ECONOMIC STUDIES

DEPARTMENT OF ECONOMICS
SCHOOL OF ECONOMICS AND COMMERCIAL LAW
GÖTEBORG UNIVERSITY
143

INCOME GENERATION IN THE AFRICAN AND COLOURED POPULATION – THREE ESSAYS ON THE ORIGINS OF HOUSEHOLD INCOMES IN SOUTH AFRICA

Sten Dieden

ISBN 91-85169-02-1
ISSN 1651-4289 print
ISSN 1651-4297 online

Mind the gap.

(Announcement often heard in the London Underground.)

To Jamie-Lee, Jean-Ray, and Caylin

CONTENTS

III Integration into the South African core economy– household level covariates…….……..1

ABSTRACT

The dissertation comprises three empirical analyses focusing on the origins of South African households' income sources. The income sources are categorized to reflect the households' varying extent of economic core sector integration. All three chapters contain analyses of household survey data from 1995 and focus on the African and coloured population.

Chapter one explores the extent to which inequality in a sample of African and coloured individuals can be attributed specific labour-market related characteristics of their households or household heads. The analyses apply the Theil-L measure of inequality to the distribution of a consumption bundle. The education level of household heads is the strongest single explanatory factor, followed by households' main income sources. The race, age categories, or gender of household heads do not account for large fractions of inequality in this sample.

Chapter two utilizes the income source categories to investigate whether inter-household variation in income sources can contribute to explain variation in income levels. The results from the estimation of three reduced form models are compared. All three models have households' log-income levels as dependent variables and share a set of household characteristics as explanatory variables. Two of the models are two-stage specifications that use provincial locations in the construction of instruments for income source categories. The third specification contains no income source variables but includes provincial locations as explanatory variables. The results show that, as compared to the specification with provincial locations, income sources can be incorporated as explanatory variables into multivariate regression analyses without considerable loss of explanatory power. Controls for endogeneity must however be applied. The partial impacts from income sources are statistically significant and their signs are in accordance with expectations. The results also suggest that households in different main income source categories differ systematically in their demographic and educational endowments.

In Chapter 3 the sample is restricted to the African and coloured households that have a main income source. The analyses aim to identify household level characteristics that are associated with differing extents of core sector integration. Two separate multinomial logit models are estimated, for urban and non-urban households respectively. The output from the analysis is utilized to compute the probabilities for a household having a main income source from either one of five categories. The results indicate that prominent covariates of low core-economy integration are low levels of education, female gender, and either old or young of working-age among the households' main income earners. A non-urban household's location in either a former "homeland" or in an agriculturally or commercially developed area also yields disparate implications for the main income source probabilities. The study also finds associations between main income sources and households' demographic compositions which are compatible with findings in previous research on endogenous household formation in South Africa.

Doubtlessly, varying extents of core-sector integration is very much an inter-racial phenomenon in South Africa. However, the analyses show that such integration varies considerably also within the previously disadvantaged population segment in South Africa. There, the varying integration displays impacts on household welfare that are separable from other characteristics, but integration appears also to be a mechanism partly determined by household or individual earner characteristics.

Other studies emphasize the importance of employment creation as a policy objective to combat South Africa's many challenges. The results here support the dire need for more employment, since the worst-off households often rely on non-labour income sources. However, even if employment would increase in South Africa's core economy within the foreseeable future, no guarantee exists that the currently most marginalized households would be integrated. For the latter to eventuate, demand for labour would have to extend into remote areas and encompass low-skilled, largely female, and young labour that may not currently afford to market their services. Geographically targeted efforts that upgrade the skills of the young and facilitate their access to the labour market are thus also favoured by the results of these studies.

ACKNOWLEDGEMENTS

While the usual disclaimers apply to faults and mistakes contained in this thesis, many individuals have been instrumental to my completion of the Ph.D. program. Quite naturally most lecturers, assistant teachers, seminar attendants, administrative staff and fellow Ph.D. students here in Goteborg have in some way contributed. I am deeply thankful to each one of you, but below I try to list those that have assisted in the most perceptible ways.

In completing these Ph.D. studies I have been blessed with not less than three mentors here in Goteborg, all of whom have read and commented wisely on several earlier drafts of the studies contained in this thesis. For this I am profoundly thankful. Also, their dedication both to research quality and to issues of social injustice all over the globe set standards to which I will always aspire. I am deeply indebted to my supervisor Arne Bigsten who has also held up his protective and financially generous guard for me, in face of the prolonged duration of my studies. Sincere thanks to Björn Gustafsson who made sure I could remain involved with South African issues from an early phase of my studies. It was also due to his very educating and pedagogical co-authorship that my name first appeared on an article in an international journal. Last but not least, my warm thanks to big-hearted Ali Tasiran. At the time when things looked most gloomy, he involved himself deeply, adeptly, and patiently in my research. He was always ready to assist.

I owe fond thanks also to Eva Jonason and Eva-Lena Neth for going out of their way to provide devoted assistance with a multitude of administrative issues, not least in connection with my applications for funding.

I am indebted to the Department for Research Cooperation of the Swedish International Development Cooperation Agency's (Sida (SAREC)) for providing funds for two years of my studies and thereby making it possible for me to conduct a lot of my research in South Africa. I am most thankful for funding also from the Centre for Social Science Research (CSSR) of the University of Cape Town (UCT) as well as from the Swedish Council for Working Life and Social Research (FAS).

The journey to this thesis started more than ten years ago, when I first came to South Africa. Innumerable people and events affected my interest in South African circumstances during the time I spent there before starting my Ph.D. studies in August 1998. Below I am going to limit the thanks those to who have contributed after I commenced my Ph.D. studies. However, the journey would never have started if it were not for a small group of South Africans, all of whom I hold in high regards and owe special thanks. Firstly, I am profoundly grateful to Dudley Horner and Francis Wilson, who arranged my unforgettable and highly educating four year spell with their unique Southern Africa Labour and Development Research Unit (SALDRU). Francis Wilson was also kind enough to write a letter of recommendation to the Ph.D. program in Göteborg and Duldey Horner mediated a request for the same from Guy Standing, who benignly met that request.

When I first arrived at SALDRU, Ingrid Woolard taught me the basics of household survey data analysis and at a much later stage also provided sharp comments to what eventually developed into research included in this thesis. My journey with the SALDRU train* also risked coming to a sudden stop before it really commenced - or may later have derailed - if it were not for Isabella Scholtz of UCT's International Office, who joined our side in the lingering struggle over research permit extensions with the South African Department of Home Affairs. When I first considered taking up Ph.D. studies, I benefited tremendously from several exchanges with Carol Kaufman (who is free from all responsibility since she tried to discourage me). Our connection was partly a positive externality from the commencement of a successful collaboration between SALDRU and the University of Michigan's Institute for Social Research. I owe thanks to David Featherman and David Lam for an educating spell there.

Special thanks are also directed to the current director of SALDRU, Murray Leibbrandt, for lucid comments to my research, for his encouragement, and for the logistical arrangements provided in his capacity as director of UCT's School of Economics. I owe deep gratitude also to Nicoli Natrass and Jeremey Seekings of the

* In those days, the aircraft carrier was a more commonly applied metaphor for SALDRU.

CSSR and its Social Surveys Unit, both of whom have been very generous, hospitable and encouraging. The CSSR provided me with both office space and a computer at UCT, as well as gave me the opportunity to publish working papers and present at the centre's seminar series. In the same connection, innermost thanks to my friend Director Haroon Bhorat of the Development Policy Research Unit, who let me present early versions of two of the papers in this thesis at the outstanding conferences arranged by his unit.

I am also indebted for favours, assistance, comments, and input into my research from the following colleagues and friends: Brenda Adams, Gadija Allsion, Anthony Black, Servaas Van der Berg, Vani Borooah, Debbie Budlender, Liz Coetser, Stefan Dercon Libby Downes, Faldie Esau, Johan Federke, Lennart Flood, Kathy Forbes, Liesl George, Trudy Hartzenberg, Rasmus Heltberg, Lennart Hjalmarsson, Stefan Klasen, Gunnila Leander, Mats Lundahl, Paul Lundall, Julian May, Anna McCord, Bernadette Ontong, Laura Poswell, Alison Siljeur, Munacinga Simatele, Guy Standing, Allsion Stevens, Donald Storrie, Liv Tørres, Matthew Welch, Martin Wittenberg, and Lynn Woolfrey.

A group of former co-students who are now at various stages of prominent academic careers have provided most helpful input of various sorts. Thanks very much to Martin Linde-Rahr, Måns Söderbom, Kerem Tezic, and Anthony Wambugu. In this respect I place two individuals in a class of their own; Ever-friendly Fredrik Carlsson who never turned down a request for assistance and my very dear friend Dinky Daruvala, the support and encouragement from whom has been truly invaluable. Also, in a special category of her own, very warm and special thanks to Oana Arnautu who has assisted and encouraged me unrelentingly over time and across space.

To the group of current co-students and friends we call "La famiglia": you all know who you are, for what and how very much I owe you - molto grazie. In that group, I would like to especially acknowledge Martine Visser who has shared the experience of a Ph.D.-student existence divided between Cape Town and Gothenburg. She has been very hospitable, always encouraging, and commented wisely to early drafts of my analyses. In the latter respect, I also direct very warm thanks to my dear friends James Hodge, Lynn Bowie, Asha Bowie-Hodge and her little sisters Anna and Ella,

for making me feel as part of their family. I will never forget the wining and dining, the coffee breaks, the story reading, the running, and the insightful comments to my work. For similar reasons I also owe deep gratitude to Sean Chandler, my equally close chum in Cape Town, from whom no favour asked has ever been too big to grant. Also in this regard, heartfelt thanks yet again to charmingly grumpy, dear, and unmatched Dudz who has been there for me for so long and in so many respects.

Deeply felt thanks also to Scandinavian friends, family and relatives: Ulrika Callstam; Andreas Dahlén; Per Ejnelind and Joseph Daruvala-Ejnelind; Henrik, Ulla Cecilia, and Cristina Dieden; Carina Gråbacke; Gunnar, Inger, Per, and Jesper Ideberg; Thomas, Kerstin, Tove and Kajsa Johansson; Jan Jörnmark; Geir and Kathrine Kjærstad; Rolf and Kerstin Larson; Knut, Eva, Emma, Marit, and Kjetil Løyland; Nils and Lisa Mörner; Nigel, Eva, Tom, Annika, and Alexander Olesen; Marta, Per, and Jan Rosenberg; Sven Silvander; Bjørn Skjellaug and Ulrik Bjørnson Arneberg; and George Stade. Fond thanks are long overdue and meagre recompense for the support and understanding from my father Johan, my brother Peter, and from my American family Harvey, Sheila, Dena, Andrea, Stuart and Kara Bjørnlie. Many of you have received questionable reciprocity in commitment from me over the last six years, yet all of you have brought me much-needed light, at crucial points in time through this period.

Last but not least, thank God this is over.

Sten Dieden

Göteborg, one dark evening in late December 2004.

INTRODUCTION

The three essays of this thesis all fall within the field of Development Economics. Investigations in this field often focus on the nature and causes of the lamentable living standards among the large fractions of the human population in the third world. In this discipline, concepts like “poverty”, “welfare”, “deprivation” and “well-being” are commonly applied, but their precise interpretation is subject to intense academic debate. There appears, however, to exist some consensus that irrespective of which of those terms is applied, income is one crucial dimension. Partly for the latter reason, all three of the impending analyses involve the origins of household incomes in South Africa, specifically in the country’s African and coloured population.[1]

South Africa’s first democratic election, in April 1994 marked the formal and longed-for end of the atrocious apartheid regime and the gory struggle against its racially discriminatory policies. However, in the wake of the apartheid era the living standards among the country’s non-white population were in many cases abysmal. Analyses of baseline microdata collected in the last quarter of 1993, attest to dire poverty, extreme inequality, high unemployment rates, and widespread lack of access to many basic facilities among non-white households. The analyses in this study utilize household data from 1995. But in course of the eighteen months passed since the country’s democratization, a reasonable observer should hope for little detectable reversal of circumstances generated over decades - in some cases centuries - of oppression.

The justification for the focus on the African and coloured population in this thesis is repeated in each essay along with other contextual issues. In summary, however, this limitation of scope is due to three circumstances; Firstly, these population groups jointly constitute approximately 85 percent of the South African population. Secondly, the groups encompass virtually all individuals at the bottom of the country’s income distribution. Finally, these population groups share similar historical

[1] Apartheid policies defined four main “racial classifications”; African, coloured, Asian/ Indian and white. The discrimination by race ran through all aspects of life and had tremendous effects on everyone’s living standards. For these reasons official statistics in South Africa still apply “racial” categories. The delimitation of the scope for all analyses in this thesis is founded on these categories.

legacies of racial discrimination from the apartheid era, although in some aspects the extent of discrimination may have varied.

The latter legacies would encompass those from *inter alia* racially biased settlement rights, rights of landownership, access to education, and restrictions on upward mobility in the labour market. Consequently, when analyzing deficient living standards in South Africa, many commonly applied explanatory factors are so closely intertwined with race characteristics, that separating out the formers' impacts from those of race becomes a serious methodological challenge. The restriction of the scope to a sample subject to similar legacies thus facilitates the identification of impacts from explanatory factors other than the unanimously associated race characteristics.

Income generation in the target group for the analyses in this work diverges from what is common on the rest of the African continent. The divergent features are firstly the very high extent to which livelihoods are generated in urban areas. Secondly, the very small contribution to household income from small-scale agriculture also in rural areas is also atypical to the continent. (Reardon (1997)) Finally, one finds a very widespread dependence on transfer incomes among both rural and urban South African households (Jooma (1991)). These features find their roots in the historical legacies from the "reserve" settlement practices that were first applied by the colonial powers in South Africa and later fortified by the apartheid regime.

The institutionalization of race discrimination prohibited landownership among the non-white population and barred Africans from settlement in urban or commercial agricultural areas, except for individuals issued with employment contracts. The African population was confined and sometimes forcedly removed *en masse* to designated areas, in which migrant employees would be reunited with their families during scarce leaves or at the end of their work contracts. These areas – the former "homelands" or "Bantustans" - were mostly but not exclusively of a rural nature. However, "rural" should not be confused with "pastoral" here. Even in the few cases where agricultural conditions initially were present in the former "homelands", they soon deteriorated due to vast overpopulation. Despite the relaxation of migration and settlement regulations already in the late 1980s, one still finds very high rural unemployment rates and pervasive poverty in many different shapes in these areas.

One benefit from restricting the sample to the African population is that one may get a clearer impression of which factors allow some of those worst-off in the population to overcome historical legacies, while others that face similar circumstances do not. However, while South Africa may have the highest level of urbanization in Sub-Saharan Africa, the country's African population is not nearly as urbanized as the other population groups. Good reasons exist to believe that educational facilities and labour market access, as well as many other factors that affect living standards, vary between urban and rural locations and thus both within and between the African and coloured population.

The notion of disparate conditions for income generation in urban and rural areas of South Africa enters from the explanatory side in the first two studies, but it is explicitly considered only in the last part of the thesis. Founded on the circumstances briefly summarized above, the analyses in this thesis approach income generation from a core-periphery perspective. Fundamental to all three analyses is a categorization of households' income sources according to the degree of core sector integration reflected by the income sources. This is categorization borrows features from a previous breakdown of the South African economic sectors into the "core economy", the "marginal modern economy" and "the peripheral economy". Slightly different versions of this categorization are applied in all three studies of the thesis.

The results from the fist paper show that, while the education level of household heads is the strongest single explanatory factor, households' types of income sources can explain almost as much welfare inequality in the sample as can education.

An application of the same categories in the second study shows that income sources from different categories have different, partial impacts on income levels when other household characteristics are controlled for. However, households' endowments of these other factors appear to vary systematically between the various categories.

The results from the analysis in the third paper suggest that households' allocation into a subset of the income source categories can to some extent be explained by a set of household characteristics, including its location and the gender, age and education levels of households' income earners.

Thus, one the hand, it is suggested that core integration is mechanism partly via which household characteristics affect welfare. However, indications are also that differing extents of such integration, as conceptualized by different types of income sources, have impacts separable from other characteristics. Many previous studies emphasize the importance of "employment creation" and list it among implied policies to combat South Africa's many challenges. The results here support the dire need for more employment, since the households' that are worst off often rely on non-labour income sources.

In this author's view, employment creation is indeed an overriding policy *objective* for South Africa. However, the phenomenon is itself not a "policy" open to "recommendation". Increased employment remains an outcome of complex interactions between many different factors, among which appropriate policies would be one. Irrespective of this – perhaps linguistic - objection, two other qualifiers must be raised based on the results of these analyses. Firstly, not all types of employment are equally beneficial to households. Secondly, if large-scale employment-increases would take place in South Africa's core economy within the foreseeable future, no guarantee exists that the currently most marginalized households would be integrated. For the latter to eventuate, demand for labour would have to extend into remote areas and encompass low-skilled, largely female, and young labour that may not currently afford to market their services.

Hence, even when controlling for historical legacies that disfavour the African and coloured population group, a core-periphery approach to income generation seems enlightening. The divide between the core and the periphery in post-apartheid South Africa appears to be partly a matter of physical distance and the financial costs implied in travelling such distance. However, the mechanisms which would facilitate the traversal of that gap need overcome legacies that operate in conjunction with distance via personal characteristics, such as age and gender, and qualifications. Pertaining to the latter, the results from all three analyses attest to the beneficial effects of education. Geographically targeted efforts that upgrade the skills of the young and facilitate their access to the labour market are thus favoured by the results of these studies.

Chapter I:

Ahoy the Good Hope? Some bearings and signals in seldom-navigated waters - on inequality in South Africa's coloured and African population.

Sten Dieden*

Abstract

Previous studies have decomposed South African income inequality into inequality between and within the population groups through which the apartheid regime operated racial discrimination. While a substantial fraction of total inequality can be attributed to differences in mean income levels between those population groups, the level of inequality within the racial groups has been found to be a larger contributor. Yet few investigations have attempted to elucidate inequality within these population groups. This study therefore explores the extent to which inequality in a joint sample of African and coloured individuals can be attributed specific labour-market related characteristics of their households or household heads. The analyses apply the Theil-L measure of inequality to the distribution of a consumption bundle in a household survey data set from 1995. The education level of household heads is the strongest single explanatory factor, followed by households' main income sources. The race, age categories, or gender of household heads do not account for large fractions of inequality in this sample.

* Department of Economics, University of Gothenburg, Sweden. sten.dieden@handels.gu.se.
While any defects or shortcomings in this work are entirely my own responsibility, I am indebted for very valuable comments to previous versions of this work to Renato Aguilar, Arne Bigsten, Björn Gustafsson, James Hodge, Ali Tasiran, and participants of the seminar series given by University of Cape Town's Centre for Social Science Research (CSSR). The financial provision by the Swedish International Development Cooperation Agency (Sida) and the CSSR's Social Services Unit is thankfully acknowledged. I also thank the CSSR for hosting me for during much of the time spent on research on this work.

