverty probability) and marginally higher for tertiary education (15% and 18% respectively).

In the "all households" and White households inequality models, the education variables make the largest block contribution to inequality of all variable sets. In all other models, this block is the second largest next to the labour market block. Within this education block, Table 4 shows that shmatric makes the largest contribution to inequality of any education factor in all of the models. Next to *shwork*, it is the second largest contributor to inequality of all of the individual factors. In the "all households" and White households models, the shmatric contribution is only marginally smaller than *shwork*. All of these findings clearly establish the importance of the educational factors in inequality. Earlier we flagged the importance of shmatric in differences between rural and urban inequality and the importance of shsec in explaining differences between African and White inequality.

Judging by the "all households" equations, **pensioners** appear to have a small impact on average household earnings. However, this is an instance of the "all households" situation representing a bad average of two disparate African and White trends. In the case of White households, a rising share of pensioners in the household makes a negative contribution to income and a positive contribution to inequality. In all African models, a rising share of pensioners makes a positive contribution to income. This is especially notable in rural areas. In addition, the impact on poverty is very strong. Earlier in this study we addressed the role of pensions in poverty alleviation work. The multivariate work now affirms our earlier contention that pensions are well targeted in terms of their welfare objectives.

Old age pensions also yield some interesting inequality results. For African households they make a *negative* but very small contribution to inequality. This is a consequence of the fact that pensions make a positive contribution to income but households with pensioners in them are in the lower income groups. This variable is therefore negatively correlated with lnY. This is somewhat surprising as it is well known that pensions play an important role in the rural social safety net for African households. However, the correlation coefficient is very close to zero implying that households with pension incomes are not the poorest of the African poor. Indeed, in African rural households the correlation is slightly positive giving this factor a small positive role in inequality.

4. Concluding Points

There is a large body of work, including our own, that teases out and describes the key dimensions of poverty and inequality in South Africa. From the policy point of view, there is a pressing need to provide a sense of the relative importance of these key dimensions. There is virtually no precedent for such work in South Africa. In this paper we have taken a first step in this direction. We have estimated and discussed four multivariate models of household income determination, household poverty and household inequality. Econometric adequacy is elusive in such household-level models but we have endeavoured to be as careful as possible in our estimations.

For this study, the major issue at stake is to understand the role of the labour market in driving income determination, poverty and inequality. Our interpretation of results has been slanted towards this angle. For example, urban and rural differences have been seen to reflect different types of employment and levels of unemployment in rural and urban labour markets. This labour market angle is justified by the results themselves. Employment of adult labour market participants is shown to be the biggest single contributor to household per capita income, household poverty avoidance and household inequality. Unemployment of such adults imposes a high cost on households. The aggregate impact of new job creation is especially significant as it involves the removal of the negative unemployment effect and the addition of the positive employment effect. On average, the net impact of this would be very close to a 100% improvement in average per capita household income and a 40% reduction in the probability of the household being poor. In addition, the aggregate contribution of these labour market variables to “all household” inequality is 12.3%.

While education is not only a labour market issue, education and the labour market are intimately related at the policy level. This paper has repeatedly shown the important, positive contributions made by household members that are educated to at least the secondary school level. Households get particularly high returns from adult members with completed secondary schooling and tertiary education. Moreover, the block of education variables is always amongst the top two contributors to household inequality.

All in all then, our models have certainly justified the fact that further studies will give detailed attention to the labour market and to the role that education plays in determining labour market outcomes. The discussion in this paper has not been particularly precise about urban versus rural labour markets or the factors determining unemployment and earnings. The reason for this is that we have yet to present a picture of the way that the labour market operates in South Africa. Within the labour market it is individuals that are employed or unemployed and it is individual characteristics that determine this as well as consequent earnings for the employed. Thus, in the labour market section we will move away from households and focus on individuals.