1 Introduction

The apartheid regime in South Africa recognized and applied different extents of racial discrimination to four different population groups.[1] Due to *inter alia* the legacies from those discrimination policies, South Africa has the seventh highest level of inequality in the world (World Bank (2004). Previous studies have applied those racial groups in additive decomposition analyses and found that substantial fractions of the inequality is attributable to differences in average income levels between those groups (Whiteford and McGrath (1998, 2000), May (2000)), Leibbrandt, Woolard and Bhorat (2000), Leibbrandt and Woolard (2001)).[2] The fraction of total inequality originating from *within* the racial groups is however always found to be larger. Yet few investigations have analysed inequality within the racial groups.[3] This study of household survey data therefore explores the extent to which inequality within in a joint sample of African and coloured individuals can be attributed specific characteristics of their households or household heads.

Several previous studies on South Africa emphasize the importance of households' access to employment and *wage income* in explaining income inequality and in evading poverty (Carter and May (1999), Leibbrandt, Woolard, and Woolard (2000), Leibbrandt, Woolard, and Bhorat (2000), van der Berg (2000), Jenkins and Thomas (2000)). The analysis undertaken here utilizes a household's main income source as an explanatory factor that reflects a household's labour market attachment. The other explanatory factors are geographical location in the rural-urban dimension and in

[1] The four main "racial classifications" recognized by the apartheid regime were: African (black), coloured, Asian/ Indian and white. The discrimination by race ran through all aspects of life and had tremendous effects on everyone's living standards. For these reasons official statistics in South Africa still apply "racial" categories. The same practice is followed here and the categories will interchangeably be referred to as "population" or "racial", "groups" or "categories".

[2] In the literature on additively decomposable income inequality, the inequality attributable to differences between mean incomes of population subgroups' is considered "explained". By definition, the total level of inequality is reached by the adding to the former, the summed inequality in the income distribution around the means *within* each subgroup (Bourguignon (1979), Shorrocks (1980, 1984), Blackorby, Donaldson and Auersperg (1981), Cowell and Jenkins (1995)). Other measures of the centre of income distribution than the arithmetric mean can also be used.

[3] Two relevant exceptions are Leibbrandt, Woolard and Woolard (2000) and Leibbrandt and Woolard (2001). The first work decomposes the Gini index in the African population group and in the second a multivariate technique developed by Fields (2003) is applied to decompose the variance in households' log per capita income in the same population group.

provinces (Leibbrandt and Woolard (1999)), and the education level and racial group of household heads, which are all commonly applied determinants in the literature on individuals' wage earnings (Moll (1996), Kingdon and Knight (1999), Mwabu and Schultz (2000)).

The decomposition methodology utilized in this study is commonly applied and has been developed by Bourguignon (1979), Shorrocks (1980, 1984), Blackorby, Donaldson and Auersperg (1981), and Cowell and Jenkins (1995). The empirical analysis uses the Theil-L index of inequality, which allows total inequality to be unambiguously split into the contribution due to differences *between* subgroups and the contribution due to inequality *within* subgroups (Shorrocks (1984)). The units of observations are individuals, to which their households' per adult-equivalent expenditure levels have been attached.

The paper proceeds from here to discuss the methodology and justify the choice of index in Section 2. In Section 3 the data and choice of welfare measure are introduced. Thereafter Section 4 justifies the scope of the study and describes the sample delimitation process. The relationship between each explanatory variable and welfare levels are illustrated with descriptive statistics in Section 5, after which follows the empirical results in Section 6. Conclusions are drawn in the final Section 7.

2 Methodology

Given some measure of welfare a decomposition of South Africa's welfare inequality presupposes a picture that measure's distribution, as a readily expressible function of the inequality *between* certain groups and (some aggregation of) the inequality *within* the same groups. A wide range of measures of inequality exist (e.g. Champernowne and Cowell (1998)), but standard methodology draw on results from Bourguignon (1979), Cowell (1980) and Shorrocks (1980,1984) [4], from which it can be concluded that the most suitable measures are those ordinally equivalent to the measures in the general entropy class.

[4] See e.g. Tsaklogou (1993) for an early application and Gustafsson and Li (2002) for a more recent.

Denoting a measure of the latter class G(·), the related ordinally equivalent group of measures can be defined as:

$$(1) \qquad I(x) = C[G_\alpha(x), n(x), \mu(x)]$$

where $n(\cdot)$ is the finite dimension of - i.e. the number of observations in - the vector x that represents the welfare distribution of which $\mu(\cdot)$ is the arithmetic mean. The function $C[\cdot]$ is a cardinalisation of $G(\cdot)$. The latter increases monotonically in the first argument and is defined by

$$(2) \qquad G_\alpha(x) := \frac{1}{\alpha^2 - \alpha}\left\{\frac{1}{n(x)}\sum_{i=1}^{n(x)}\left[\frac{x}{\mu(x)}\right]^\alpha - 1\right\}$$

where the parameter α can be assigned any real value. For high positive levels of α, the index is sensitive to welfare changes in the upper level of the distribution and for an index which is sensitive to redistributions at the lower level, α should take on a negative value. Shorrocks (1984) shows that among the additively decomposable inequality measures, the index derived from α = 0, Theil-L, index, is the most satisfactory and allows total inequality to be unambiguously split into the contribution due to differences between subgroups and the contribution due to inequality within each subgroup.

Given a set of S groups the decomposition of $G(x)$ is undertaken using the group-means in the following general manner:

$$(3) \qquad G(x) = G_W + G_B$$

where

$$(4) \qquad G_W := \sum_{s=1}^{S} u_s^\alpha w_s^{1-\alpha} G_s(x_s)$$

and

$$(5) \qquad G_B := G(\mu_s)$$

where , μ_s is the mean welfare level in group s, with u_s and w_s as the income and population shares of group s respectively, and μ is the vector of S group means. [5]

The units of observation, subject to inequality, in this case are human beings. From that perspective the $G_0(x)$ index has the appealing feature that inequality within each group is weighted by the population fraction in that group. Hence, the index is deemed the most suitable for the purposes of this analysis.

When α = 0, l'Hôpital's rule yields

$$(5') \qquad G_B = \frac{1}{w_N}\sum_{s=1}^{S} w_s \ln\frac{\mu}{\mu_s}$$

and

$$(4') \qquad G_W = \frac{1}{w_N}\sum_{s=1}^{S}\sum_{i=1}^{n(s)} w_i \ln\frac{\mu_s}{y_i^s}$$

where w_i is the sampling weight attached to observation i and

$$w_s = \sum_{i=1}^{n(s)} w_i \quad \text{and} \quad w_N = \sum_{s=1}^{S} w_s$$

Cowell and Jenkins (1995) derive a measure of explained inequality, R_B, as

$$(6) \qquad R_B := \frac{G_B}{G} = 1 - \frac{G_W}{G}$$

The same authors show that this concept of explained inequality extends to the analysis of more than one determinant of inequality at the time, since through the

[5] In the inequality literature is $G_{\alpha=0}()$ and $G_{\alpha=1}()$ are also commonly and respectively referred to as Theil-L and Theil-T measures of inequality. It follows from equation 4) that the two measures differ in how total within-group inequality is computed. When the $G_0(x)$ is applied each group's inequality is weighted by the population fraction in that group, whereas when the $G_1(x)$ is applied the weights constitute each group's share of total expenditures.

specification of a refined subpartition, $\Pi_{a \text{ and } b}$, of an original partition Π_a or Π_b, it must be true that

$$R(\Pi_{a \text{ and } b}) \geq R(\Pi_a) \qquad \text{and} \qquad R(\Pi_{a \text{ and } b}) \geq R(\Pi_b)$$

Thus, a succession of subpartitions yields a consistent representation of the importance of the characteristics that define the consecutive subpartitions.

3 Data and choice of welfare measure

Data

In 1995 Statistics South Africa undertook its annual October Household Survey with questionnaire-based interviews on a wide range of living standards issues using a stratified and clustered sample of 30 000 households, representing all households in the country and containing nearly 131 000 inhabitants (the "OHS" sample). Two months later 28 585 of the same households were revisited in a more detailed Income and Expenditure Survey (the "IES" sample and henceforth the surveys or samples will jointlybe referred to as the "OHS/IES 95" data.)[6]

In the surveys a household is defined by "a person or a group of people dependent on a common pool of income who normally occupy a dwelling unit or a portion thereof and who provide themselves with food or the necessary supplies or arranged for such provision." A member resides four nights a week in the household. The sample for the surveys was stratified by province, urban and non-urban area and population group. Altogether, 3 000 enumerator areas (EAs) were drawn as primary sampling units, within each of which ten households were visited. The data concerning households were weighted by the estimated number of households in each stratum. (Statistics South Africa (1997)). The analyses in this study of a subsample of the full OHS/IES95 are conducted with the supplied household weights renormalized to sum

[6] At the time of the writing of this essay a similar, nationwide South African data set from year 2000 had been released. However, since the reliability of the 2000 data was also still under evaluation by South Africa's Statistics Council and since also the other analyses in my thesis are undertaken on the 1995 data, the latter was deemed preferable to the current analysis. One reason for the disputed comparability of the two data sets, is that nominal incomes were lower in the four lowest per adult-equivalent income quintiles in year 2000 than 1995, whereas the total expenditure data do not display such characteristics (see Tables A1 and A2). Some of the indications of the robustness of the core results of this analysis are based on the 2000 data and provided in Appendix 1.

to unity, as suggested by Deaton (1997) when faced with missing survey data. It should be noted that, given nine provinces in South Africa, two types of areas and four population groups, the full data sets may be considered representative of 72 groups, 36 of which are African or coloured. The inference to a population level of results based on partitions into larger numbers of groups is thus limited.

Welfare measure

Welfare is a complex phenomenon that involves multiple dimensions of deprivation and lack of goods and services is yet one of those dimensions (Sen (1985, 1987)). Even so, there is a good deal of consensus on the value of using a consumption aggregate as a welfare metric of living standards (Deaton and Zaidi (2002)). This study follows that tradition and uses a consumption-aggregate based on household expenditure data as a summary measure.[7] The aggregate is constructed according to the guidelines put forth by Deaton and Zaidi (2002) and contains the summed subtotals of household expenditures in 1995 Rand on the following categories as defined by Statistics South Africa (1997): food, beverages, tobacco, personal care, fuel and power, household operation, housing costs, remuneration for domestic workers, footwear, clothing, medical care, transport, telecommunication, education, and reading material.

Individuals are used as units of analysis rather than households since it is difficult to conceive of households experiencing welfare, rather than the individual members (Deaton (1997)). Attached to each sampled individual is its household's total annual expenditure on the categories in the consumption aggregate, divided by the household size as calculated in terms of adult-equivalence, yielding the households "per-adult-equivalent expenditure".[8]

[7] A common justification for the use of consumption is that current consumption is a function of permanent income (Slesnick (1993), but as pointed out by Deaton (1997), the empirical support for the permanent income hypothesis is at best mixed. See Slesnick(1993, 1998) and Chaudhuri and Ravallion (1993) for discussions of the choice between income or consumption as welfare metrics.

[8] Leibrrandt and Woolard (2001) investigate several adult equivalence scales for South Africa using the OHS/IES95 data and impacts appear to be miniscule. The authors proceed using the scaled applied by May, Carter and Posel (1995) i.e.: $E=(A+0.5K)^{0.9}$, where E is number of adult equivalents, A number of adults and K is the number of children 15 years old or younger. This study applies the same procedure. Information about the quintiles for the full samples based on this concept is displayed in Table A3 in Appendix 1.

The assumption of equal division among household members (whether in the format of adult equivalents or not) is as shown by e.g. Haddad and Kanbur (1990) questionable. It is also recognized that that welfare measures in some respect are too limited and, as developed in Sen's (1985, 1987) work, other indicators such as life expectancy, infant mortality and literacy would be better (Ravallion (1996). However, both these latter issues are beyond the scope of this investigation.[9]

4 Target group and sample delimitation

Target group

This study aims to find explanations for inequality among individuals in the African and coloured population. The two population groups are defined as the individuals that live in households where the head belongs to either the African or coloured race group. The objective of the study is justified by the figures in Table 1, which shows the distributions of all individuals sampled by the OHS/IES95 by per-adult-equivalent expenditure quintiles and population group.

Table 1 Composition of per adult-equivalent household expenditure quintiles in the full OHS/IES95 sample, by population group.

Quintile	African	coloured	Asian/ Indian	white	All	Quintile cut-off Points (1995 Rand)	
						Lower	Upper
1	96.7	3.2	0.0	0.1	100.0	62	1 496
2	92.7	7.2	0.1	0.1	100.0	1 496	2 468
3	86.9	11.4	1.1	0.6	100.0	2 468	4 139
4	73.1	15.8	4.5	6.7	100.0	4 139	9 313
5	29.4	8.0	7.7	54.8	100.0	9 313	760 069
All	75.8	9.1	2.7	12.5	100.0	62	760 069
Number (millions)	30, 0	3, 6	1,1	4,9	39, 7		
Total expenditure share (%)	43.6	8.4	5.3	42.7	100.0		

Source: OHS/IES95, own computations, weighted figures. Absolute population numbers are weighted sample estimates in millions of individuals. n =125 112

[9] See Klasen (1997, 2000) for two multi-dimensional approaches to deprivation in South Africa.

As can be seen, African and coloured individuals constitute more than 95 percent of the individuals in the three lowest brackets, while the corresponding shares of the total population is approximately 85 percent. The summed expenditures of the African and the coloured subsamples however, amount to just over 50 percent of the total. At the same time, the white population fraction is miniscule in the three first quintiles, in the neighbourhood of seven percent in the fourth and only becomes substantial in the highest quintile. While the Indian population fraction is small, the group is over-represented in the highest brackets. Similarly, the Indian/Asian fraction of total expenditure is twice as large as its population share, whereas the expenditure share of the white sample is more than three times as large as its population share.

A closer investigation of inequality within the African and coloured subsample is warranted for at least three reasons; Firstly, the subsample represents the *overwhelming majority* of South Africans and virtually *all* individuals at the lowest end of the expenditure distribution. Secondly, the members of these groups face similar historical legacies. The identification of the factors that are associated with inequality within that sample may thus provide some insight into the nature of inequality at the lower segment of the expenditure distribution, where policy measures to reverse past injustices are most needed. Finally, using the same data, Leibbrandt, Woolard and Bhorat (2001) report the contributions to the total level of inequality in households' per-adult equivalent income from inequality within those groups to 56.8 percent as measured by the same inequality measure. Hence, explained fractions of within-group inequality in the subsamples will add considerably to the total explained inequality in South Africa.

Sample delimitation

For the purposes of this study and for the above reasons, only individuals that live in households where the head belongs to the African or coloured population groups were selected. Furthermore, for reasons which are motivated in the next section, the origins of households' main sources of income by *inter alia* broad economic sectors are to be used as explanatory characteristics for inequality. Since the quality of the information on individuals' labour market characteristics were greater in the OHS module than in the IES, it was deemed desirable to extract that information from the former base.

Households in the two data sets are easily matched, since their unique codes were identical in both data sets. However, the within-household codes for individuals differed across the surveys. Persons that were captured with any amount of income in the IES module therefore had to be matched to the OHS data according to household, age, gender and race. By this procedure 97.5 percent of the utilized sample were matched. Another 773 earners was identified by allowing either age to mismatch by two years, with race and gender matching perfectly, or race to have been miscaptured, with age and gender matching perfectly. This procedure yielded in all 30 906 earners identified in both data sets. The sample delimitation process is illustrated in Table 2. All results in the remainder of the analysis are weighted figures, based on the 86.5 percent (92 717) of individuals that resided in households that met the first criterion and where all income earners covered by the IES module were identified in both data sets..

Table 2) Sample delimitation process

Sample	Number of individuals	Share of total revisited sample	Share of revisited African and coloured sample
Total OHS/IES sample	125 112	100.0	
African and coloured OHS/IES sample	107 229	84.9	100.0
Final sample	92 717	74.1	86.5
Source OHS/IES95, own computations, unweighted figures.			

Quintiles based on per adult-equivalent expenditures in current Rand were designed for this sample and information about the expenditure in each quintile is presented in Table 3. A first impression of the welfare inequality in this sample is given by the ratio of the average expenditures in the *fourth* quintile to the first being 4.3, while the corresponding ratio is 11.4 for the *fifth* and first quintiles. Hence, the most distinct change in expenditure levels occurs between the two highest quintiles. In absolute terms the within-quintile expenditure span is by far the largest in the fifth. However, the range of relative expenditures is just slightly wider in the fifth quintile than it is in the first, with ratios of the highest to lowest expenditure at 23.3 and 22 respectively. The relative ranges are considerably narrower in the other three quintiles.

Table 3) Mean, minimum and maximum per adult-equivalent expenditure, by quintile (1995 Rand)

Quintile	Mean	Min	Max
1	984	62	1 369
2	1 738	1 369	2 141
3	2 646	2 141	3 241
4	4 248	3 241	5 663
5	11 255	5 663	133 037
All	4 174	62	133 037

5 Descriptive statistics and partition-defining characteristics

Given the importance of households' access to employment and *wage income* detected in previous research on South African inequality (Leibbrandt, Woolard, and Woolard (2000), Leibbrandt, Woolard, and Bhorat (2000), van der Berg (2000), Jenkins and Thomas (2000)), this study applies partitions into subcategories along, on the one hand, households' main income source category (henceforth "Main income source") - as a reflection of its labour market attachment – and, on the other, characteristics that are commonly used determinants for individuals' wage earnings (Willis (1986), Moll (1998), Kingdon and Knight (1999), Mwabu and Schultz (2000)).

While the concept of a household head is non-trivial, the definition used by Statistics South Africa enumerator's manual for the October Household Survey is applied: a head of household can either be male or female, and is the person who assumes responsibility for the household (Budlender (1997)). Assuming implicitly that the head is a significant earner of income, the implied determinants for the households wage or non-wage earnings characteristics are proxied by the population group, highest educational achievement, gender, and age category of the household head, as well as the household's location in rural or urban areas and province of residence (henceforth "Race", "Education", "Gender", "Location" and "Province" when referred to as explanatory variables).

As discussed by Leibbrandt and Woolard (2001), one can expect variables of this nature to "move together" in the South African setting. The reasons for the

presumably high degree of correlation are found in the historical legacies of racially discriminatory practices which span across areas such access to education, labour market regulations, migration, settlement and rights of landownership (e.g. Wilson and Ramphele (1989)).

Hence, a high degree of overlap in fractions of explained inequality by these characteristics would be expected in samples containing *all* the South African population groups. For an impression of the extent to which one can expect race to be of individual significance as an explanatory variable in this sample the reader is referred to Table 4. The table shows that the fraction of coloured individuals constitutes less than ten percent throughout the third quintile and is just over 20 percent in the fifth. At R 6 253 per month, the mean expenditure in the same population group is some 50 percent higher than that of the African at R 3 920. Brief introductions to of each of the five remaining explanatory factors for inequality are introduced below, with descriptive statistics that serve to justify their application.

Table 4) Percentage-wise composition of sample quintiles, by Race.

Location	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	All	Mean expenditures
African	97.2	93.5	90.1	84.9	79.9	89.1	3920
Coloured	2.8	6.6	9.9	15.1	20.1	10.9	6253
Total	100.0	100.0	100.0	100.0	100.0	100.0	4174

South African households' income sources

The South African literature usually distinguishes, by one set of labels or another, between at least four broad groups of household income sources: private transfers, public transfers, self-employment, and wage income (e.g. Carter and May (1999), Leibbrandt, Woolard, and Bhorat (2000)). According to Leibbrandt, Woolard, and Bhorat (2000), income generation processes differ above and below the poverty line, in that the contributions of *wages* to total income are smaller among the poor and vice versa for remittances and state transfers. One conclusion made by the authors is that wage income is central in the determination of both poverty status and poverty depth. On the same note Bhorat (2000) shows that households with earners that are exclusively either domestic workers or agricultural workers have relatively high poverty propensities. Also of high relevance to this study, van der Berg (2000) shows that the shares of remittance income decline in higher income-consumption quantiles

and wage-income shares increase, both in general and as households' main sources of income.

With respect to the definition of relevant categories of income sources, in a study of poverty and labour market participation, Van der Berg (1992) decomposes the sectors of employment for the South African labour force into three groups. The categorization is based on the extent to which to workers and dependants "participate in the modern consumer economy". The three groups are:

- **the core economy** sectors – manufacturing, government, other industry and services
- **the marginal modern economy** – commercial agriculture, domestic services, mining
- **the peripheral economy** – subsistence agriculture, informal sector, unemployed

According to Van der Berg (1992) "… *part of the labour force in the modern economy are to a larger degree no longer poor. Poverty in its most extreme form now mainly occurs in the peripheral sectors* […], *but is also widespread amongst workers and dependants relying on earnings from the primary and low-wage sectors*."

In this study, the classification of households' income sources are inspired by the above work, but categories within the marginal modern sectors have been created according subsector origin and public and private transfer incomes implicitly represent household income generation in the "peripheral" segment. Here, the "core" thus includes all sectors *except the Primary sectors), Domestic services* and *Mining and quarrying*. The Core sector category furthermore encompasses households with Capital income and all types of self-employment income as main income sources. In addition to these income sources it is recognized that households also derive "indirect income" and "*diversifying*" households are defined as those without a unique main income source that meets a contribution requirement discussed below. The income source categories are described in greater detail and in as close approximation as possible of the wording in the IES95 questionnaire in Appendix 1.[10]

[10] It has been noted by Leibbrandt et al (2000), that the IES95 data do not capture agricultural activities for own consumption well. In this study's sample from the IES95, 9.7 percent of all households were

Main income source definition

The definition of a main income source is not trivial.[11] One possible route is to construct the definition by the source's contribution to total household income. Some ambiguity necessarily enters the decision of where to draw the cut-off contribution-line. This study uses a minimum contribution (regardless of the number of members that raise the income) of 66.7 percent to total household income, an appeal of which is that the main income source contributes at least twice as much as any other income source.

Table 5 shows the impacts on the distribution of individuals across the various main income categories from where the cut-off contribution is drawn. In the second row of the table can be seen that roughly 75 percent of the households had a main income source and that approximately half the main income sources originate in the core sectors. The second largest category is Public Transfers with 16 percent of the households, followed by Private Transfers with six percent. Both the Mining and Quarrying and the Domestic services categories are small with two percent each, while the Primary sectors and Indirect income groups contain four and five percent respectively.

The figures in Table 6 attest to the notion that income generation activities as defined by these categories vary across the expenditure distribution. The table shows the composition of the sample quintiles with respect to households' main income sources. The figures show e.g. that the fractions of households which rely on Core or Mining-and-Quarrying sector wage incomes increase dramatically from the lower income brackets to the higher. The opposite is true for the two categories of households that

---

recorded with either slaughtered domestic animals or harvested crops in the year preceding the interview. Profit from agricultural activities should be registered in the IES questionnaire under "self-employment", but only 1.2 percent of the households that had slaughtered or harvested had records of any self-employment profits at all. The above figures presumably understate the importance of agriculture, which according to May (1996) assumes several important functions as inter alia a supplementary source of nutrition and as a safety net for vulnerable households in South Africa. But left with little choice other than taking the data at face value, agricultural production is not treated as a separate source of income. The individuals in the few households that would have agricultural income as their main source are included in the core economy category along with other types of self-employment.

[11] For the analysis of livelihoods in a dynamic setting Ardington and Lund (1996) raise a valid objection to the use of a "dominant source of income" in that such sources may be of a temporary nature.

rely on Public transfers or Private transfers, as well as for households with main income from the Primary sectors and Domestic services.

Mean expenditure levels by Main income source reflect the above distribution, with annual averages in the neighbourhood of R 2 000-2 500 for individuals in households with either transfers or wage incomes from the Primary or Domestic service as main income sources. Members of households which rely on Indirect income, Core or Mining and Quarrying sector incomes are associated with mean expenditures in the range of R 5 000-6 000, while those in Diversifying households constitute a middle category with average annual expenditures just below R 3 500.

Table 5) Percentage fraction of individuals in households by main income source category and various main-income cut-off contributions levels.

Main income contribution to total household income	Main income source category								
	No main income source	Core sectors	Mining and quarrying	Primary Sectors	Domestic Services	Public transfers	Private Transfers	Indirect income	Sum
50%	6	43	2	8	3	20	8	10	100
66.7%	**26**	**39**	**2**	**5**	**2**	**16**	**7**	**4**	**100**
75%	37	34	2	4	1	14	6	2	100
90%	55	25	1	2	1	11	5	0	100
100%	75	14	0	1	1	7	3	0	100

Table 6) Percentage-wise composition of sample quintiles by Main income sources.

Main income Source	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	Total	Mean expenditures
Diversifying	28.9	30.9	29.7	24.1	17.1	26.1	3435
Core sectors	13.7	23.4	36.0	**50.7**	**69.0**	38.6	5983
Mining & quarrying	1.0	1.2	1.9	**3.1**	**3.5**	2.2	5249
Primary sectors	**6.1**	**6.0**	6.4	4.3	0.8	4.7	2404
Domestic services	2.5	2.2	1.9	1.4	0.6	1.7	2506
Public transfers	**33.1**	**24.8**	13.2	6.9	1.6	15.9	1923
Private transfers	**11.4**	**8.6**	7.6	4.4	1.2	6.6	2232
Indirect income	3.3	3.0	3.3	5.1	6.3	4.2	5860
Total	100.0	100.0	100.0	100.0	100.0	100.0	4174

The two geographical dimensions of inequality

In the study two sets of groups are defined by Province and by Location. Table 7 shows that 60 percent of the households in the sample are rural but also that the Location-wise composition of the quintiles differ considerably. In the lowest

bracket the fraction of rural individuals is nearly 85 percent while in the highest bracket the corresponding fraction is 28.3 percent. The differences in composition shift much more gradually across the three first quintiles while the rural fraction decreases by twenty percentage points between both the third and fourth and the fourth and fifth quintiles. It is also noteworthy that the urban mean expenditure at R 6 124 is more than twice that of the rural at R 2 878.

Table 7) Percentage-wise composition of sample quintiles by Location.

Location	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	All	Mean expenditures
Rural	**84.3**	**75.6**	65.5	46.6	28.3	60.1	2878
Urban	15.7	24.4	34.5	**53.4**	**71.7**	39.9	6124
Total	100.0	100.0	100.0	100.0	100.0	100.0	4174

From the figures in Table 8 can be seen that two provinces, KwaZulu-Natal and the Eastern Cape, each contain fractions of almost 20 percent of the sample. In a second category of size are Limpopo and Gauteng with 14.3 percent and 12.3 percent respectively. The Northern Cape contains the smallest sample fraction with only 1.8 percent, while the remaining three provinces contain shares of in the range of 8-9 percent. The differences in the provincial composition of the quintiles are perhaps best illustrated using, on the one hand, the poorest provinces of the Eastern Cape and the Free State, and on the other the richest, Gauteng and the Western Cape. The fractions of the poorest two provinces in Table 8 are considerably smaller in the highest bracket, at 9.5 percent and four percent respectively, than in the lowest with 29.7 percent and 12.2 percent. Vice versa applies to the two richest provinces, with 29.6 percent and 14.4 percent in the fifth quintile and 2.2 percent and 1.1 percent in the first.

In terms of average expenditures, the Eastern Cape and the Free State are both found at the bottom with less than R 3 000 per month. Slightly higher mean expenditure levels are found in Mpumalanga and the Northern Cape at approximately R 3 300. The average in the Northern Province is R 200 higher than the latter two and KwaZulu-Natal is higher than the Northern Province by the same amount. With over R 7 500 Gauteng is at a considerably higher level than that of the second highest

province, the Western Cape, at just over R 5 800. The average expenditure in Limpopo is below the nationwide average by just over R 50.

Table 8) Percentage-wise composition of households' sample quintiles by Province.

Province	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	All	Mean expenditures
W Cape	**1.1**	4.6	**8.2**	14.2	**14.4**	8.5	**5834**
E Cape	**29.7**	24.9	**17.6**	12.3	**9.5**	18.8	**2989**
N Cape	1.7	2.3	2.2	1.9	1.1	1.8	3348
Free State	**12.2**	8.6	**6.5**	4.9	**4.0**	7.2	**2814**
KZN	15.5	21.1	24.8	21.8	15.4	19.7	3769
NW Province	11.9	11.4	9.3	7.1	7.1	9.4	3517
Gauteng	**2.2**	3.8	**8.6**	17.4	**29.6**	12.3	**7558**
Mpumalanga	7.8	9.4	9.6	8.2	5.2	8.0	3273
Limpopo	17.8	14.0	13.4	12.3	13.8	14.3	4122
Total	100.0	100.0	100.0	100.0	100.0	100.0	4174

Education of household heads

The association between the head's education level and the expenditure distribution is depicted in Table 9. As can be seen, almost 75.4 percent of the household heads in the sample have primary education or less and approximately two-fifths of those have no education at all. Almost 15 percent of all the heads have *some* secondary education, whereas only 11.5 percent have *completed* or above secondary education, out of which 5.2 percentage points have more than secondary education.

Table 9) Percentage-wise composition of sample quintiles by Education.

Education level of household head	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	All	Mean expenditures
None	**46.8**	37.6	30.8	18.8	**7.9**	28.4	2388
Primary	46.1	50.3	50.3	50.3	29.9	45.4	3304
Some secondary	5.5	9.5	13.6	20.4	24.4	14.7	5537
Complete secondary	**1.5**	1.9	3.3	6.6	18.4	6.3	8671
Above secondary	**0.2**	0.6	1.9	4.0	19.3	5.2	12177
Total	100.0	100.0	100.0	100.0	100.0	100.0	4174

The fraction of households that are headed by individuals with none or primary education each diminishes upwards in the income distribution from approximately 46 percent to around nine percent and 30 percent, respectively, in the fifth quantile. The opposite is true for the fractions of households with better educated household heads that increase from 1.5 percent and 0.2 percent, in order of educational achievement, to 18.4 percent and 19.3 percent, respectively, in the highest bracket. Also here, the

compositions within the three lower quintiles are reasonably similar and the change between the fourth and fifth quintiles is more dramatic than between the third and fourth.

There are also considerable differences in mean expenditure levels between the Education subgroups. The highest expenditures are found among households with heads that have more than secondary education, the average of which at R 12 177 is five times higher than for the category with none-educated heads. The mean expenditure of the second highest education category is found at R 8 671, which in turn is R 3 134 above the mean of the households with heads that have some secondary education. Households with heads that have only primary education have a mean expenditure level of R 3304.

Gender and age of household heads

Table 10 illustrates that the fraction of female headed households host just over 35 percent of the sample, but the fraction decreases gradually by a total of almost ten percentage points, from 43 percent in the poorest quintile to the 33.8 percent in the fourth quintile. In the richest quintile however, the corresponding fraction is only 24.3 percent. The average expenditure level of male headed households at R 4 629 is almost 40 percent higher than that of the female headed at R 3 352.

Table 10) Percentage-wise composition of sample quintiles by Gender.

Gender of household head	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	All	Mean expenditures
Female	**43.0**	**40.7**	36.4	33.8	24.3	35.6	3352
Male	57.0	59.3	63.6	66.2	**75.7**	64.4	4629
Total	100.0	100.0	100.0	100.0	100.0	100.0	4174

With respect to age categories, the figures in Table 11 show that households with heads aged 41-59 years old contain 45 percent of the sample. The youngest and oldest categories, below 25 and above 60, each host 2.7 percent and the remaining two subgroups thus take in approximately one-quarter each. The fractions of the youngest and second oldest categories do not differ dramatically across the quintiles either and remain at approximately 2.5 percent and 45 percent. The fraction of individuals that

live in households with elderly heads is reduced dramatically however, from almost 35 percent in the poorest quintile to just below 11 percent in the highest. The pattern is the diametrically opposite for the second youngest age category which increases from 17.4 percent in the first quintile to over 42 percent in the fifth. In both cases the shifts in composition are most dramatic from the fourth to the fifth quintile.

Table 11) Percentage-wise composition of sample quintiles by Age category.

Age category of household head	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	All	Mean Expenditures
≤ 25	2.6	2.7	2.8	2.9	2.2	2.7	4006
26 – 40	17.4	20.5	27.2	**31.0**	**42.2**	27.7	5513
41 – 59	45.0	43.6	43.3	46.1	44.8	44.6	4162
60 ≤	**34.9**	**33.2**	26.8	20.0	**10.8**	25.1	2741
Total	2.6	2.7	2.8	2.9	2.2	2.7	4174

With respect to expenditure levels, considerable differences exist between on the one hand the expenditures of households the heads of which are in the oldest age category and those in the second youngest category, in the age span of 26 to 40. The latter's average if found at R 5 513, while the formers' is at R 2 741. The expenditure level among individuals that live in households with heads in the second oldest age category is R 12 below the nationwide average and R 156 higher than that of the youngest age category.

Simultaneous application of several explanatory variables

Finally, reasons exist to believe that partitions both by Education and Main income sources to some extent may capture geographical variation; Firstly, during the apartheid era - when most heads in the sample were of schooling age – the access to and the quality of Education for Africans was subject to geographical variation (Wilson and Ramphele (1989). Secondly, it is conceivable that households' Main income sources are determined *inter alia* by the education of household members that raise the income as well as by the household's geographical location. Also, Education may affect the level of earnings from individual income sources. In order to get an impression of the extent to which the partitions by Education and Main income source capture geographical variation and/or overlap each other, the results from four

other partitions, based on combinations of the latter four explanatory characteristics will also be presented in the section with empirical results.

6 Empirical analyses

Table 12 contains the results from the decomposition of the Theil-L index along the partitions defined by the various explanatory characteristics. The table is designed in the following manner: The total level of inequality is found in the first column of the table. Each row contains the results from the partition along one specific characteristic. Within each row are found two sets of figures, the top ones are percentage fractions of total inequality and the lower ones contain absolute index values. The third column contains the fraction and index value of *explained inequality* pertaining to each partition, followed in the fourth column by the total *within-group* fraction and index value. The subsequent columns contain the contribution to total inequality from inequality within each group defined by the partition, as well as the index value for each group. (As an indication of the robustness of results Table A4 in Appendix 2 contains the corresponding values for the decomposition of the $G_1(x)$ (Theil-T) index and Table A5 and A6 contain the corresponding decompositions for a data set from year 2000.)

The index value for total inequality in the sample is 0.393 and the rows in the table are found in rank order of each partition's fraction of explained inequality. Two points are of contextual interest in this respect. Firstly, Leibbrandt, Woolard and Bhorat (2000) report a Theil-L statistic of 0.706 for the *total* sample of individuals in the same survey – including also the Asian/Indian and white subsamples - with the identical adult equivalence scale applied to household *income*. Hence, inequality in this study's sample is considerably lower than in the full sample. Secondly, while the Theil-L and Gini-index of inequality do not in general yield identical results, South Africa has the seventh highest level of income inequality in the world as measured by the former applied to the full population (Worldbank (2004)).[12]

As can be seen from the third column, the explained inequality from the gender of household head is the lowest at 2.9 percent, followed first by heads' population

[12] Leibbrandt, Woolard, and Woolard (2000) decompose the Gini-index for South Africa

Table 12) Inequality in the sample as measured and decomposed by the Theil-L index; percentage fractions of between- and within-group inequality partitions defined by one characteristic. Absolute index values in parentheses.

Total index value	Partition	Between-group inequality	Total within-group	Within-group contributions and absolute levels of inequality								
0.393	Gender	2.9 (0.012)	97.1 (0.382)	Female 30.2 (0.333)	Male 66.9 (0.409)							
	Race	3.0 (0.012)	97.0 (0.381)	African 88.0 (0.389)	Coloured 9.0 (0.323)							
	Age	7.9 (0.031)	92.1 (0.362)	≤25 2.5 (0.375)	26-40 27.9 (0.396)	41-59 44.7 (0.395)	60≤ 17.0 (0.265)					
	Province	13.1 (0.051)	86.9 (0.342)	W Cape 5.7 (0.265)	E Cape 17.8 (0.373)	N Cape 1.3 (0.281)	Free State 6.0 (0.327)	KZN 15.5 (0.308)	NW Prov 9.1 (0.383)	Gauteng 9.3 (0.299)	Mpumalanga 5.5 (0.269)	Limpopo 16.7 (0.461)
	Location	17.9 (0.070)	82.1 (0.323)	Rural 46.7 (0.306)	Urban 35.4 (0.349)							
	Main income source	23.5 (0.092)	76.5 (0.301)	Diversifying 22.5 (0.338)	Core sectors 32.4 (0.330)	Mining & Quarrying 1.5 (0.266)	Primary sectors 2.3 (0.192)	Domestic services 1.0 (0.234)	Public transfers 8.0 (0.197)	Private transfers 3.7 (0.219)	Indirect income 5.2 (0.486)	
	Education	30.8 (0.121)	69.2 (0.272)	None 17.6 (0.244)	Some primary 31.4 (0.272)	Some secondary 11.3 (0.302)	Complete secondary 5.1 (0.314)	Post secondary 3.9 (0.292)				

groups and then heads' age categories, with 3.0 percent and 7.9 percent respectively. The Province and Location partitions yield higher explained fractions at 13.1 percent and 17.9 percent, respectively, while differences in households' Main income source categories explain 23.5 percent of the inequality. The highest value is found for the education levels of household heads which account for 30.8 percent of the total inequality.

Equation (4') shows how the total within-group fraction of inequality for a given partition is a weighted-sum of the inequality within each subgroup and from equation (6) follows that the percentage fraction of within-group inequality is always 100 minus the between-group fraction. Hence, within-group inequality is higher than 90 percent in the three first partitions, approximately 85 percent in the fourth, near three-quarters in the Main income source partition and just below 70 percent when subgroups are defined by Education. With the exception of the Province partition do the largest subgroups also contribute the largest fraction to within-group inequality levels. However, only for the partitions by Gender, Race, and Age are the largest subgroups also associated with the highest *levels* of inequality.