There are two additional findings from our modelling that warrant flagging. First, the provincial analysis reveals some interesting dimensions. The analysis of provincial poverty shares that we undertook in an earlier section of this study concluded that the provincial shares are very sensitive to the choice of poverty measure. In the multivariate context, the relative income and poverty rankings appear to be quite different from the conventional views and quite unstable across different equations. Thus, the multivariate models certainly provide additional support for our earlier cautionary note. This is important because provinces are institutional intermediaries in the social service delivery process in South Africa and, to some extent, provincial budget allocations are based on measurement of need *across provinces*. The policy question that arises is what are the currently used needs rankings and how are they derived? The multivariate models have also shown that there are important urban-rural differences in mean income and poverty *within* provinces. Appropriate intra-provincial service delivery rules are therefore going to be key in ensuring successful anti-poverty policy no matter how provincial shares are derived.

Finally, the multivariate models confirm the ongoing importance of race as a fundamental factor structuring South African poverty and inequality. This is in line with our earlier racial decomposition work and greatly strengthens this work by showing that race retains its direct importance even after controlling for its indirect influence on access to education, location and employment opportunities. This is a daunting indicator of the magnitude of the project South Africa faces to redress our racial legacy.

Table 1: The Determinants of Household Income

	Percentage Contributions to Household Income (Ordinary Least Squares Estimates on Log Per Capita Income)				
	All Households	White	African	African Urban	African Rural
Head:					
Femres	**-.2701286**	**-.305298**	**-.2556111**	**-.2040247**	**-.293596**
Femabs	**-.1709232**	**-.1741878**	**-.161133**	-.0551329	**-.2313877**
Maleabs	**.1396368**	**.0767289**	**.1643966**	**.1768707**	**.1557935**
Composition:					
Kid7	**-.1655144**	**-.2724278**	**-.1500638**	**-.1692102**	**-.1431333**
Kid15	**-.1589389**	**-.22314**	**-.1479725**	**-.1844129**	**-.1336192**
F16_59	**-.064871**	**-.0924504**	**-.0685099**	**-.0706807**	**-.0616601**
M16_59	**-.048516**	**-.0432652**	**-.0493758**	**-.0388929**	**-.0519207**
Ad60	.0070972	**-.1528326**	.012329	-.0311293	**.0341542**
Province:					
Eastern Cape	**-.1615817**	-.0732565	**-.1120541**	-.0519413	**-.4851876**
North Cape	**-.2101004**	**-.1728977**	**-.2477614**	-.1297678	**-.6558229**
Free State	**-.3017424**	**-.1825402**	**-.2846519**	**-.1610042**	**-.6839104**
Kwazulu/Nat	**.091771**	-.0249695	.1710477	**.1880675**	-.1776282
North-West	-.0413994	-.0243796	**.0075066**	.0422307	**-.3501167**
Gauteng	**.167128**	**.1195842**	**.2092098**	**.2341174**	-.2290835
Mpumalanga	**-.0654942**	**-.1153256**	-.0090146	**-.1568967**	-.2903379
North Prov	**.0921316**	.0586052	**.1553172**	**.2308377**	**-.2055513**
Urban	**.2386275**	**-.2051718**	**.2764435**		
Race:					
Coloured	**.1652885**				
Asian	**.4500624**				
White	**.8450188**				
Education:					
Shno_ed	.0177529	.6961734	.0072023	.1054994	-.0387509
Shprim	**.0764243**	.2909542	.0791113	.1252036	.0621953
Shsec	**.3641685**	**.4393242**	**.3187812**	**.3098628**	**.305834**
Shmatric	**.5072614**	**.4111087**	**.5068706**	**.5086073**	**.4591706**
Shtert	**.4202615**	**.3251523**	**.5269322**	**.4554513**	**.6646113**
Labour Mkt:					
Shwork	**.6812947**	**.4659047**	**.7659807**	**.8308757**	**.735619**
Shunemp	**-.3260304**	**-.3031824**	**-.2934526**	**-.3835213**	**-.2292114**
Shmig	**.1736706**	.3392358	**.1644313**	**.1753879**	**.1205743**
Welfare:					
Shoap	**.0747939**	**-.5387748**	**.3447923**	**.2879481**	**.3781855**
Cons	**8.067799**	**9.433772**	**7.938406**	**8.10466**	**8.333174**
	N=28 578	N=5 224	N=18 476	N=7 744	N=10 732
	Prob>F=0.000	Prob>F=0.000	Prob>F=0.000	Prob>F=0.000	Prob>F=0.000
	AdjR2=0.70	AdjR2=0.37	AdjR2=0.56	AdjR2=0.44	AdjR2=0.37

Note: Bold Coefficients are significant at the 1% level using unweighted sample data

Table 2: The Deteminants of Household Earnings at Different Quantiles of the Error Distribution (Shares)

	Quantile Regression **ALL HOUSEHOLDS**			Quantile Regression **AFRICAN**			Quantile Regression **WHITE**		
	Median	(0.1)	(0.9)	Median	(0.1)	(0.9)	Median	(0.1)	(0.9)
Head:									
Femres	**-0.26**	**-0.26**	**-0.32**	**-0.24**	**-0.24**	**-0.30**	**-0.29**	**-0.21**	**-0.39**
Femabs	**-0.17**	**-0.13**	**-0.24**	**-0.16**	**-0.12**	**-0.24**	**-0.14**	.01	**-0.23**
Maleabs	**0.15**	**0.24**	.09	**0.15**	**0.20**	.12	.10	.20	.03
Composition:									
Kid7	**-0.18**	**-0.17**	**-0.16**	**-0.16**	**-0.16**	**-0.14**	**-0.31**	**-0.29**	**-0.29**
Kid15	**-0.17**	**-0.17**	**-0.16**	**-0.16**	**-0.17**	**-0.15**	**-0.22**	**-0.23**	**-0.20**
F16_59	**-0.07**	**-0.08**	**-0.06**	**-0.07**	**-0.10**	**-0.06**	**-0.09**	-.06	-.11
M16_59	**-0.05**	**-0.06**	**-0.05**	**-0.06**	**-0.08**	**-0.04**	-.04	-.06	-.05
Ad60	.02	.02	-.02	.02	**0.04**	-.02	**-0.15**	**-0.18**	-.12
Province:									
Eastern Cape	**-0.14**	**-0.16**	**-0.06**	**-0.11**	**-0.17**	-.12	-.07	-.07	-.17
North Cape	**-0.19**	**-0.28**	-.11	**-0.16**	**-0.35**	**-0.26**	**-0.19**	**-0.22**	-.16
Free State	**-0.34**	**-0.35**	**-0.20**	**-0.32**	**-0.37**	**-0.25**	**-0.23**	-.10	-.17
Kwazulu/Nat	**0.10**	.06	**0.16**	**0.17**	.10	.14	-.04	**-0.14**	-.04
North-West	-.03	-.07	**0.08**	.00	-.04	-.02	-.11	-.10	.09
Gauteng	**0.17**	**0.16**	**0.18**	**0.19**	**0.18**	.09	**0.09**	**0.14**	.02
Mpumalanga	-.04	**-0.18**	.07	.02	**-0.18**	.07	**-0.11**	.03	**-0.27**
North Prov	**0.09**	**-0.12**	**0.38**	**0.12**	-.13	**0.37**	-.06	.03	.013
Urban	**0.25**	**0.24**	**0.19**	**0.27**	**0.24**	**0.26**	**-0.14**	-.00	-.77
Race:									
Coloured	**0.17**	**0.16**	**0.15**						
Asian	**0.43**	**0.50**	**0.55**						
White	**0.79**	**0.88**	**0.94**						
Education:									
Shno_ed	.03	.06	**0.00**	.03	.03	-.02	.86	**1.84**	.04
Shprim	**0.07**	**0.12**	.08	.07	.11	.07	.45	.51	.08
Shsec	**0.37**	**0.30**	**0.41**	**0.29**	**0.28**	**0.34**	.17	**0.50**	.22
Shmatric	**0.43**	**0.29**	**0.54**	**0.53**	**0.38**	**0.61**	**0.29**	**0.20**	**0.40**
Shtert	**0.54**	**0.50**	**0.57**	**0.62**	**0.41**	**0.59**	**0.35**	**0.42**	**0.46**
Labour Mkt:									
Shwork	**0.68**	**0.71**	**0.62**	**0.79**	**0.68**	**0.74**	**0.51**	**0.68**	**0.27**
Shunemp	**-0.35**	**-0.38**	**-0.20**	**-0.28**	**-0.35**	**-0.15**	**-0.34**	**-0.51**	-.07
Shmig	**0.19**	**0.17**	**0.14**	**0.18**	**0.22**	**0.11**	.19	.18	-.09
Welfare:									
Shoap	**0.15**	.00	.05	**0.40**	**0.45**	.18	**-0.52**	**-0.28**	**-0.54**
Cons	**8.05**	**7.35**	**8.80**	**7.93**	**7.38**	**8.74**	**9.45**	**7.99**	**11.31**
	N=2858 3	N=2858 3	N=2858 3	N=1848 1	N=1848 1	N=1848 1	N=5224	N=5224	N=5224
	=0.41	*R2*=0.44	*R2*=0.36	*R2*=0.	*R2*=0.27	*R2*=0.33	*R2*=0.22	*R2*=0.25	*R2*=0.17