Among the smaller subgroups which display inequality levels that are distinctly higher than the nationwide figure are found the province of Limpopo and the Main income category Indirect income. The former contains 16.7 percent of the sample and has an inequality index value of 0.46, whereas 5.2 percent of the sample reside in households that belong to the latter income category, which includes highly varying types of income sources. With respect to Limpopo province, it's noteworthy that the average expenditure in the province was very close to the nationwide mean (see Table 8).

Several subgroups also display considerably *lower* inequality than the nationwide level. Among the groups with the very lowest inequality levels are the four provinces of the Western Cape, Northern Cape, Gauteng and Mpumalanga; the five Main income source categories Mining and Quarrying, Primary sectors, Domestic services, and the two transfer categories; the two Education categories Some primary and Post secondary; and finally the above-60 age category. The index value for all these groups

are in the approximate range of 0.20-0.30. A plausible explanation for these low levels of inequality, in accordance with the previous section's descriptive statistics, is that the Western Cape, Gauteng, Mining and Quarrying, and the Post secondary education categories all predominantly contain observations at the *upper* end of the expenditure distribution, while the other subgroups contain observations clustered at the *lower* end.

Table 13) Explained fractions of inequality; partitions defined by multiple characteristics.

Partition defining Characteristics	Explained inequality	Number of groups	
		Theoretical	Observed
Province and Location	25.2	18	18
Province, Location, and Main income source	39.8	144	141
Province, Location and Education	44.4	90	90
Province, Location, Education and Main income source	53.0	720	574

The partition defined by Education category yielded the highest fraction of explained inequality above, followed by the partitions by Main income source, Location and Province. The results from four other partitions, that are defined along more than one dimension, are shown in Table 13. In the first row of results can be seen that, when applied simultaneously to define 18 subgroups, Province and Location jointly explain 25.2 percent of the inequality in the sample. When each of these subgroups were refined by hypothetically eight Main income source subgroups each, 141 observed subgroups were returned, that jointly explain 39.8 percent of the inequality. Hence, 14.6 percentage points of explained inequality were added. If the 18 geographical subgroups were rather refined by Education, 90 groups were defined and observed, which added 19.2 percentage points to yield an explained fraction of inequality of 44.4 percent. Finally, refining further by applying both Education and Main income source to the combined geographical partition returned 574 observed groups and a fraction of explained inequality of 53 percent.

Hence, implications from the latter set of results are that, when applied solely Education and Main income sources both capture some of the inequality explained by Location and Province. However, when partitions are defined by Location and

Province jointly *and* either Education or Main income sources, both of the latter individually capture inequality that is not explained by geographical variation. In reality it is furthermore plausible that household heads' Education is a determinant of both (i) the households' type of main income source and (ii) the returns from that main income source. The results in the fourth row of Table 13 are perhaps most readily interpreted as evidence of point (i) but the results are likely to feature also inequality due to point (ii). The applied methodology can not resolve this issue.

7 Conclusions

In commenting their results from the decomposition of income inequality in the US Cowell and Jones (1995) consider their explained fractions in the ranges of 20 to 30 percent "not much". This study has utilized the Theil-L measure to decompose expenditure inequality in a sample of black and coloured South African individuals, sampled by a household survey from 1995. The results from partitions defined by *one* characteristic are higher than "not much" in only case here, namely in that of the education level of household heads, which accounts for 30.8 percent of the inequality. However, partitions that take several factors into account return explained fractions up to 53 percent. Hence, at least some of the inequality in this sample can be explained.

Somewhat surprisingly, the explanatory power of race in these samples is relatively low, which is to some extent true for geographical location in both the rural-urban dimension and in provinces, when applied separately. However, the explanatory power of geography increases to 25.2 percent when province and rural-urban location are applied jointly, which suggests a more meaningful perspective on the spatial dimension of welfare in South Africa. When the same partition was further refined by households' types of main income source, explained fractions rose to 39.8 percent, while a refinement by household heads' education level, rather than main income sources, yielded explained fractions of 44 percent. Applied simultaneously to the joint geographical partition, education and main income sources yielded the abovementioned highest fraction of explained inequality.

Thus, among the explanatory variables applied in this study, the education levels of household heads stand out as the single most important associate of differing positions

in the expenditure distribution. However, further research into the determinants of households' types of main income sources is also warranted, partly by its relatively high explanatory power in this setting. A further interpretation of the results is that main income sources add explanatory power to what is attributable to Education and geography. Hence, other factors than the latter two may determine the allocation of main income sources to households.

Other researchers have shown that a substantial contribution to total inequality in South Africa arises from inequality within the African and coloured population. The results in this study show that a considerable fraction of that within-population-group inequality *can* be explained by further refinement of partitions into a not overwhelming number of subgroups (a partition by e.g. Race and Education applied to a national level would yield 20 subgroups.). Currently most analyses of South African inequality are undertaken without investigation beyond the too narrow focus on population groups. Such an approach neglects several dimensions of inequality. The dimensions of households' core-economy integration and of education legacies from the apartheid era are not only informative to our understanding of the phenomenon. Relatively small differences in lengths of education affect the distribution of welfare among those worst off and this may be addressed by policy as may amounts and eligibilities for transfers.

References

Ardington, E., and Lund, F. (1996). "Questioning rural livelihoods", Ch.2 in Lipton *et al* (1996).

Bhorat, H., Leibbrandt, M., Maziya, M., Van der Berg, S. & Woolard, I. (2001). *Fighting poverty – labour markets & inequality in South Africa*. Cape Town, South Africa: UCT Press.

Blackorby, C., Donaldson, D., and Auersperg, M. (1981) "A new procedure for the measurement of inequality within and among population subgroups", *Canadian Journal of Economics*, 14:4.

Bourguignon, F. (1979). "Decomposable income inequality measures", *Econometrica*, 47.

Budlender, D. (1997). "The debate about household headship." *Statistics South Africa Report*, Statistics South Africa, Pretoria.

Champernowne, D. G., and Cowel, F. A., (1998) *Economic inequality and income distribution*, Cambridge, Cambridge University Press, UK.

Chaudhuri, S., and Ravallion, M. (1994). "How well do static welfare indicators identify the chronically poor?", *Journal of Public Economics*, 53:3.

Cowell, F. A. (1980). "On the structure of additive inequality measures", *Review of economic studies*, 47.

Cowell, F. A., and Jenkins, S.P.(1995), "How much inequality can we explain? a methodology and application to the United States.", *Economic Journal*, 105: 429.

Deaton, Angus, and Salman Zaidi (2002). "Guidelines for constructing consumption aggregates for welfare analysis". *LSMS Working Paper No. 135*, World Bank, Washington.

Fields, G. (2003) "Accounting for income inequality and its change: A new method, with application to the distribution of earnings in the United States," *Research in Labor Economics.*

Gustafsson, B., and Li, S. (2002). "Income inequality within and across counties in rural China", *Journal of Development Economics*, 69:1.

Haddad, L., and Kanbur, R., (1990). "How serious is the neglect of intra-household inequality", *Economic Journal*, 100.

Hoogeveen and Özler (2004), *Not separate, not equal, poverty and inequality in post-apartheid South Africa*, unpublished World Bank mimeo.

Kingdon, G., and Knight, J. (1999). "Unemployment and wages in South Africa: A spatial approach." *Working Paper WPS/99.12, Centre for the Study of African Economies*, Oxford University.

Klasen, S., (1997). "Poverty, inequality and deprivation in South Africa: an analysis of the 1993 SALDRU survey", *Social Indicators Research*, 41.

Klasen, S. (2000). "Measuring poverty and deprivation in South Africa", *Review of Income and Wealth*, 46:1.

Lanjouw, P., Prennushi,G. and Zaidi, S. (1996). "Building blocks for a consumption-based analysis of poverty in Nepal". The World Bank. Unpublished manuscript.

Leibbrandt, M. and Woolard, I. (1999) "A comparison of poverty in South Africa's nine provinces", *Development Southern Africa,* 16:1.

Leibbrandt, M., Woolard, I., and Bhorat. H. (2000)."Understanding contemporary household inequality in South Africa", *Studies in Economics and Econometrics*, 24:3

Leibbrandt, M, Woolard, C, & Woolard, I, 2000. "The contribution of income components to income inequality: a decomposable Gini analysis." *Journal of African Economies* 9

Leibbrandt, M & Woolard, I, 2001. Household incomes, poverty & inequality in a multivariate framework in *Bhorat et al* , 2001.

Lipton, M., De Klerk, M., and Lipton, M. (Eds), (1996) *Land, labour & livelihoods in rural South Africa*. Durban, South Africa: Indicator Press.

May, J., Carter, M., and Posel, D. (1995). "The composition and persistence of poverty in rural South Africa: an entitlements approach". *Land and Agriculture Policy Centre Policy Paper 15*. Land and Agriculture Policy Centre, Johannesburg, South Africa.

Moll, P., (1996). "The collapse of primary schooling returns in South Africa 1960-90", *Oxford Bulletin of Economics and Statistics*, 58:1.

Mwabu, G., and Schultz, P. (2000). "Wage premia for education and location, by gender and race in South Africa", *Economic Development and Cultural Change*, 48:2.

Ravallion, M.(1996). "Issues in measuring and modelling poverty", *Economic Journal* 106.

Sen, A. (1985). *Commodities and capabilities,* Amsterdam, North Holland.

Sen, A. (1987). The standard of living, Cambridge, Cambridge University Press.

Shorrocks, A. F. (1980). "The class of additively decomposable inequality measures", *Econometrica*, 48:3.

Shorrocks, A. F. (1984). "Inequality decomposition by population subgroups", *Econometrica*, 52:6.

Slesnick, D. (1993). "Gaining ground: poverty in the postwar United States.", *Journal of Political Economy*, 101.

Slesnick, D. (1998). "Empirical approaches to the measurement of welfare", *Journal of Economic Literature*, 36.

Statistics South Africa, 1997. Documentation supplied with OHS & IES1995 raw data sets. Statistics South Africa, Pretoria.

Tsaklogiu, P. (1993), "Aspects of inequality in Greece", *Journal of Development Economics*, 40.

van der Berg, S, 1992. "Redressing economic inequality in South Africa" in *Redistribution and affirmative action – working on the South African political economy*, Hugo, P. (Ed.). Halfway House, South Africa; Southern Book Publishers.

van der Berg, S, 2000. "Social policies to address poverty" in Bhorat *et al* (2001).

Whiteford, A. and McGrath, M., (1998). "Income inequality over the apartheid years." *SANER Working Paper*, No. 6, South African Network for Economic Research, Cape Town, South Africa.

Whiteford, A. and Mvan Seventer, D. E.(2000) "South Africa's changing income distribution in the 1990s", *Studies in Economics and Econometrics*, No 24:3.

Willis, R. J. (1986). Wage determinants: a survey & reinterpretation of human capital earnings functions. In Ashenfelter, O. & Layard, R. (Eds), *Handbook of labor economics*, Volume I, Elsevier Science.

Wilson, F. and Ramphele, M. (1989), *Uprooting poverty: the South African challenge.* Cape Town, South Africa: David Philip Publishers

Worldbank (2004) *World Development Indicators 2004*, World Bank, Washington DC.

Appendix 1

Income originating from the core economic sectors (henceforth "Core sector income")**:** *salaries and wages*[13] *from secondary sectors and tertiary sectors* including s*elf-employment income*, in the form of net profit from business or professional practice/activities conducted on a full time basis; and *capital income* from the letting of fixed property, royalties, interests, dividends and annuities.[14]

Primary sector income: *salaries and wages from agriculture, fishing, and forestry*.

Mining and quarrying sector income: *salaries and wages from mining and quarrying*.

Domestic services income: *salaries and wages from private households*.

Private transfers: alimony, maintenance and similar *allowances from divorced spouses or family members living elsewhere* and regular allowances from family members living elsewhere.

Pensions and public transfers: *pensions resulting from own employment, old age and war pensions, social pensions* or allowances in terms of *disability grants, family and other allowances, or from funds* such as e.g. the Workmen's Compensation, Unemployment Insurance, Pneumoconioses and Silicosis funds.

[13] In the "salaries and wages" concept is included bonuses and income from over time, commissions and directors fees, part-time work and cash allowances in respect of transport, housing and clothing.

[14] The secondary sectors encompass the Statistics South Africa (1997b) "Major sector divisions": Manufacturing; Electricity, gas and water; and Construction. The tertiary sectors constitute the "Private services" and "Community, social and personal services" excluding "Private households with employed persons". "Private services" is made up of the major divisions: Wholesale and retail trade, repair of motor vehicles, motor cycles and personal and household goods, hotels and restaurants; Transport, storage and communication; and Financial intermediation, insurance, real estate and business services.

Indirect income: income derived from [i] hobbies, side-lines, part-time activities, or the sales of vehicles, property etc; [ii] payments received from boarders and other members of the household; [iii] the pecuniary value of goods and services received by virtue of occupation; [iv] gratuities and lump sum payments from pension, provident and other insurance or from private persons; [v] 'other income' withdrawals, bursaries, benefits, donations and gifts, bridal payment or dowries and all 'other income'.

Appendix 2

This appendix provides indications of the robustness of the results in Table 12 and contains the results from the decomposition of inequality in the sample by the Thei-T measure. In addition to the latter, decomposition was undertaken with both measures to a data set from year 2000, generated by a nationwide questionnaire-based survey similar to the OHS/IES95, but with Statistics South Africa's biannual Labour Force Survey having taken the place of the then ceased annual October Household Survey. The comparability of the two data sets has been disputed in the South African research community. One reason for this uncertainty is that nominal incomes were lower in the four lowest per adult-equivalent income quintiles in year 2000 than 1995, whereas total expenditure data which do not display such characteristics. Tables A1 and A2 illustrate the latter issues.

The problem of matching individuals was much smaller with the LFS/IES2000 than with the OHS/IES95 data. In the former 96.2 percent of the observations that met the population group criterion are used. However, in addition to zero expenditures for 106 observations, 3438 observations were lost from the same data set due to either missing remuneration or industry data for wage earners or due to missing education or age data for household heads. The sample delimitation process is illustrated in Table A3.

Table A4 contains the decomposition results from the Theil-T index applied to the 1995 data. The decomposition results for the year 2000 data are shown in tables A5-6.

(Results from the application of other inequality measures on both data sets yield similar results and are available from the author.) In the 2000 data the total levels of inequality are considerably higher. However, in all three decompositions the relative rank and approximate differences in explained inequality for the various partitions are similar to those in Table 12. In the Theil-T decomposition of the 1995 data the explanatory power of race and Main income source is however lower by approximately one-tenth and one-fifth respectively.

In both decompositions of the 2000 data the explanatory power of race is roughly 50 percent higher than in the 1995 Theil-L decomposition. Applying the same index to the 2000 data shows that the explanatory power of Gender is almost twice as high, while that of Location is around ten percent higher. The fraction of explained inequality yielded by Education is however lower by one-tenth. For the Theil-T decomposition the difference in inequality explained by Gender is approximately 60 percent higher than in the Theil-L decomposition for 1995. Both the explanatory power or Age and Main income source are also lower, in the former case by some 15 percent and in the latter by around one-fifth.

Table A1) Per adult-equivalent household income quintiles in the full OHS/IES95 and LFS/-IES2000 samples; weighted population sizes and annual incomes in nominal Rand.

Sample	Quintile	Weighted number of individuals	Mean Income	Minimum income	Maximum income
1995	1	8 014 923	1 434.1	148.8	2 098.3
	2	8 015 225	2 813.6	2 098.3	3 638.7
	3	8 015 152	4 874.6	3 639.1	6 511.6
	4	8 014 762	10 010.8	6 511.6	15 538.1
	5	8 015 588	41 255.5	15 538.1	2 657 998.0
	All	40 075 650	12 078.1	148.8	2 657 998.0
2000	1	8 358 799	1315.8	11.0	2 035.4
	2	8 359 303	2745.0	2 035.4	3 571.6
	3	8 359 271	4827.8	3 571.6	6 430.6
	4	8 358 789	9805.4	6 430.6	15 564.9
	5	8 359 527	46786.6	15 564.9	2203 030.0
	All	41 795 689	13096.5	11.0	2203 030.0
Source: OHS/IES95 and LFS/IES2000, own computations, weighted figures. n1995 =125 112 n2000=101 803					

Table A2) Per adult-equivalent household expenditure quintiles in the full OHS/IES95 and LFS/IES2000 samples; weighted population sizes and annual expenditures in nominal Rand.

Sample	Quintile	Weighted number of individuals	Mean Income	Minimum income	Maximum Income
1995	1	8 015 065	1 354.2	162.2	1 982.4
	2	8 015 152	2 685.9	1 982.4	3 504.0
	3	8 015 078	4 719.1	3 504.0	6 314.9
	4	8 014 905	9 725.5	6 314.9	15 104.2
	5	8 015 450	40 427.9	15 104.2	2 657 998.0
	All	40 075 650	11 782.8	162.2	2 657 998.0
2000	1	8 359 097	1 611.5	0.0	2 395.7
	2	8 359 166	3 149.0	2 395.7	3 977.1
	3	8 358 449	5 190.5	3 978.0	6 699.5
	4	8 359 738	9 894.5	6 699.5	15 215.0
	5	8 359 239	46 109.7	15 215.0	2 740 995.0
	All	41 795 689	13 191.2	0.0	2 740 995.0
Source: OHS/IES95 and LFS/IES2000, own computations, weighted figures. n1995 =125 112 n2000=101 803					

Table A3) The year 2000 sample delimitation process

Year	Sample	Number of individuals	Share of total revisited sample	Share of revisited African and coloured sample
2000	Total LFS/IES sample	101 803	100.0	
	African and coloured LFS/IES sample	93 842	92.2	100.0
	Final sample	90 298	88.7	96.2
Source LFS/IES2000, own computations unweighted figures.				