Note: Bold Coefficients are significant at the 1% level using unweighted sample data

Table 3: The Determinants of Household Poverty

	Contributions to the Probability of Being Poor (Probit Marginal Effects Estimates at Mean Values)				
	All Households	White	African	African Urban	African Rural
Head:					
Femres	**.0781478**	**.0219131**	**.1136566**	**.0373425**	**.157935**
Femabs	**.0497033**	**.0009678**	**.0745141**	**-.018173**	**.1414804**
Maleabs	**-.0442584**	No Poor HH	**-.0731561**	**-.0255452**	**-.0992155**
Composition:					
Kid7	**.0313683**	-.0000713	**.0487308**	**.0270438**	**.0541386**
Kid15	**.0271427**	.0000927	**.041356**	**.0325001**	**.0407004**
F16_59	**.0300596**	.0004054	**.0482446**	**.0333742**	**.0471798**
M16_59	**.0209923**	.0000933	**.0333056**	**.017946**	**.0366794**
Ad60	-.0040401	.0000364	.0025527	**.0123253**	**-.0050099**
Province:					
Eastern Cape	**.1273211**	.0000488	**.1081468**	**.0489737**	**.386569**
North Cape	**.1421202**	.0002972	**.1319067**	**.0425478**	**.3694423**
Free State	**.2431827**	-.0001416	**.2410857**	**.0827216**	**.4602548**
Kwazulu/Nat	-.030258	.0006555	**-.1136445**	**-.0783361**	**.1881312**
North-West	**.065554**	-.0004	.026282	**.0221716**	**.2995115**
Gauteng	**-.0529552**	-.0000628	**-.1514906**	**-.1175519**	**.2343683**
Mpumalanga	**.0409939**	No Poor HH	-.0023541	**.0693216**	**.2528469**
North Prov	-.0090761	No Poor HH	**-.0744267**	**-.0402356**	**.2264175**
Urban	**-.1129563**	**-.0021825**	**-.1653923**		
Race:					
Coloured	**-.060042**				
Asian	**-.1206235**				
White	**-.204393**				
Education:					
Shno_ed	-.0195861	No Poor HH	-.0370429	**-.1045847**	**.0118241**
Shprim	**-.0487069**	No Poor HH	**-.0716319**	**-.0946805**	**-.0431914**
Shsec	**-.1347915**	-.0007906	**-.2072631**	**-.1246659**	**-.2074922**
Shmatric	**-.1780758**	-.0003662	**-.2896607**	**-.1680733**	**-.3226574**
Shtert	**-.1150711**	-.0003629	**-.1861881**	**-.1513875**	**-.1750088**
Labour Mkt:					
Shwork	**-.2817501**	-.0012705	**-.4445696**	**-.3294419**	**-.4563091**
Shunemp	**.1094686**	.0008389	**.1611013**	**.1513047**	**.114812**
Shmig	**-.0614882**	No Poor HH	**-.1003269**	**-.0736301**	**-.0914207**
Welfare:					
Shoap	**-.1600474**	.0009599	**-.3029865**	**-.2021589**	**-.304651**
Cons					
	N=28 578	N=4 485	N=18 476	N=7 744	N=10 732
	Prob>chi2= 0.000	Prob>chi2= 0.000	Prob>chi2 =0.000	Prob>chi2 =0.000	Prob>chi2= 0.000
	PseudoR2= 0.39	PseudoR2= 0.31	PseudoR2= 0.29	PseudoR2= 0.32	PseudoR2= 0.21

Note: Bold coefficients are significant at the 1% level using unweighted sample data

Table 4: Contributions of Different Factors to Household Income Inequality, for All Households and by Race

	Shares of Total Inequality				
	All Households (Gini=0.638)	**White (Gini=0.466)**	**African (Gini=0.552)**	**African Urban (Gini=0.519)**	**African Rural (Gini=0.526)**
Head: Femres Femabs Maleabs	0.016	0.017	0.026	0.023	0.031
Composition: Kid7	0.050	0.023	0.055	0.047	0.06
Kid15	0.053	0.032	0.060	0.062	0.059
F16_59	0.017	0.000	0.021	0.026	0.018
M16_59	0.005	-0.004	0.003	0.004	0.003
Ad60	0.000	0.029	-0.001	0.004	-0.002
Province: Eastern Cape North Cape Free State **Kwazulu/Nat** North-West Gauteng Mpumalanga North Prov	0.030	0.019	0.033	0.046	0.023
Urban	0.045	0.006	0.047		
Race: Coloured Asian White	0.170				
Education: Shno_ed	-0.001	-0.001	0.000	-.004	0.002
Shprim	0.006	0.001	0.006	0.006	0.004
Shsec	0.061	0.006	0.041	0.035	0.028
Shmatric	0.093	0.081	0.060	0.074	0.036
Shtert	0.038	0.049	0.035	0.038	0.032
Labour Mkt: Shwork	0.100	0.082	0.147	0.174	0.128
Shunemp	0.022	0.005	0.021	0.037	0.016
Shmig	0.001	0.001	0.007	0.010	0.004
Welfare: Shoap	-0.001	0.021	-0.001	-0.003	0.002
Residual	0.296	0.633	0.443	0.421	0.557

Table 5: The Contribution of Explanatory Factors to Household Income Inequality for South African Households of All Races.

	All Households (Gini 0.52) (log variance 1.455) s(lnY)=1.206			
	Log Per Capita Income (OLS) (a_j)	Standard Deviation of Factor $s(Z_j)$	Correlation of Factor with LnY *Cor*[Z_j,lnY]	Contribution to Inequality (By 2.a.)
Head:				
Femres	**-.2701**			
Femabs	**-.1709**			0.016
Maleabs	**.1396**			
Composition:				
Kid7	**-.1655**	1.008	-0.360	0.050
Kid15	**-.1589**	1.071	-0.377	0.053
F16_59	**-.0649**	1.053	-0.297	0.017
M16_59	**-.0485**	0.980	-0.114	0.005
Ad60	.0071	0.635	-0.105	0.000
Province:				
Eastern Cape	**-.1616**			
North Cape	**-.2101**			
Free State	**-.3017**			
Kwazulu/Nat	**.0918**			0.030
North-West	-.0414			
Gauteng	**.1671**			
Mpumalanga	**-.0655**			
North Prov	**.0921**			
Urban	**.2386**	0.498	0.457	0.045
Race:				
Coloured	**.1653**			
Asian	**.4501**			0.170
White	**.8450**			
Education:				
Shno_ed	.0178	0.263	-0.304	-0.001
Shprim	.0764	0.294	0.345	0.006
Shsec	**.3642**	0.400	0.503	0.061
Smartic	**.5073**	0.376	0.587	0.093
Shtert	**.4203**	0.257	1.206	0.038
Labour Mkt:				
Shwork	**.6813**	0.374	0.472	0.100
Shunemp	**-.3260**	0.244	-0.331	0.022
Shmig	**.1737**	0.198	0.039	0.001
Welfare:				
Shoap	*.0748*	0.142	-0.057	-0.001
Cons	**8.0678**			
Residual	1.000	0.656	0.544	0.296

Bold Coefficients are significant at the 1% level using the unweighted sample data.

Table 6: The Contribution of Explanatory Factors to Household Income Inequality for African and White Households and the Contribution of Explanatory Factors to Differences in African and White Income Inequality.