Table A4) Inequality in the 1995 sample as measured and decomposed by the Theil-T index; partitions defined by one characteristic

Total index value	Partition	Between-group inequality	Total within-group	Within-group contributions and absolute levels of inequality								
0.430	**Gender**	2.6 (0.01111)	97.4 (0,41840)	**Female** 25.0 (0.37478)	**Male** 72.4 (0.43589)							
	Race	3.1 (0.01195)	96.9 (0,41612)	**African** 84.2 (0.43232)	**Coloured** 12.7 (0.33308)							
	Age	6.9 (0.02981)	93.1 (0.39970)	**≤25** 2.6 (0.44073)	**26-40** 34.3 (0.40303)	**41-59** 45.0 (0.43462)	**60≤** 11.2 (0.29207)					
	Province	13.0 (0.05586)	87.0 (0.37365)	**W Cape** 8.3 (0.30001)	**E Cape** 14.2 (0.45358)	**N Cape** 1.1 (0.32635)	**Free State** 4.0 (0.35644)	**KZN** 14.4 (0.34680)	**NW Prov** 8.1 (0.44060)	**Gauteng** 15.8 (0.30430)	**Mpumalanga** 4.2 (0.28473)	**Limpopo** 16.9 (0.51618)
	Location	16.4 (0.07056)	83.6 (0.35895)	**Rural** 34.5 (0.35768)	**Urban** 49.1 (0.35984)							
	Main income source	19.6 (0.08419)	80.4 (0.34532)	**Diversifying** 20.0 (0.40029)	**Core sectors** 43.6 (0.33872)	**Mining & Quarrying** 1.6 (0.24794)	**Primary sectors** 1.2 (0.19605)	**Domestic services** 0.5 (0.22480)	**Public transfers** 4.0 (0.23640)	**Private transfers** 2.0 (0.23741)	**Indirect income** 7.4 (0.54114)	
	Education	32.4 (0.13930)	67.6 (0.29021)	**None** 10.1 (0.10114)	**Some primary** 24.3 (0.24332)	**Some secondary** 14.1 (0.31071)	**Complete secondary** 9.1 (0.29793)	**Post secondary** 9.9 (0.27997)				
Source OHS/IES95, own computations, weighted figures. n = 92 717												

Table A5) Inequality in the 2000 sample as measured and decomposed by the Theil-L index; partitions defined by one characteristic data.

Total index value	Partition	Between-group inequality	Total within-group	Within-group contributions and absolute levels of inequality							
0.485	**Gender**	5.7 (0.028)	94.3 (0.454)	**Female** 37.3. (0402)	**Male** 57.0 (0.496)						
	Race	4.8 (0.023)	95.2 (0.458)	**African** 87.7 (0.468)	**Coloured** 7.4 (0.370)						
	Age	7.6 (0.036)	92.4 (0.445)	**≤25** 3.3 (0.401)	**26-40** 27.7 (0.478)	**41-59** 44.3 (0.503)	**60≤** 17.2 (0.322)				
	Province	13.4 (0.065)	86.6 (0.417)	**W Cape** 6.0 (0.340)	**E Cape** 15.1 (0.460)	**N Cape** 1.3 (0.420)	**Free State** 5.7 (0.493)	**KZN** 17.5 (0.415)	**NW Prov** 7.4 (0.495)	**Gauteng** 18.2 (0.389)	**Mpumalanga** 4.9 (0.363)	**Limpopo** 10.5 (0.418)
	Location	20.1 (0.097)	79.9 (0.385)	**Rural** 31.9 (0.327)	**Urban** 48.0 (0.436)						
	Main income source	23.7 (0.114)	76.3 (0.368)	**Diversifying** 19.9 (0.406)	**Core sectors** 33.1 (0.414)	**Mining & Quarrying** 1.3 (0.322)	**Primary sectors** 1.6 (0.232)	**Domestic services** 2.2 (0.289)	**Public transfers** 6.8 (0.209)	**Private transfers** 5.5 (0.292)	**Indirect income** 6.0 (0.674)	
	Education	27.2 (0.131)	72.8 (0.351)	**None** 19.8 (0.385)	**Some primary** 32.9 (0.340)	**Some secondary** 11.2 (0.337)	**Complete secondary** 5.6 (0.348)	**Post secondary** 3.3 (0.328)				
Source LFS/IES2000, own computations, weighted figures. n = 90 298												

Table A6) Inequality in the 2000 sample as measured and decomposed by the Theil-T index; partitions defined by one characteristic data.

Total index value	Partition	Between-group inequality	Total within-group	Within-group contributions and absolute levels of inequality								
0.533	Gender	5.0 (0.027)	95.0 (0.506)	Female 28.5. (0.454)	Male 66.5 (0.496)							
	Race	5.1 (0.027)	94.9 (0.505)	African 83.0 (0.534)	Coloured 11.9 (0.367)							
	Age	6.5 (0.035)	93.5 (0.498)	≤25 3.2 (0.416)	26-40 35.4 (0.499)	41-59 42.6 (0.545)	60≤ 12.3 (0.397)					
	Province	12.5 (0.066)	87.5 (0.466)	W Cape 9.9 (0.360)	E Cape 10.8 (0.543)	N Cape 1.3 (0.488)	Free State 4.6 (0.537)	KZN 13.4 (0.484)	NW Prov 8.5 (0.670)	Gauteng 26.8 (0.432)	Mpumalanga 4.6 (0.404)	Limpopo 7.7 (0.519)
	Location	16.7 (0.089)	83.3 (0.444)	Rural 17.9 (0.179)	Urban 65.4 (0.654)							
	Main income source	19.6 (0.104)	80.4 (0.429)	Diversifying 19.7 (0.526)	Core sectors 45.4 (0.417)	Mining & Quarrying 3.1 (0.502)	Primary sectors 0.8 (0.226)	Domestic services 1.1 (0.278)	Public transfers 3.1 (0.241)	Private transfers 2.6 (0.321)	Indirect income 4.6 (0.692)	
	Education	29.6 (0.158)	70.4 (0.375)	None 13.0 (0.487)	Some primary 25.3 (0.386)	Some secondary 12.1 (0.338)	Complete secondary 10.8 (0.371)	Post secondary 9.1 (0.301)				
Source LFS/IES2000, own computations, weighted figures. N = 90 298												

Chapter II:

Homing in on the core – household incomes, income sources, and geography in South Africa

Sten Dieden*

Abstract

The focus of this study is on household income generation among previously disadvantaged households in South Africa. Previous research has found that poverty among South African households was associated with the extent to which workers and their dependants were integrated into the South African core economy. This study investigates whether a similar conception can be ascertained in multivariate regression analysis. Households' income sources are divided into categories that reflect differing extents of association with the core economy. Ensuing further justification by results from descriptive analyses, the income source categories are utilized as explanatory variables to investigate whether inter-household variation in income sources can explain variation in income levels. For the latter purposes, the results from the estimation of three reduced form models are compared. All three models have households' log-income levels as dependent variables and share a set of household characteristics as explanatory variables. Two of the models are two-stage specifications that use provincial locations in the construction of instruments for income source categories. The third specification contains no income source variables but includes provincial locations as explanatory variables. The results show that, as compared to the specification with provincial locations, income sources can be incorporated as explanatory variables into multivariate regression analyses without considerable loss of explanatory power. Controls for endogeneity must however be applied. The partial impacts from income sources are statistically significant and their signs are in accordance with expectations. For some income sources the magnitudes of the impacts are not in correspondence with what may be expected from the descriptive analysis. The latter results suggest that households in different main income source categories also differ systematically in their demographic and educational endowments. When assimilated with results from the descriptive analyses, the estimated partial impacts from the different provinces support this interpretation.

* Department of Economics, University of Gothenburg, PO Box 640, SE 405 30 GOTHENBURG. Sweden. Ph. +46 31 773 41 98 Fax 0946 31 773 2648. sten.dieden@handels.gu.se.

While any defects or shortcomings in this work are entirely my own responsibility, I'm deeply indebted to Arne Bigsten, Stephan Klasen, Paul Lundall, Laura Poswell, Ali Tasiran, and participants at University of Cape Town's School of Economics seminar series for very valuable comments to previous versions of this work. The financial provision by the Swedish International Development Cooperation Agency (Sida) and by the University of Cape Town's Centre for Social Science Research (CSSR), that also hosted me during much of the time spent on research for this work, is thankfully acknowledged.

1. Introduction

As a legacy of racially discriminatory dispossession of land rights and forced removals, little agricultural self-employment is found among South Africa's rural non-white households, while dependence on transfer incomes is prevalent and unemployment rates are high (SALDRU (1994), Jensen (2002)). Hence, the conditions for household income generation appear atypical to the rest of the continent and many South African households seem to face severe constraints in their livelihood generation (Reardon (1997), Kingdon and Knight (2004)). Previous research on South Africa emphasises the role of households' access to *wage income* in avoiding poverty and in accounting for income inequality (Bhorat, Leibbrandt, Maziya, Van der Berg, and Woolard (2001)). A further refined perspective was adopted by Van der Berg (1992), who pronounced that poverty among South African households was associated with the extent to which workers and their dependants were integrated into the South African core economy. This study investigates whether a conception similar to the latter can be ascertained in multivariate regression analysis of the income levels among previously disadvantaged households in South Africa. The households' income sources are divided into categories which reflect differing extents of association with the core economy. The same categories are subsequently utilized to investigate whether inter-household variation in income sources can explain variation in income levels.

South Africa is a vast country where the physical geographical conditions for income generation vary distinctly from one region to another. This variation is further augmented by legacies from colonial and apartheid policies that fostered uneven spatial economic development (Wilson and Ramphele (1989)).[1] When income sources are applied to explain variation in income levels good reasons exist to suspect that causality may be running both ways between the dependent and explanatory variables. In order to investigate for such statistical endogeneity, the empirical analysis in this study utilizes the perception that geographical location may affect household income levels via variations in the accessibility of different income sources across locations.

[1] Direct impacts from both urban/rural and provincial location on household welfare in South Africa are well documented (e.g. Leibbrandt and Woolard (1999), Klasen (1997, 2000)).

This study's analysis of South African household survey data from 1995 augments previous research in several ways. Firstly, descriptive analyses show that the vast majority of the households under scrutiny derive more than two-thirds of their income from *one* category of income sources. Secondly, the results from studies that recognize the importance of access to wage income in this context are processed by the estimation of separate impacts for wage-income of different origins as well as for two transfer income categories and for "indirect income". In addition, the study's categorization of South African households by their income sources provides a composite appreciation of key facets of deficient household incomes in the country.

The empirical analysis involves a comparison of the results form three reduced form WLS regression specifications. All specifications have a set of household characteristics as explanatory variables in common. Two of the specifications are novel to the South African literature in that they contain households' income sources as explanatory variables. In these specifications, dummy variables for provincial location are utilized as first-stage, *instrument variables*, in order to test and control for the simultaneous determination of income sources and income levels. In order to get an impression of the extent to which utilization of province dummies as instruments come at a cost of lost explanatory power in the second-stage regression, the third specification utilizes the province dummies juxtaposed to the other explanatory variables in a one-stage regression model.

The paper proceeds as follows: Section 2 introduces South African income source categories and relates these to households' core integration. Section 3 is a brief review of South African research on poverty and income sources in the broader African context. The data, sample delimitations and the main income source definition are discussed in Section 4. A discussion founded on descriptive statistics links the main income source concept to some aspects of households' income generation in Section 5. Section 6 discusses the reduced form approach to modelling household incomes. The explanatory variables applied in this study are introduced and some analytical concerns are raised. Section 7 motivates this study's utilization provincial locations as instruments for main income sources. The empirical approach is introduced in Section 8 and this is followed by the empirical investigation in Section 9. Finally, conclusions are drawn in Section 10.

2. South African households' income sources

The South African literature usually distinguishes between at least four broad groups of household income sources, which may be classified as private transfers, public transfers, self-employment, and wage income (Carter and May (1999)). In a study of poverty and labour market participation Van der Berg (1992) decomposes the sectors of employment for the South African labour force into three groups. The categorization is based on the extent to which to workers and dependants "participate in the modern consumer economy". The three groups are:

- **the core economy** sectors – manufacturing, government, other industry and services
- **the marginal modern economy** – commercial agriculture, domestic services, mining
- **the peripheral economy** – subsistence agriculture, informal sector, unemployed

According to Van der Berg (1992) "… *part of the labour force in the modern economy are to a larger degree no longer poor. Poverty in its most extreme form now mainly occurs in the peripheral sectors* […], *but is also widespread amongst workers and dependants relying on earnings from the primary and low-wage sectors*." The analyses in this study and the classification of households' income sources in particular are inspired by the above work. However, here income from the marginal modern sectors is decomposed into its subsectors, while public and private transfers separately represent income generation in the "peripheral" segment.

The "core" concept in this study thus includes all sectors *except the Primary sectors, Domestic services* and *Mining and Quarrying*. Income from capital and self-employment are also attributed to the core. In addition to these income sources is also recognised "indirect income". Below follows a more detailed list of the income sources in each category with descriptions phrased in as close approximation as possible of the wording in the IES95 questionnaire:

Income originating from the core economic sectors (henceforth "Core sector income")**:** *salaries and wages*[2] *from secondary sectors and tertiary sectors* including s*elf-employment income*, in the form of net profit from business or professional practice/activities conducted on a full time basis; and *capital income* from the letting of fixed property, royalties, interests, dividends and annuities.[3]

Primary sector income: *salaries and wages from agriculture, fishing, and forestry*.

Mining and Quarrying sector income: *salaries and wages from mining and Quarrying*.

Domestic services income: *salaries and wages from private households*.

Private transfers: alimony, maintenance and similar *allowances from divorced spouses or family members living elsewhere* and regular allowances from family members living elsewhere.

Public transfers: *pensions resulting from own employment, old age and war pensions, social pensions* or allowances in terms of *disability grants, family and other allowances, or from funds* such as e.g. the Workmen's Compensation, Unemployment Insurance, Pneumoconioses and Silicosis funds.

Indirect income: income derived from [i] hobbies, side-lines, part-time activities, or the sales of vehicles, property etc; [ii] payments received from boarders and other members of the household; [iii] the pecuniary value of goods and services received by virtue of occupation; [iv] gratuities and lump sum payments from

[2] Included in the grouping "salaries and wages" are bonuses and fixed or contributed income commissions and directors fees, part-time work and cash allowances in respect of transport, housing and clothing.

[3] According to Statistics South Africa (1997b) the secondary sectors include: Manufacturing, Electricity, gas and water and Construction. The tertiary sectors constitute the "Private services" and "Community, social and personal services" excluding "Private households with employed persons". "Private services" is made up of the following divisions: Wholesale and retail trade, repair of motor vehicles, motor cycles and personal and household goods, hotels and restaurants; Transport, storage and communication; and Financial intermediation, insurance, real estate and business services.

pension, provident and other insurance or from private persons; [v] 'other income' withdrawals, bursaries, benefits, donations and gifts, bridal payment or dowries and all 'other income'.

Finally, in the aggregate, all income sources other than "Indirect income" will be referred to as "direct" income sources.

3. Previous research on income sources and income levels in South Africa

The increased collection of microdata since the early 1990 has led to a considerable amount of quantitative research being conducted on income poverty and inequality in South Africa, some of which is contained in Møller (1997), May (2000) and Bhorat *et al* (2001). Detailed work on the income sources and livelihoods among South African households is found also in Lipton, de Klerk and Lipton (1996). On a broader scale, an overview of rural livelihoods and diversity in the third world is provided by Ellis (2000).

Many household attributes that are associated with low household incomes in South Africa apply also in many other parts of sub-Saharan Africa. Among such attributes are found low levels of education, low or high age, and female gender of the household head. Also large household sizes and/or many dependants as well as location in rural areas are associated with low incomes and income levels are also subject to inter-regional variations (e.g. Coulombe and Mckay (1993), Leibbrandt and Woolard (1999), Geda, de Jong, Mwabu and Kimenyi (2001), Bigsten, Kebede and Shimeles (2003)). As could be expected, given South Africa's historical legacies, most of the above South African poverty analyses also attest to race as a dominant determinant of poverty (Carter and May (1999)).

Several recent studies that apply multivariate analysis to South African data emphasise the importance of households' access to *wage income* in explaining income inequality and in evading poverty (Carter and May (1999), Bhorat *et al* (2000)). Furthermore, according to Leibbrandt, Woolard, and Bhorat (2000), income generation processes differ above and below their poverty line, in that contributions of wages to total income are lower below their poverty line, whereas contributions from

remittances and state transfers are higher. One conclusion made by the authors is that wage income is central in the determination of both poverty status and poverty depth. On the same note Bhorat (2000) shows that households have relatively high poverty propensities where earners are exclusively either domestic workers or agricultural workers. A point highlighted by van der Berg (2000) which has even higher relevance to this study is that shares of remittance income decline with higher income-consumption quantiles while wage-income shares increase, both in general and as households' main sources of income. Evidence from this study to confirm these trends will be discussed in Section 5.

4. Data, main income source definition and sample delimitations

In October 1995 Statistics South Africa conducted questionnaire-based interviews on a wide range of living standards issues with a sample of 30.000 households, intended to represent all households in the country and containing nearly 131.000 inhabitants. Two months later 28 585 of the households were revisited in a more detailed investigation of their income and expenditure. These two surveys are often referred to as the October Household Survey/Income and Expenditure Survey 1995 (henceforth "OHS/IES 95").

The sample for the two surveys was stratified by province, by urban and non-urban areas, and by population group. Altogether, 3 000 enumerator areas were drawn as Primary sampling units in each of which ten households were visited. The data on households are supplied with weights in accordance with the number of households in each stratum. Statistics South Africa recommend that, when the two surveys are linked to each other the weights for the Income and Expenditure Survey should be applied to both (Statistics South Africa 1996, 1997a, 1997b). The above procedure is applied to the present analyses, but with the weights renormalized to sum to unity (Deaton (1997)).

In the two surveys a household is defined by "a person or a group of people dependent on a common pool of income who normally occupy a dwelling unit or a portion thereof and who provide themselves with food or the necessary supplies or arrange for such provision." A household "member" by definition resides at least four nights a

week in the household. The income concept applied in this study refers to annual income and controls for household size (number of members) as measured by per-adult-equivalents [4]. Table 1 shows the distribution of *all* households sampled by the IES95 across per-adult-equivalent income deciles by population group.[5]

Table 1: Households distribution across population groups, by per-adult equivalent annual income deciles (full OHS/IES95 sample).

Per adult-equivalent income decile	Population group				Total
	African	Coloured	Asian	White	
1	*96.4*	*3.6*	0.0	0.0	100.0
2	*94.2*	*5.5*	0.2	0.2	100.0
3	*90.0*	*9.1*	0.4	0.5	100.0
4	*86.6*	*11.6*	1.1	0.7	100.0
5	*81.1*	*13.9*	2.2	2.8	100.0
6	76.3	13.8	3.6	6.3	100.0
7	67.0	12.5	6.0	14.5	100.0
8	49.7	9.2	6.7	34.4	100.0
9	24.8	5.0	4.8	65.4	100.0
10	29.2	3.8	2.6	64.5	100.0
All	**69.5**	**8.8**	**2.8**	**18.9**	**100.0**
Weighted figures, n= 28 585					

This study uses a sub-sample consisting of 19.914 of the revisited households, the selection of which were based on two criteria. Firstly, since 95 percent or more of the households in the five lowest deciles in Table 1 belong either to the African or the coloured population groups, this study focuses on households where the head belongs to one of these racial groups. The second criterion is related to the identification of individuals in both surveys. Since the quality of the information on individuals' labour market characteristics is greater in the OHS module than in the IES, it was deemed desirable to extract labour market information from the former. Households in the two data sets are easily matched, since they were equipped with matching identifiers in

[4] This study uses the adult equivalence scale applied by May, Carter and Posel (1995) i.e.: $E=(A+0.5K)^{0.9}$, where E is number of adult equivalents, A the number of adults and K is the number of children 15 years old or younger. Leibrrandt and Woolard (2001) explore the impacts on incidence of poverty by several adult equivalence scales and find that South Africa's poverty rates among African and Coloured and rural and urban dwellers remains astonishingly unchanged, even when large adjustments are made to the scale parameters.

[5] Apartheid policies defined four main "racial classifications"; African, coloured, Asian/ Indian and white. The discrimination by race ran through all aspects of life and had tremendous effects on everyone's living standards. For these reasons official statistics in South Africa still apply "racial" categories, and here the same approach will be followed (referring to the same categories as "groups").

both data sets, whereas individuals were not. Individuals that were captured with any amount income in the IES module were therefore matched to the OHS data by means of households' unique identifiers, age, gender and race.

The final sample in the analyses, including only the households where all income earners were identified in both data sets, consists of 89 percent of the households that met the first criterion. Since the matching procedure would be more complicated the higher number of earners a household contains, the selection into this sample could be biased towards households with few earners. More detail on the matching procedure is found in Appendix 1.

A main income source can be defined by the fraction of total income that originates from that source-category. Table 2 contains only the households that met the first two criteria and shows how the distribution of these households across various main income source categories is affected by alternative definitions according to cut-off contributions. Hence, if a main income source is defined by a contribution of 50 percent or more to total household income, 5 percent of the households would not have a main income source. If the cut-off contribution is set at 90 percent, the fraction of households without a main income source increases to 52 percent, the mirror reflection of which is that 48 percent of the sample raise 90 percent or more of their income from one income source category. [6] Analogously for the 100 percent definition, more than one-quarter of the households derive *all* their income from one category. Further, a phenomenon robust to definitions is that households with core sector main income encompass roughly half the households with a main income source, followed by a fairly stable fraction of one-quarter to one-fifth of the households relying on public transfers.

Thus, regardless of which contribution defines a main income source *many households seem to rely to a high extent on a single source of income*. Yet, some ambiguity necessarily comes into the decision of where to draw the cut-off

[6] The magnitude of the fraction of Diversifying households that do not rely on a main income source is of some interest. A multitude of motives for and consequences of livelihood diversification exist (e.g. Ellis (2000)). While this investigation includes diversifying households as a main income source category, the analyses will remain incomplete in that no explanation is sought for why some households are more diversified than others.

contribution. Here the cut-off contribution is set at 66.7 percent of total household income. An appeal of this definition is that the main income source contributes twice as much to total household income as any other source and is unquestionably of considerable importance to the household.[7] In some respects the main income source may be considered a crude indicator of how a household's income is generated, in that the definition disregards e.g. the number of members involved and the contributors' individual characteristics. Appendix 2 provides some further indication as to the gravity of those objections.

The figures in the second column of Table 2 show that by the applied 66.7 percent criterion, 24 percent of the households fall in the category "No main income source" (henceforth "Diversifying" households), which implies that 76 percent of the households in the final sample *do* have a main income source. Out of the latter fraction, exactly half derive that income from the Core sectors. One fifth of the households with a main income source, or 15 percent of the applied sample, rely on Public transfers, which is approximately twice as many as those dependent on Private transfers. The share of the sample deriving their main income from the Primary sectors is 6 percent, two percentage points below which one finds the Mining and Quarrying and the Indirect income categories. The households that have salaries and wages from Domestic services as their main income source constitutes the smallest category at 2 percent of the sample.

The figures in Table 3 attest to low extents diversification. With the sole exception of "Indirect income" which is utilised among almost two-thirds of the sample, none of the other income source categories are accessed by as much as half the sample. However, the propensity for "Indirect income" to be a *main income source* is very low, with a small fraction of utilizing households in the two-thirds-or-more category. The figures in the contribution interval one-third-to two-thirds suggests that among 19

[7] In a dynamic perspective Ardington and Lund (1996) raise a valid objection to the use of a "dominant source of income" for the analysis of livelihoods since sources may be of a temporary nature.

Table 2: Percentage of households by their main income source category, for various main income cut-off contribution levels.

Main income contribution to total household income	Main income source category								
	No main income source	Core sectors	Mining and Quarrying	Primary Sectors	Domestic services	Public transfers	Private Transfers	Indirect income	Total
50%	5	41	4	10	3	19	8	10	100
66.7%	**24**	**38**	**4**	**6**	**2**	**15**	**7**	**4**	**100**
75%	33	34	3	5	2	14	6	2	100
90%	52	25	2	2	1	12	5	0	100
100%	72	16	1	1	1	7	3	0	100
Unweighted figures, n=19914.									

Table 3: Percentage of households with income from income source categories and contributions to total household income.

Income source	%-age share of households deriving income from source	Contribution (γ) to total income among households with source			Total	Fraction with source as main income source
		$0 < \gamma \leq 1/3$	$1/3 < \gamma < 2/3$	$2/3 \leq \gamma$		
Core sector	49	6	16	77	100	38
Mining/Quarrying	5	4	13	83	100	4
Primary sectors	15	18	43	40	100	6
Domestic services	11	53	28	19	100	2
Public transfers	31	27	23	50	100	15
Private transfers	17	39	22	39	100	7
Indirect income	65	75	19	6	100	4
Unweighted figures. n=19 914						

percent of the households that access Indirect income, its relative contribution is large enough to allocate the household into the Diversifying category. In the same column can be seen that substantial fractions of the diversifying households often access wage income from the Core or Primary sectors and Public transfers. The highest propensities to be main income sources are found in the Core sectors, Mining and Quarrying sectors, and the Public transfers categories where the source provides the main income in, respectively 77, 83, and 50 percent of the households with access

With respect to income from agricultural production it has been noted by Leibbrandt *et al* (2000), that agricultural income has not been well captured by the IES data. In the final sample here, 9.7 percent of the households had either slaughtered domestic animals or harvested crops in the last year. While profit from agricultural activities should be registered in the IES questionnaire under "self-employment", only 1.2 percent of the households that had slaughtered or harvested had records of *any* self-employment profits at all. Still, agricultural production for own consumption assumes several other important functions as *inter alia* a supplementary source of nutrition and as a safety net for vulnerable households in South Africa (May (1996)). Thus, the survey figures may understate the importance of agriculture. However, left with little choice other than taking the data at face value, agricultural production is not listed as a separate source of income. The few households that would have agricultural income as their main source are included in the core economy category among households with main income from other types of self-employment.

In conclusion there exist at least two reasons to consider the applied definition of main income source a useful concept in the description of households' income generation: Firstly, the contribution of total income from the main income source is twice as large as from any other source. Secondly, individual categories of direct income are typically accessed by small fractions of the sample which suggests that diversification is fairly limited.

5. Main income sources and income levels

This part of the study constitutes a descriptive analysis of the associations between variation in households' main income sources and income levels. Table 4 shows the distribution of the households in the sample across ten household income brackets according to the households' main income sources. The brackets are defined by the cut-off income levels between households per adult-equivalent income deciles in the *full* IES95 sample (including the Asian/Indian and white sample). Accordingly, the figures in the table can be read as, for instance, 16 percent of the households in this study that have a primary sector main income, belong to the poorest ten percent of the households in the full OHS/IES95 sample.

Table 4: Households' distribution across population per-adult equivalent household income deciles, by main income source category.

Main income source category	Income bracket										Sum	Mean income
	1	2	3	4	5	6	7	8	9	10		
Diversifying	11	17	17	16	13	11	7	4	2	1	100	6 023
Core sectors	*3*	*4*	*7*	*11*	12	16	*17*	*15*	*11*	*4*	100	12 854
Mining/quarrying	*1*	*1*	*4*	*4*	9	9	*27*	*29*	*14*	*2*	100	14 536
Primary sectors	*16*	*15*	*17*	*19*	14	12	5	2	0	0	100	4 462
Domestic services	*22*	*14*	*19*	*13*	11	13	7	3	0	0	100	4 458
Public transfers	*32*	*24*	*17*	*10*	12	2	1	0	1	0	100	3 031
Private transfers	*31*	*22*	*17*	*14*	8	5	2	1	0	0	100	3 265
Indirect income	9	12	13	16	9	13	9	7	6	6	100	11 490
All	*12*	*13*	*13*	*13*	12	11	10	8	5	2	100	8 408
Unweighted figures. n=19 914												

If one adds up the figures in the four lowest income brackets in Table 4, the overall fraction households in those brackets is found at 51 percent in the bottom row. The corresponding sum for households in either transfer income category is almost 85 percent, while for the Primary sectors and Domestic services categories the analogous fractions are approximately two-thirds. The share of Core sector households in the first four brackets is relatively low at one-quarter and that of the Mining and Quarrying sector is just over 10 percent. For the latter two categories, 60 percent and almost three-quarters respectively, are found in the fifth through eighth income brackets. Among the diversifying households some 60 percent are found in the first four brackets, with another quarter found in the consecutive two brackets. The

distribution of households that rely on “Indirect income” seem to follow closely to the all-over distribution of households in the sample.
In the last column of Table 4 are also listed the mean per adult-equivalent income levels among the households in the various main income source categories. The mean incomes reflect the distributions across the income brackets of the households within the different main income source categories.The mean incomes of households with Core sector or Mining and Quarrying main income sources are found at R12 854 and R14 536, which are both more than twice as high as the Diversifying households that average at R6 023. The households with main incomes from either Domestic services or the Primary sectors both have mean incomes very close to R4 460, whereas the Publics transfers and Private transfer main incomes on average yield R3 031 and R3 265 respectively. Given the similarity in the distribution across income brackets of the households in the Indirect income category to the full sample’s, it is surprising to find the mean in the Indirect Income at all of R 11 490, which is considerably higher than the all-over mean at R 8 408. An explanation may be found in the high variety of income sources included in the category.

The investigation of main income sources as explanatory factors for income levels is thus motivated by the apparent statistical associations between a household’s main income source and its position in the income distribution. The Core and Mining and Quarrying sector households in general appear considerably better off than households in the other categories. Households with transfer main income sources are to a high extent clustered among the very poorest, which is true also for households relying on main income from the Primary sectors or Domestic services. The mean incomes of households in the various income source categories also reflect the rank order in terms of income levels implied from the differing distributions across income brackets.

6. The reduced form approach to modeling household income levels - explanatory variables and analytical concerns

The objective of this study is to investigate if income sources, in conjunction with other household characteristics, can contribute to explain variations in households’ income levels. The value of the information attained by that investigation depends on

how well the household income generation process is modelled. While estimating the determinants of a different dependent variable – household welfare – Glewwe (1991) makes two points of relevance to the analytical approach of this study; the regression of income levels "on various explanatory variables assumed to be pre-determined or exogenous […] is simply a reduced form estimate of various structural relationships". Thus, at least two challenges enter the formulation of a model for household income generation. Firstly, in reality there may exist several links between the household and the realms of income generation. Secondly, empirical methodology should be designed to control for the potential lack of statistical exogeneity of the explanatory variables.

Modelling income generation and explanatory variables

The formulation of a structural model in the shape of an equation system, that specifies all conceivable links between a household and modes of per-adult equivalent income generation, would be preferable from a methodological viewpoint and include equations for e.g. labour force participation, fertility, migration decisions, earnings functions, and household production functions. Theoretical guidance exists for the formulation of models that represent such relationships *individually*. However, existing theory is lacking for how to best *combine* such relationships into a system of structural equations. Hence, for purposes similar to this study's, the reduced form has become common in the development economics literature.

From the above perspective, one requirement is that the applied right-hand side variables in as much as possible capture the links between the household on the one hand, and on the other, the labour market, access to public and/or private transfers, and the dependency ratios. A reduced form model for South African household incomes has been developed by Leibbrandt and Woolard (2001) who analyze log per-capita income in the OHS/IES95 data set and justify their choice of explanatory variables in detail. Motivated primarily by those authors' successful application, this study borrows most of the non-income source explanatory variables from their model. The list of variables common to all specifications in this study follows below with summary restatements of the motivation provided by Leibbrandt and Woolard (2001);

- Since previous analyses of South Africa have repeatedly shown that race is a dominant and persistent indicator of both poverty and inequality, a dummy variable for households belonging to the African population group is included.

- It has also been shown in other work on South Africa that the number of household members and specifically children are larger in less prosperous households (Dieden and Gustafsson (2003)). The explanatory variables therefore include the number of household members in age and gender categories. Age and gender categories are defined as follows: Children aged 0 -7 and 8 -15, females aged 16-59, and males aged 16-64, and elderly (above the upper limit of working age for both genders).

- Education appears in most specifications of individual earnings functions and has shown to be influential also at the household level in developing countries (Appleton (2001a)). The applied specification therefore includes shares of households' adults (16 years old or older) in categories for highest level of educational achievement. Education categories are designed for tertiary education, complete secondary, some secondary, some or complete primary education. The left-out category is the share of adults with no education.

- The extent of successful integration in the allocation of members into labour market employment and the burden to the household of non-employed members are captured by shares of households' adults that are unemployed or non-active by the expanded definition for unemployment.[8] The left-out labour market status category is the share of adults in employment.

- Earlier work has shown that incomes vary considerably between South Africa's rural and urban areas. Hence, all specifications include a dummy variable for rural location.

The inclusion of dummy variables representing each of South Africa's nine provinces (with KwaZulu-Natal as the reference province) in one of the specification is justified

[8] As opposed to the official definition of unemployment, the expanded definition encompasses also the non-working working-age population who are willing to work but have given up searching for employment due to the belief that there are no jobs available to them. By the official definition, the latter category would be non-participants.

by their different regional economies discussed in the next section. With respect to the explanatory variables that have been listed this far, expectations are that the signs of their coefficient estimates would match closely to those estimated by Leibbrandt and Woolard (2001). Hence, African population group should have a negative impact on income as would higher numbers of members, regardless of age and gender, with the exception of elderly. Positive impacts on income levels are expected from increasing shares of adults with higher levels of education. The opposite is expected for increasing shares of non-active or unemployed adults and for rural location. With respect to the estimates for provincial dummies, the analyses by Leibbrandt and Woolard (2001) returned no significant difference in income levels between the Western Cape, KwaZulu-Natal and Mpumalanga, and the only province with a positive level effect (as compared to KwaZulu-Natal) was Gauteng. The negative impacts were strongest for the Northern Cape and the Free State, and in rank from there followed the Eastern Cape, the North West Province, and Limpopo.

The variables representing households' utilization of income sources are included in the other two specifications. The inclusion of these variables is an attempt to investigate whether partial impacts on income levels exist, that originate in the utilization of income sources from the different categories, when controlling for other household characteristics that are assumed to affect income levels. In the latter group of variables are found those variables that may capture allocations to the categories of income sources and/or the return from these. The two specifications with income sources differ in the means by which income source categories are included. One of these specifications contains dummy variables for each Main income source category. The last specification contains six variables representing the continuous fractions of total income derived from each source. The specification thus serves as a control for whether the signs of the estimated effects for main income sources are also found for marginal increases in the shares of total income from the various sources. With respect to the expected *magnitudes* of the partial impacts of the various income categories, the outcome depends crucially on how well the other explanatory variables explain allocation or access to the income source categories. It appears intuitively appealing that impacts would match the *signs* the differences in their mean income levels, but no certain case can be made for such an outcome.

In summary a linear reduced form relationship between the variables is assumed to be of the following format:

$$Y = \mathbf{XB} + I_P \sum_{j=1}^{J} P_j \pi_j + I_S \sum_{m=1}^{M} S_m \xi_m + I_F \sum_{m=1}^{M} F_m \psi_m$$

where Y is the household's income level, $\boldsymbol{X}$ a k x 1 vector of the household's demographic and educational characteristics. The variable, P_j is an indicator taking on unit value if the household is located in province j and S_m is an indicator of whether the household derives income from source category m. The variable F_m represents the fraction of the household's income originating from source m. The 1x k vector $\boldsymbol{B}$ contains the slope parameters for each of the household characteristics in X, while π_j , ξ_m and ψ_m are slope parameters for province j and main income source category m and income fraction from the same category. The variable I_P is an indicator variable that takes on the value one if provinces are used as explanatory variables and zero otherwise. The variables I_S and I_F are analogous indicators for the income source variables.

Analytical concerns

This subsection discusses two complications that arise from the utilization of income sources as explanatory variables in regression analysis. The first concern is with the interpretation of coefficients for these variables and the second complication pertains to their possible statistical endogeneity.

Firstly thus, the current values of a number of the explanatory variables – such as labour force participation and income sources utilized –would be outcomes of structural relationships that model household-specific choices. Hence, the variables cannot be perceived as proper *determinants* of household income. An analysis, like this study, which does not identify the latter processes and determinants is in that sense incomplete (Glewwe (1991)). Consequently, parameter estimates for income source variables should be understood as explaining the variation in household income *conditional on the past decisions and events through which the household has been assigned its current main income source*.

The literature in this genre also recognizes that the assumption of exogeneity may not be realistic for many typical explanatory variables. Two common sources of endogeneity in applied econometrics are the omission of (unobservable but relevant) explanatory variables and the simultaneous determination of at least one explanatory variable along with the dependent variable (Wooldridge (2002)). In the latter category Appleton (2001b) points to e.g. land holding, adult household members' education levels (Behrman (1991)), and household demographics (Schulz (1983)). The analyses in this study attempts to control for the endogeneity of income sources, but there are limits as to what may be inferred and caution must be exercised in drawing conclusions.

With respect to the endogeneity of income sources, one reason to be wary is that income levels may affect the accessibility of certain income sources to households. Firstly, financial constraints may apply to increasing the range or returns of income sources for a household. This would e.g. apply for costs that are incurred by searching for employment away from the area of residence or by capital investments for self-employment. In addition, households' income levels may influence the extent to which they are entitled to means-based public grants. Similarly, the income levels of prospective private transfers receivers may also affect the decisions by remittance senders.[9] Plausibly, not all public transfers are subject to households' needs tests and factors other than receivers' income levels may affect the senders' decisions. In the end, however, it is still conceivable that causality runs in both directions.

As will be explained in more detail in Section 8, in order to control for endogeneity in the empirical analysis a household characteristic which is a strong covariate of household's income sources is needed. But the covariate should not itself be determined by household income levels. This study utilizes provincial location for that purpose and Section 7 serves to motivate the choice.

[9] See e.g. Stark (1995) for a discussion of transfer behaviour or Posel (2001) for a South Africa specific study of several hypotheses regarding transfer behaviour.

7. Main income sources and provincial labour markets

The multivariate analysis depends crucially on the correlation between households' geographical location by province and their main income sources. It is implicitly suggested that the latter variation originates in the provinces' labour market conditions. Transfer income dependence would be expected to be more prominent where unemployment is high and/or participation rates are low. Similarly the provinces' composition with respect to employment by major economic sector should be reflected in households' wage main income sources. Descriptive statistics in this section serve to illustrate these occurrences.