		African Households (Gini 0.55) (log variance 0.98) s(lnY)=0.990				**White Households** (Gini 0.47) (log variance 0.65) s(lnY)=0.81				**Contribution to Differences Between African/White Inequality**	
		Log Per Capita Income (OLS) **(aj)**	Standard Deviation of Factor s(Z$_j$)	Correlation of Factor with LnY ***Cor*[Z$_j$,lnY]**	Contribution to Inequality **(By 2.a.)**	Log Per Capita Income (OLS) **(aj)**	Standard Deviation of Factor s(Z$_j$)	Correlation of Factor with LnY ***Cor*[Z$_j$,lnY]**	Contribution to Inequality **(By 2.a.)**	Gini **(By 3)**	Log-Variance **(By 3)**
Head:	Femres	**-.2556**				**-.3053**					
	Femabs	**-.1611**			0.026	**-.1742**			0.017	8	4
	Maleabs	**.1644**				.0767					
Composition:	Kid7	**-.1501**	1.072	-0.336	0.055	**-.2724**	0.645	-0.105	0.023	23	12
	Kid15	**-.1480**	1.125	-0.359	0.060	**-.2232**	0.760	-0.152	0.032	21	12
	F16_59	**-.0686**	1.124	-0.268	0.021	**-.0925**	0.668	0.003	0.000	13	6
	M16_59	**-.0494**	1.036	-0.050	0.003	-.0433	0.690	0.099	-0.004	4	1
	Ad60	.0123	0.618	-0.161	-0.001	**-.1528**	0.707	-0.215	0.029	-16	-6
Province:	Eastern Cape	**-.1121**				-.0733					
	North Cape	**-.2478**				**-.1729**					
	Free State	**-.2847**				**-.1825**					
	Kwazulu/Nat	**.1710**				-.0250					
	North-West	.0075			0.033	**-.0244**			0.019	10	6
	Gauteng	**.2092**				**.1196**					
	Mpumalanga	-.0090				-.1153					
	North Prov	**.1553**				.0586					
	Urban	**.2764**	0.490	0.343	0.047	**-.2052**	0.277	-0.083	0.006	27	13
Education:	Shno_ed	.0072	0.290	-0.207	0.000	.6962	0.031	-0.039	-0.001	0	0
	Shprim	.0791	0.319	0.235	0.006	**.2910**	0.039	0.053	0.001	3	2
	Shsec	**.3188**	0.395	0.321	0.041	**.4393**	0.087	0.119	0.006	23	11
	Smartic	**.5069**	0.296	0.396	0.060	**.4111**	0.406	0.392	0.081	-5	2
	Shtert	**.5269**	0.196	0.331	0.035	**.3252**	0.384	0.316	0.049	-4	1
Labour Mkt:	Shwork	**.7660**	0.366	0.518	0.147	**.4659**	0.393	0.362	0.082	49	27
	Shunemp	**-.2935**	0.264	-0.265	0.021	**-.3032**	0.126	-0.113	0.005	10	5
	Shmig	**.1644**	0.228	0.179	0.007	.3392	0.067	0.032	0.001	4	2
Welfare:	Shoap	**.345**	0.133	-0.025	-0.001	**-.5388**	0.165	-0.190	0.021	-12	-4
	Cons	**7.9384**				**9.4338**					

Residual		1.00	0.659	0.666	0.443	1.00	0.641	0.796	0.633	-58	7

Bold Coefficients are significant at the 1% level using the unweighted sample data.

Table 7: The Contribution of Explanatory Factors to Household Income Inequality for African Urban and Rural Households and the Contribution of Explanatory Factors to Differences in African Urban and Rural Income Inequality.