In terms of physical geography the nine provinces of the present day South Africa are very different, with considerable variation in economic activities. As can be seen in Table 6, the four most populous provinces – the Eastern Cape (E Cape), KwaZulu-Natal, Gauteng and Limpopo – contain nearly 65 percent of the working-aged population[10], but with very dissimilar distributions across rural and urban areas. In the E Cape, KwaZulu-Natal, the North West Province (NW Province), Mpumanlanga, and Limpopo, most of the population is rural, although KwaZulu-Natal contains the metropole Durban, the third largest city in South Africa. At the other extreme are found the largely urbanised provinces of the Western Cape (W Cape) and Gauteng, which are the two leading provinces economically. They respectively host Cape Town and the conurbanised area of Johannesburg, Witwatersrand and Pretoria, in the proximity of which are found also many of the former gold mines.

The third "Cape" province, the Northern Cape (N Cape) is scarcely populated but highly urbanized. The province contains largely desert and savannah areas, but also some of the country's vast diamond findings near its capital, Kimberly. From there the bushy highland landscape, the "Karoo", extends into the largely agricultural but also relatively urbanized Free State, the provincial capital of which is Bloemfountain, the country's legislative capital during the apartheid era. Other fertile farming areas are found south and east of the coastal mountain ranges in the Eastern and W Cape and in KwaZulu-Natal, which in turn also host the prosperous and industrial coastal cities

[10] By the gender specific age-criteria Old Age Pension access South Africa, working-aged are defined as 16-59 years for women and 16-64 for men.

Port Elizabeth, Cape Town and Durban, all of which are among the largest ports of the African continent.

As shown in Tables 6 – 8 provincial location is a covariate of a number of labour market features. Starting in Table 6, the figures in the first column indicate that four provinces contain almost 65 percent of the total working-age sample (henceforth the "most populous provinces"). Hence, the conditions in these provinces have a large impact on the extent to which provinces covary with Main income source categories. Table 6 also illustrates how the working-age population in one of the most populous provinces, Gauteng, is mostly urban. As can be seen in Table 7, the participation rate in Gauteng is alos high and the expanded unemployment rate is among the lowest ones, while its official ditto is just below average. Excluding employment in the Primary sectors, Households, and Mining and Quarrying in Table 8, one finds 79 percent of the employed in Gauteng in the Core sectors with another 9 percent in Mining and Quarrying.

On the other hand, in Limpopo and the E Cape, two of the other most populous provinces, rural dwellers dominate the working-age population, the participation rates are low, and the provinces have the two highest rates of expanded unemployment. It is, however, noteworthy that the official unemployment rate at 27 percent in the E Cape is one-and-a-half times that of Limpopo. The fractions of Core sector employment in the two provinces are of similar size at approximately two-thirds. Half of that employment is in both cases found in Public service which leaves the provinces ranked as number one and two in terms of such employment.

In the remaining most populous province, KwaZulu-Natal, the rural dwellers constitute 70 percent of the working- age population. Both unemployment rates are high and the employed are underrepresented among the working-aged, but not by as much as in Limpopo or the E Cape. At 68 percent the province' fraction of Core sector employment is large and both the Private and Public services sectors as well as the Secondary sectors rank as number three among the provinces.
Table 9 shows the distribution of Main income categories in the provinces. In accordance with the above features one finds 62 percent of all households in Gauteng supported by Core sector employees and another 10 percent with main income

sources from Mining and Quarrying. On the other hand, dependence on transfer incomes is very large in the E Cape and Limpopo, at 42 percent and 32 percent respectively, while less than one-third of the households in either province have Core sector main incomes. KwaZulu-Natal has the fourth highest fraction of households depending on either type of transfers, but at 21 percent the share is distinctly lower than that of Limpopo. Two-fifths of the households in KwaZulu-Natal are supported by Core sector income earners and, perhaps reflecting the province' compound nature of low urbanization but a large core, its 28 percent fraction of Diversifying households is the third largest.

Table 6: Sample shares of working-age population distribution across rural and urban areas, by provinces.

Province	Rural	Urban	All	Share of working-age sample
W Cape	17	83	100	9
E Cape	67	33	100	17
N Cape	32	68	100	2
Free State	46	54	100	7
KZN	70	30	100	19
NW Prov.	70	30	100	9
Gauteng	7	93	100	14
Mpumalanga	79	21	100	8
Limpopo	92	8	100	13
All	57	43	100	100
Total no.	11 492 000	15 043 000	26 535 000	
Weighted figures, n= 52 919.				

Table 7: Sample shares of working-age population and employed with labour force participation and unemployment rates across provinces.

Province	Official participation rate	Official unemployment Rate	Expanded unemployment Rate	Share of employed
W Cape	65	15	22	13
E Cape	36	27	46	12
N Cape	54	22	32	2
Free State	55	13	28	10
KZN	45	24	39	17
NW Prov.	46	17	35	9
Gauteng	63	18	28	19
Mpumalanga	43	18	38	8
Limpopo	34	19	40	10
All	47	19	35	100
Total no.	14 019	2 423	5 469 000	10 093 000
Weighted figures, n= 52 919.				

Table 8: Distribution of employment among identified earners in the sample by sectors and provinces.

Province	Primary Sectors	Mining/ Quarrying	Secondary Sectors	Private services	Public Services	House-holds	Self-Employment	Total	All
W cape	20	1	29	22	17	9	3	100	13
E Cape	21	1	11	16	31	13	7		12
N Cape	38	7	9	14	14	16	2		2
Free State	34	8	7	11	15	22	2		10
KZN	15	1	21	20	25	11	6		17
NW Prov.	23	10	12	19	19	11	6		9
Gauteng	3	9	24	29	21	9	6		19
Mpumalanga	30	7	18	15	13	13	5		7
Limpopo	19	6	9	16	34	8	8		9
All	18	5	18	20	22	12	5		100
Weighted figures, n = 18 776.									

Table 9: Distributions of main income source categories and mean income levels across provinces.

Province	Main income source category									Mean income
	Diversifying	Core sectors	Mining/ Quarrying	Primary Sectors	Domestic Services	Public transfers	Private transfers	Indirect income	All	
W Cape	23	52	0	9	2	10	1	2	100	10 090
E Cape	21	27	1	4	2	28	14	4	100	5 846
N Cape	29	23	4	16	3	17	3	5	100	6 350
Free State	37	24	7	5	3	13	4	7	100	6 261
KZN	28	40	1	5	2	15	6	4	100	8 084
NW Prov.	28	32	8	6	1	13	8	4	100	8 099
Gauteng	15	62	10	2	3	5	1	3	100	14 035
Mpumalanga	22	35	6	16	3	12	4	1	100	5 719
Limpopo	21	30	3	6	1	20	12	7	100	8 195
All	24	38	4	6	2	15	7	4	100	8 408
Weighted figures, n =19 914.										

With respect to some of the other provinces, the W Cape, which hosts 9 percent of the working age sample, shares many of the labour market features of Gauteng. The province has no households with Mining and Quarrying main incomes, but approximately half the households in the W Cape have Core sector main incomes, while 9 percent rely on Primary sector income. The NW Province hosts a fraction of the working-aged which is similar to W Cape's and the shares of participants and rural dwellers are similar to those of KwaZulu-Natal. However, the fraction of employees in the Core sectors in the NW Province is lower, as is the approximately one-third share households with corresponding Main income sources. Among the employees in the same province one-tenth are found in Mining and Quarrying sectors, with a similar fraction of households' Main income sources.

Almost one-quarter of the employees in the NW Province are found in the Primary sectors, but the share of households that depend on the same sectors for the main income is only 6 percent. A similar tendency applies to the Free State. Attesting to the low propensity of such sectors to provide main incomes, shown in Table 3, the extents of Diversification are high in both these provinces, as well as in the population-wise miniscule N Cape. However, primary sector employment is high also in Mpumalanga, but the province' share of diversifying households is the seventh lowest. Rather, Mpumalanga's 16 percent fraction of households with Main income sources from the Primary sectors ranks as the highest in that category along with the N Cape.

In conclusion, some extent of regularity can be detected between the mean income levels of the various provinces and their composition with respect to Main income sources. Incomes are highest in Gauteng and the W Cape, at R14 035 and R10 090 respectively, where main incomes from Core sector are most common. At the opposite end one finds the E Cape with high dependence on transfers and the average income at R 5 846. In the N Cape and the Free State average income levels are also low. This may be partly explained by the small fractions of households supported by employees in the Core sectors, by high prevalence of Diversification and Primary sector main incomes, as well as by the provinces' displaying the fifth highest dependency on Public transfers. The lowest mean income of R 5 846 is found in Mpumalanga, however it does not appear to be associated with any other distinct features than the large fraction of households that rely on Primary sectors for their

main income. The remaining three provinces, all have main incomes in the close proximity of R8 100. Thus, while the relationship between provincial mean income levels and composition of Main income sources may be somewhat imprecise, the latter composition itself varies discernibly across provinces.

8. Empirical approach

The empirical analysis in this section is undertaken by the comparison of results from three different multivariate regression model-specifications. The first specification includes province dummy variables and serves as a benchmark (henceforth "the geography specification"), whereas the other two are two-stage specifications that include different representations of income sources as explanatory variables.

One of the specifications with income source variables uses dummy variables for the household's main income category (henceforth "the dummies specification") and the other uses continuous fractions of income derived from *all* of the seven categories of income (henceforth "the fractions specification"). As discussed in Section 6, the analyses must be undertaken with tests and, if necessary, controls for the endogeneity of the income source variables.

The analyses are undertaken by weighted least squares regression analyses, in which a transformation function between the log per adult-equivalent household income levels and household characteristics is postulated. The general relationship is modelled as:

$$1) \qquad Y_i = X_i\beta_m + D_s^1\sum_{j=1}^{9} P_{ji}\pi_j + D_s^2\sum_{k=1}^{8} M_{ki}\eta_k + D_s^3\sum_{m=1}^{7} C_m\varphi_m + \varepsilon_i \qquad \begin{matrix} i = 1,2...n \\ s = 1,2,3 \end{matrix}$$

where Y_i represents log annual per adult-equivalent income for household i and X_i is a vector of household characteristics variables common to all specifications. D^1 is an indicator variable with value one in the geography specification and zero elsewhere. Analogously D^2 and D^3 take on unit value for the dummies specification and income share specification respectively, and zero elsewhere. The province dummies are

symbolized by P, where P_{ji} takes on unit value if household i resides in province j. The symbol M applies to the main income source category dummy variables, and M_{ki} takes on unit value if income from category k contributes 66.7 percent or more to the total income in household i. The continuous income fraction derived from source m is represented by C_m. The empirical model also contains the three vectors of slope parameters π, η, and φ, for the provinces, main income source categories and fractions of income from the various sources respectively.

The error term, ε_i in equation 1) is usually assumed IID with zero mean across observations and uncorrelated with the explanatory variables. In this respect a further complication arises from the household surveys' two-stage, stratified sampling design and the delimitation of the sample analysed here. The first population-group related criteria for selection into this study renders the subsample no longer representative of the whole South African population, for the representation of which the weights were designed. As a consequence of the second (identification) criteria, there is no guarantee that the subsample is even representative of the corresponding population groups. However, the weights supplied with the data contain information about the *relative* representativity of observations, based on population proportions of observations in each strata and cluster. While not returning a representative sample, the application here of the original weights renormalized to sum to unity is a feasible attempt to correct for the relative over-representation of some households. The application of the weights furthermore allows for the incorporation of controls for stratification and clustering effects into the analyses, as recommended by Deaton (1997) when a survey sample contains unusable values.[11]

Testing and controlling for endogeneity [12]

In equation 1) an explanatory variable x_k is said to be endogenous if it is correlated with the error term ε (i.e. $E(x_k \varepsilon) \neq 0$). Endogeneity usually arises in applied

[11] In general, stratification will typically enhance the precision of sampling estimates, while clustering usually will increase standard errors. The reason for the latter is that households living in the same cluster are usually more similar to one another than are households living in different clusters, due to covariation in behaviours or characteristics related to e.g. agro-climatic conditions, prices or ethnicity. Hence, less information is obtained when several households are sampled from the same cluster, than would be the case if they were randomly sampled from different clusters, and the precision of estimates thus depends on the correlation within clusters of quantities being measured. In the presence of such correlation, estimators need be used that incorporate weights and reflect lower degrees of freedom in tests of significance (Deaton, (1997)).

[12] The section on endogeneity draws heavily on Wooldridge (2002: 50-51, 118-120,472-478)

econometrics in one (or more) of three ways; omitted variables, measurement error and simultaneity (Wooldridge (2002)). While the distinction between these three forms of endogeneity may not always be sharp, the concern here is with the last issue. If y is determined by x_k, but x_k also determined partly by y, then x_k and ε will be correlated.

The regression based test of endogeneity applied here has been developed by Hausman (1978, 1983). With the endogeneity suspect, x_k, relabelled y_2, the set-up is in brief as follows;

2) $$y = \mathbf{Z}_1\boldsymbol{\delta} + \lambda y_2 + u$$

3) $$y_2 = \mathbf{Z}\boldsymbol{\gamma} + v_2$$

where $\mathbf{Z}_1$ is a vector of explanatory variables, the δ and γ vectors and the scalar λ are slope parameters, u and v_2 are vectors of unobserved IID disturbance terms with zero mean. Equation 2) is the population model of interest (a simplification of equation 1)) and equation 3) is the linear projection of y_2 on a vector $\mathbf{Z}$ of exogenous explanatory variables. For the identification of 2) and 3) when y_2 is endogenous, crucial assumptions are that the variables in the $\mathbf{Z}_1$-vector are a subset of $\mathbf{Z}$ which in turn contains at least one element not in $\mathbf{Z}_1$. The latter element must be partially correlated with y_2, but not simultaneously determined with y.The maintained exogeneity of Z implies crucially that $E(\mathbf{Z}'u) = 0$ while the concern here is with the validity of $E(y_2u) = 0$. Since $E(\mathbf{Z}'u) = 0$ and assuming $E(y_2v_2) = 0$ Wooldridge (2002) shows that that y_2 is endogenous if $E(u\, v_2) \neq 0$.

The linear projection of u onto v_2 in error form can be written

4) $$u = v_2\rho_1 + e_1$$

where $\rho_1 = E(v_2u) \big/ E(v_2^2)$ and it can be shown that $E(v_2e_1) = 0$ and $E(\mathbf{Z}'e_1) = 0$.

With 4) inserted into 3) exogeneity of y_2 can be maintained only if $\rho_1 = 0$ in:

2') $$y = \mathbf{Z_1}\boldsymbol{\delta} + \lambda y_2 + \rho_1 v_2 + e_1$$

Following an OLS regression of equation 2') a *t*-test on the variable v_2 provides a test of the null hypothesis: $\rho_1 = 0$. The problem that v_2 is not observed is solved by replacing v_2 with the residuals from an OLS regression of the first-stage equation (equation 3)). The test easily extends into an *F*-test of several endogeneity suspects, where the incorporation of each endogeneity suspect into the system requires an additional first-stage equation with an additional exogenous element in Z not in Z_1. In the cases of continuous dependent variables, the endogeneity of a variable y_2 may be controlled for by *replacing the variable with its predicted value from the first-stage OLS regression* or in the case of binary endogenous variables, with its corresponding predictions from a probit first-stage regression.

In Section 4 it was shown that only Indirect income was accessed by more than half the sample. This is means that the share derived from each of the other income source categories is equal to zero for more than half the households. Hence, modelling contributions from the income source categories constitutes a typical sample selection problem, if the same variables which explain the *magnitude of the fraction* of total income derived from a specific income source also explain a households' *utilization* of the source. Similarly, the fact that a household accesses a certain income source does not by necessity imply that the income source is the household's *main* income source. Hence, the analogous sample selection problem arises if the same variables which explain why a utilized income source becomes a *main income source* would explain a households' *utilization* of the source.

Under the above circumstances, the estimated coefficients for the first-stage equations would be biased and predictions faulty unless measures are taken to control for sample selection. Hence, for both the binary and continuous income variables, the first-stage equations utilize two-step selection-correction procedures (Heckman (1979), Breen (1996)). The share of adult females in the household and a dummy variable indicating migrant household head are used in order to ensure identification in the Heckman-procedures. Summary statistics of all explanatory variables are found in Table 10.

Table 10: Summary statistics of sample characteristics and explanatory variables.

VARIABLE	Mean	Std.dev.	Values if main income source	
			Mean	Std.dev.
Log per adult equivalent income	8.533	0.957		
African	0.885	0.319		
Number of children 0-7 in household	0.826	1.048		
Number of children 8-15 in household	0.941	1.107		
Number of female adults in household	1.412	1.082		
Number of male adults in household	1.242	1.008		
Number of elderly in household	0.313	0.586		
Share of adults with no education	16.308	28.728		
Share of adults with primary education	42.460	37.852		
Share of adults with secondary education	25.107	30.972		
Share of adults with matriculation	10.660	22.779		
Share of adults with tertiary education	5.465	18.547		
Share of working-age adults unemployed	18.106	28.732		
Share of working-age adults not participating in labour force	52.697	37.909		
Rural location	0.547	0.498		
W Cape	0.091	0.287		
E Cape	0.179	0.383		
N Cape	0.021	0.144		
Free State	0.087	0.281		
KwaZulu-Natal	0.174	0.379		
North-West Province	0.098	0.297		
Gauteng	0.149	0.356		
Mpumalanga	0.070	0.255		
Limpopo	0.132	0.339		
Diversifying (No main income source)	0.239	0.427		
Core sectors main income source	0.379	0.485		
M & Q sectors main income source	0.040	0.197		
Primary sectors main income source	0.059	0.236		
Domestic services main income source	0.021	0.144		
Public transfers main income source	0.154	0.361		
Private transfers main income source	0.066	0.249		
Indirect income main income source	0.041	0.198		
Percentage fraction of total income from Core sectors	39.736	43.798	92.271	9.720
Percentage fraction of total income from Mining and Quarrying	4.022	18.267	90.002	9.0615
Percentage fraction of total income from Primary sectors	8.798	23.052	84.357	10.161
Percentage fraction of total income from Domestic services	4.392	15.386	86.380	11.464
Percentage fraction of total income from Public transfers	19.393	34.414	93.048	10.002
Percentage fraction of total income from Private transfers	9.063	24.939	94.458	8.824
Percentage fraction of total income from Indirect total income	14.597	21.173	77.231	7.689
Share of adult females in the household	35.042	22.383		
Migrant head	0.077	0.266		
Unweighted figures. n=19914				

9. Empirical results

This discussion of the empirical results will commence with a comparison of the general fit of the three regression models. Thereafter will follow a summary presentation of the results for the variables common to all three specifications, after which the focus will rest with the impact of income source categories on income levels and what may learned about these impacts from comparisons to the results from the geography specification.

The results from the endogeneity tests did not support the *exogeneity of the income source variables in either specification* at any pertinent level of significance. (The test results are found in Appendix 3.) The analysis therefore proceeds with the observed income source variables replaced by the first-stage predictions. The output from all three model specifications is presented in Table 11. A future sophistication of this analysis is a log-likelihood estimator which simultaneously computes all three steps of the estimation procedure (including the selection-correction procedure in the first-stage equations). Currently, the predicted income source variables are incorporated through a *non-simultaneous* two-step procedure, which leaves the second-stage standard errors smaller than would a simultaneous estimator. Hence, the test-statistics are not strictly valid (Wooldridge (1999)). In order to alert the reader of this caveat the relevant cells in Table 11 are shaded grey. (The same caveat and notation applies to Table A3.1.)