		African Urban Households (Gini 0.52) (log variance 0.96) s(lnY)=0.978				**African Rural Households** (Gini 0.53) (log variance 0.80) s(lnY)=0.897				**Contribution to Differences Between African Urban/Rural Inequality**	
		Log Per Capita Income (OLS) (aj)	Standard Deviation of Factor s(Zj)	Correlation of Factor with LnY Cor[Zj,lnY]	Contribution to Inequality (By 2.a.)	Log Per Capita Income (OLS) (aj)	Standard Deviation of Factor s(Zj)	Correlation of Factor with LnY Cor[Zj,lnY]	Contribution to Inequality (By 2.a.)	Gini (By 3)	Log-Variance (By 3)
Head:	Femres	**-0.204**				**-0.294**					
	Femabs	**-0.055**			0.023	**-0.231**			0.031	77	-2
	Maleabs	**0.177**				**-0.156**					
Composition:	Kid7	**-0.169**	0.928	-0.285	0.047	**-0.143**	1.144	-0.328	0.06	114	-2
	Kid15	**-0.184**	0.983	-0.334	0.062	**-0.134**	1.191	-0.334	0.059	-17	8
	F16_59	**-.071**	1.116	-0.324	0.026	**-0.062**	1.129	-0.238	0.018	-63	7
	M16_59	**-0.039**	1.024	-0.090	0.004	-0.052	1.043	-0.048	0.003	-6	1
	Ad60	-0.031	0.557	-0.213	0.004	**0.034**	0.650	-0.088	-0.002	-51	3
Province:	Eastern Cape	-.0519				**-.4852**					
	North Cape	-.1298				**-.6558**					
	Free State	**-.1610**				**-.6839**					
	Kwazulu/Nat	**.1881**				-.1776					
	North-West	.0422			0.046	**-.3501**			0.023	-196	17
	Gauteng	**.2341**				**-.2290**					
	Mpumalanga	**-.1569**				**-.2903**					
	North Prov	**.2308**				-.2056					
Education:	Shno_ed	0.105	0.226	-0.165	-.004	-0.039	0.315	-0.137	0.002	50	-3
	Shprim	0.125	0.255	0.198	0.006	0.062	0.341	0.154	0.004	-24	2
	Shsec	**0.310**	0.378	0.978	0.035	**0.306**	0.380	0.217	0.028	-56	7
	Smartic	**0.509**	0.350	0.405	0.074	**0.459**	0.238	0.294	0.036	-318	27
	Shtert	**0.455**	0.242	0.341	0.038	**0.665**	0.154	0.277	0.032	-54	7
Labour Mkt:	Shwork	**0.831**	0.356	0.576	0.174	**0.736**	0.361	0.432	0.128	-381	42
	Shunemp	**-0.384**	0.266	-0.356	0.037	**-0.229**	0.263	-0.243	0.016	-175	15
	Shmig	**0.175**	0.266	0.204	0.010	0.121	0.199	0.138	0.004	-51	4
Welfare:	Shoap	**0.288**	0.125	-0.079	-0.003	**0.378**	0.138	0.035	0.002	42	-3
	Cons										
Residual		1.000	0.635	0.649	0.421	1.000	0.669	0.746	0.897	1209	-29

Bold Coefficients are significant at the 1% level using the unweighted sample data

References

Bhorat, H. and Leibbrandt, M. 1999. “Correlates of Vulnerability in the South African Labour Market.” DPRU working paper 99/27.

Bourguignon, F., Fournier, M., and Gurgand, M. "Distribution, Development and Education:Taiwan 1979-1992." Mimeo. Paris: Delta, 1998.

Fields, G. 1998b. "Accounting for Income Inequality and its Change". Mimeo. Ithaca: Cornell University.

Glewwe, P. "Investigating the Determinants of Household Welfare in Cote d'Ivoire." *Journal of Development Economics* 35: 307-37.

Huppi, M. and Ravallion, M. 1991. “The Sectoral Structure of Poverty During an Adjustment Period: Evidence from Indonesia in the Mid-1980s.” World Development 19(12): 1653-1678.

Leibbrandt, M and Woolard, I. 1996 "Rural Labour Markets, Migrant Labour and Rural Poverty in South Africa." Report for the International Labour Office.

Leibbrandt, M., Bhorat, H. and Woolard, I. 1999. “Understanding Contemporary Household Inequality in South Africa.” DPRU working paper 99/25.

Ravallion, M. 1996. Poor Areas. In Giles, D. and Ullah, A. (Eds.) *The Handbook of Applied Economic Statistics*. New York: Marcel Dekkar.

Rousseeuw, P. and Leroy, A. 1987. *Robust Regression and Outlier Detection.* New York: John Wiley and Sons.

StataCorp. 1997. State Reference Manual (Release 5). College Station: Stata Press.

Willis, R. 1987. “Wage Determinants: A Survey.” In Ashenfelter, O. and Layard, R. *Handbook of Labor Economics*. Amsterdam: North Holland: 525-602.