With respect to the fit of the models, the vast majority of the estimates are significant at the one percent level. Five estimates are non-significant. There are two estimates that are significant at the five percent level and one at the ten percent level. The values of the coefficient of determination are similar for all three specifications with the highest value at 0.558 for the geography specification and 0.548 for both income source specifications. Hence, while the dummies specification contains a higher number of less significant estimates and the geography specification explains one percent more of the variation in the dependent variable, households' income sources appear to in effect contribute to explaining the variation in log per adult-equivalent income levels as well as does provincial locations.

Table 11: Least squares regressions with predicted main income source variables.

Dependent variable: log per adult equivalent income			
	GEOGRAPHY SPECIFICATION	**DUMMIES SPECIFICATION**	**FRACTIONS SPECIFICATION**
F-value	570.07	573.58	618.26
Prob > F	0.0000	0.0000	0.000
R-squared	0.5582	0.5475	0.5478
VARIABLE	**Coeff. (Std.Err.)**	**Coeff. (Std.Err.)**	**Coeff. (Std.Err.)**
African	**-0.176*****(0.026)	**-0.227*****(0.026)	**-0.179*****(0.024)
Number of children 0-7 in household	**-0.082*****(0.006)	**-0.067*****(0.006)	**-0.095*****(0.006)
Number of children 8-15 in household	**-0.073*****(0.005)	**-0.063*****(0.006)	**-0.087*****(0.006)
Number of female adults in household	**-0.083*****(0.006)	0.015 (0.010)	**-0.039*****(0.008)
Number of male adults in household	**-0.015*****(0.006)	0.015 (0.010)	**-0.039*****(0.008)
Number of elderly in household	**-0.072*****(0.010)	**0.083*****(0.026)	**-0.090*****(0.024)
Share of adults with primary education	**0.002*****(0.000)	**0.000*** (0.000)	0.000 (0.000)
Share of adults with secondary education	**0.006*****(0.000)	**0.002*****(0.000)	**0.002*****(0.000)
Share of adults with matriculation	**0.011*****(0.000)	**0.006*****(0.000)	**0.007*****(0.000)
Share of adults with tertiary education	**0.018*****(0.000)	**0.012*****(0.000)	**0.013*****(0.001)
Share of working-age adults unemployed	**-0.002*****(0.000)	**-0.002*****(0.000)	**-0.002*****(0.000)
Share of working-age adults non-participants	**-0.008*****(0.000)	**-0.010*****(0.000)	**-0.008*****(0.000)
Rural location	**-0.296*****(0.020)	**-0.073*****(0.027)	**-0.090*****(0.026)
W Cape	**-0.160*****(0.036)		
E Cape	**-0.297*****(0.025)		
N Cape	**-0.380*****(0.047)		
Free State	**-0.465*****(0.033)		
North-West Province	**-0.161*****(0.034)		
Gauteng	**0.073**** (0.033)		
Mpumalanga	**-0.143*****(0.041)		
Limpopo	-0.001 (0.043)		
(^) Diversifying (No main income source)		**-2.016*****(0.194)	
(^) M & Q sectors		**0.764*****(0.097)	**0.008*****(0.001)
(^) Primary sectors		**-0.602*****(0.118)	**-0.011*****(0.001)
(^) Domestic services		**-0.998*****(0.261)	**-0.015*****(0.002)
(^) Public transfers		**-0.330*****(0.062)	**-0.004*****(0.001)
(^) Private transfers		**-0.210**** (0.087)	**-0.003*****(0.001)
(^) Indirect income		0.440 (0.602)	**-0.011*****(0.002)
Intercept	**9.305*****(0.043)	**9.716*****(0.056)	**10.012*****(0.064)

Significance levels are based on standard errors that take the clustered and stratified nature of the sample into account.
Estimates in **bold** are significant at the 10% level or higher. ***/** / * Estimate significant at 1%/5%/ 10% level
(^) Predicted variable. Weighted data. n = 19 914.

Six of the eight coefficient estimates for provincial location are significant at the one percent level and a seventh is significant at the five percent level. Out of the seven estimates for the Main income source dummy variables, five are significant at the one percent level, one at the five percent level and one estimate - for the Indirect income category - is not significant. All coefficients for the income fractions from the different income sources are significant at the one percent level – including a negative impact from Indirect income. Hence, in all but one case do the *main income source categories have significant partial impacts on households' log per adult-equivalent income levels* (henceforth "income" levels). The estimated impacts in the fractions specification have the corresponding signs with the one aforementioned exception.

Among the significant coefficients for the variables common to all specifications, all but one *sign* - for the number of elderly in the household – are identical across specifications. The signs of the estimated coefficients for the variables common to all specifications are in congruence with the signs expected and those attained by Leibbrandt and Woolard (2001). Since those authors discuss their results in detail the interested reader is referred to that study.[13]

The estimated impact from African population group on income levels is negative across all specifications. The same is true for increased numbers of children in both age categories. The geographical and the fractions specifications share two sets of results which deviate for the dummies specification; increasing *numbers* of working-age adults of either gender display significant negative impacts on income, while the number of elderly in the household has a significant positive effect. In the dummies specification, the first two estimates are non-significant, whereas the presence of elderly is associated with a significant decrease in income.

Increasing shares of adults in all educational categories *above* primary education show positive estimated effects in all specifications. With respect to fractions of adults with primary education, the only significant coefficient is positive and found in the geography specification. Increasing shares of working-age adults in both the unemployed and the non-active labour market categories have negative effects across all specifications. The strength of the impact is however four to five times higher for the non-active than for the unemployed. Finally, the impact of rural location is negative in all specifications but is more than three times as strong in the geography specification as in the other two. Hence, it appears as if *differences in income sources accounts for much of the difference in income attributed to rural location* in the geography specification.

[13] As mentioned the geography specification in this study is similar – although not identical - to that devised by Leibbrandt and Woolard (2001). The estimates in the geography specification are sign-wise in general congruence with those of derived by Leibbrandt and Woolard (2001). The results here differ in that significant, negative estimates are found for the number of elderly in the household and a significant difference is found between the partial impacts from residence in the W Cape as compared to KwaZulu-Natal. Leibbrandt and Woolard (2001) use the W Cape as the provincial base category.

The analysis now turns to the impacts on income levels from the Main income source categories. The Mining and Quarrying sectors category has the only positive estimated level effect. For all the other categories the estimated coefficients are negative. The analysis of these results is aided by the figures in Table 12. The first column of figures in the table shows the difference in mean incomes between the main income category under consideration and the Core sectors category, with the difference expressed as a fraction of the latter mean income (see Table 4). A negative sign indicates that the mean income was lower than in the core sector category. For instance, the figure "-53" is to be interpreted as the mean income for the category in question was 53 percent lower than the mean income in the Core sectors category.

Table 12: Differences in observed mean income and estimated partial percentage level effect on income for the different Main income categories as compared to the Core sectors category

Main income source category	Difference in observed mean Income (%)	Estimated partial (level) effect (%)
Diversifying	-53	**-87**
M & Q sectors	13	**115**
Primary sectors	-65	**-45**
Domestic services	-65	**-63**
Public transfers	-76	**-28**
Private transfers	-75	-19**
Indirect income	-11	55

The second column contains the partial coefficient estimate in percentage form for the various Main income source categories.[14] (The legend for significance in Table 11 applies and is repeated for the estimates from which the effects are computed.) The differences in magnitudes between the two statistics provide an indication of the extent to which the other variables in the specification explain the differences in observed incomes. The estimated partial coefficients are of lower absolute magnitude than the difference in mean incomes both for the two transfer categories and for the households in the Primary sectors category. Those results may be interpreted as if it were not for their *systematically relatively unfavourable endowments of controlled-for*

[14] The coefficients for the impacts of incomes sources in the dummy and fractions specifications are not directly comparable. It follows trivially from differential calculus that percentage effects on income levels from the income-fractions variables can be arrived at by multiplying the estimates by one-hundred. In order to arrive at the percentage impact from the dummy variables unity should be subtracted from the antilog of the estimate and the difference multiplied by one hundred (Halvorsen and Palmquist (1980)).

characteristics (henceforth "other characteristics"*)*, the households in the latter three categories would have been less worse-off as compared to households in the Core sector households.

Along the same line of reasoning the results indicate that the Mining and Quarrying sector households would be *even better off* were they not systematically endowed with *more unfavourable* other characteristics. The opposite appears to apply for Diversifying households which would have been even *worse* off, were it now for the relatively *favourable* endowments of other characteristics. No such phenomenon appears to apply for the Domestic services category and the lacking significance of the estimated coefficient for the Indirect income allows no certain interpretation. These results are largely confirmed by the results from a corresponding analysis for the estimates in the fractions specification, the details of which are found in Appendix 4.

With respect to the provincial locations in the geography specification, the only positive coefficient estimate is that for residence in Gauteng. A household's location in the W Cape, the NW Province, or Mpumalanga is however associated with partial negative impacts on income. The strongest negative impacts are those of the E Cape and the Free State, while the magnitude of coefficient estimate for the N Cape is roughly half-way between the latter two and the coefficient for Limpopo is not significant.

Table 13 is analogous to Table 12 and displays the results for the various provinces as compared to KwaZulu-Natal. With the exceptions of the E Cape and Mpumalanga, the estimated partial impacts for all provinces are less than the corresponding differences in mean incomes. Hence, if it were not for the more favourable endowments of other characteristics, the households in Gauteng would have been *less better off*, the households in the Western Cape *worse off*, and those in the N Cape, Free State and NW Province would have been *even worse off*, as compared to the left out province. However, the households in the E Cape and Mpumalanga would relatively speaking have *been better off* if it were not for their systematically more unfavourable endowments.

Table 13: Differences in observed mean income and estimated partial percentage level effect on income for the different provinces as compared to KwaZulu-Natal

Province	Difference in observed mean income (%)	Estimated partial effect on income(%)
W Cape	25	**-15**
E Cape	-28	**-26**
N Cape	-21	**-32**
Free State	-23	**-37**
North-West Province	0	**-15**
Gauteng	74	8**
Mpumalanga	-29	**-13**
Limpopo	1	0

A multitude of reasons may exist for differences across the provinces in both remunerations in various activities and household characteristics. However, the W Cape and Gauteng have the largest fractions of households in the Core sectors category, while the E Cape and Mpumalanga have the largest fractions of transfer dependent households. The above results are thus all consistent with *households in the Core sector category systematically having more favourable endowments of other characteristics than do the transfer dependent households*.

Finally, this analytical approach does not identify the explanatory variables that explain the differences between the observed differences in mean incomes and the partial impacts of the various main income source categories (or provinces). However, the differences in magnitudes of the estimated coefficients for rural location and the education categories across the specifications are of interest in this context. The absolute magnitudes of the impacts for both characteristics are smaller in the income source specification. Thus, some of the effects attributed to these variables in the geography specification may originate in the roles rural location and the education levels of adults play in households' allocation to main income source categories.

10. Conclusions

An objective of this study has been to investigate whether the extent to which African and coloured South African households are integrated into the core economic sectors affect their income levels. The results from both descriptive and multivariate analyses confirm this conception, contingent the notion that households which rely on

employees in the Mining and Quarrying sectors should be considered integrated in the core economy.

Four results that were found in initial descriptive analyses were utilized in the design of this study's analytical approach; Firstly, a vast majority of the households in the sample derive two-thirds or more from a single income source - a "main income source" - and secondly, most households do not access other regular sources of income. Thirdly, households' main income sources appear to be closely associated with their positions in the income distribution. Fourthly, the distribution of main income sources within provinces differs considerably across provinces.

Descriptive analyses based on a classification of main income sources according to households' core-sector integration showed that the least integrated households, which depend on transfer incomes of either public or private origin, are concentrated at the lower end of the income distribution with the lowest mean incomes. Households which were found at the periphery of the core, with labour income from either the Primary sectors or Domestic services, were only slightly better off than those in the transfer categories. Diversifying households are found in a slightly better-off position than the latter two categories. The households that derive their main income from the South African Core sectors or from the Mining and Quarrying sectors were located in higher positions in the income distribution.

The results from multivariate regression analyses, in which controls were applied for the endogeneity of income sources to income levels, showed that virtually all variables representing households' income sources were found to have *significant partial impacts on households' log-income levels*. Secondly, the impacts *differ* across the different Main income source categories and also specifically between those which constitute wage-income of different origins. Hence, *not only access to wage income is important to South African household income levels, but the sector of origin for that wage is also of considerable consequence*.

Three different regression specifications were applied in the analysis. Two of these included income source variables in different formats while the third included provincial dummy variables rather than income source variables. In comparison,

income sources appeared to *contribute to the explanation of variation in log- per-adult-equivalent incomes virtually as well as do provincial locations*. A much stronger negative association between rural areas and income levels was found in the regression specification with provincial location, than in those with income source variables. Thus, the results suggest that *differences in households' income sources account for much of the difference in income levels between rural and urban areas*.

The estimated partial impacts from several of the income sources were compared to the differences in mean income levels between the different Main income source categories and the Core sectors category. The partial impacts from several of the income sources most prevalent in lowest income quintiles were *less* negative than the difference in observed mean incomes. The positive, partial impact associated with the Mining and Quarrying sectors category however, was greater than were the corresponding difference in mean income. Thus, implications are that if it were not for their *systematically relatively unfavourable endowments of controlled-for characteristics*, the households in the Public and Private transfers categories and in the Primary sectors, would be less worse-off relative to the households in the Core sectors category, while the households in the Mining and Quarrying sectors would be even better off. An analogous comparison of the relative differences between estimated partial effects and mean income levels in the different provinces supported the above interpretation.

Thus, the results of the analyses in this paper suggest that households' main income sources may serve as "flags" that signal different household constellations, as are sometimes thought to do the gender or presence of the household head. Indications from the comparison of the different model specifications are that households' location in the rural-urban dimension and education levels among adults important are, in addition to provincial location, important statistical associates of households' types of main income sources.

The results of this analysis support the relatively recently undertaken increases in minimum wages for domestic and farm workers, provided that the measures do not have adverse effects on employment in those sectors. Similarly are measures that facilitate and increase both public transfers, conditional on their appropriate funding, also supported. With respect to implications for further research, investigations into the

processes that underlie household formation and the allocation of income sources to households are likely to improve our understanding of income generation at the lower end of the South African income distribution. The same applies also to development of models for the possibly simultaneous determination of income levels, income sources, and household characteristics among the South Africa's indigent households.

Taken together, the results from this study's endogeneity tests, two-stage regression analysis, and the comparison of the latter to the descriptive statistics suggest that income generation among some South African households is subject to severe constraints. Households that rely on transfers or wage-income from the primary sectors are associated with low average per adult-equivalent incomes. At the same time, the predicament of these households also appears to be associated with rural location, certain provinces and, low education levels among their adult members. Hence, the geographical distance between these households and the core economic sectors may imply search costs which are too high for their existing income levels. In addition, it is conceivable that the legacies of spatially biased educational provision under the apartheid era render the members of such households very poorly qualified for market-labour. Thus, the nature of economic growth that would generate employment for the poorest would have to create jobs accessible to marginalized and/or peripherally located, unskilled labour.

References

APPLETON, S., 2001a. "What can we expect from UPE?" in P. Collier and R. Reinnikka (eds.) *Uganda's recovery: The role of farms, firms, and Government,* Washington DC: World Bank.

APPLETON, S., 2001b. "'The Rich Are Just Like Us, Only Richer': Poverty Functions or Consumption Functions.", *Journal of African Economie*s 2001:10.

ARDINGTON, E & LUND, F,1996. "Questioning rural livelihoods", Ch.2 in Lipton *et al* (1996).

BEHRMAN, J.R.,1991. Women's Schooling and Non-Market Productivity: a Survey and Reappraisal, *Women in Development Working Paper No.1*, Washington DC, Worldbank.

BHORAT, H., 2000. "Public expenditures and poverty alleviation: simulations for South Africa" in *Development Southern Africa* 17:5

BHORAT, H, LEIBBRANDT, M, MAZIYA, M, VAN DER BERG, S & WOOLARD, I , 2001. *Fighting Poverty – Labour Markets & Inequality in South Africa.* Cape Town, South Africa: UCT Press.

BIGSTEN, A., KEBEDE, B. AND SHIMELES, A., 2003. "Growth and Poverty Reduction in Ethiopia: Evidence from Household Panel Surveys" in *World Development* 31:1

BREEN, (1996). "Regression models: Censored, sample selected, or truncated data." *Sage University Papers Series on Quantitative Applications in the Social Sciences*, 07-106. Thousand Oaks, CA: Sage Publications.

BUNDY, C, 1988. *The Rise and fall of the South African peasantry*. Cape Town, South Africa: David Philip Publishers.

CARTER, M & MAY, J, 1999. Poverty, Livelihood & Class in Rural South Africa. in *World Development* 27.

COULOMBE, H. & MCKAY, A. 1993. The Causes of Poverty in Mauritania in T.A. Lloyd and W.O. Morrissey (Eds) *Poverty, Inequality and Rural Development*, London, McMillan.

DEATON, A, 1997. *The analysis of household surveys: A microeconometric approach to development policy.* Baltimore, MD: John Hopkins University Press.

DIEDEN, S., & GUSTAFSSON, B. (2003). "Child poverty in South Africa: an assessment based on microdata for 1995" in *International Journal of Social Welfare*, 12:4.

ELLIS, F, 2000. Rural Livelihoods and Diversity in Developing Countries. Oxford, UK: Oxford University Press.

GEDA, A., DE JONG, M., MWABU, G. AND KIMENYI, M.S., 2001. *Determinants of poverty in Kenya: A household level analysis.* ISS Working paper, The Hague, Netherlands.

GLEWWE, P.,1991. "Investigating the determinants of household welfare in Côite d'Ivoire"in *Journal of Development Economics* 35

HALVORSEN, R., & PALMQUIST, R. (1980). "The Interpretation of dummy variables in semi-logarithmic equations", *American Economic Review*, 70:3.

HAUSMAN, J.A., 1978. Specification Tests in Econometrics, *Econometrica* 64.

HAUSMAN, J.A., 1983."Specification and Estimation of Simultaneous Equation Models" in Z Grilliches and M.D. Intriligator (Eds.) *Handbook of Econometrics*, Vol. I, Amsterdam, North Holland.

HECKMAN, J., 1979. "Sample selection bias as a specification error", *Econometrica* 47.

JENSEN, R, 2001. "Do Private Transfers 'displace' the benefits of Public Transfers? Evidence from South Africa". *Journal of Public Economics*,1.

KINGDON, G. & KNIGHT, J. (2004), Unemployment in South Africa: The Nature of the Beast, *World Development*: 32:3.

KLASEN, S.,1997. Poverty, Inequality and Deprivation in South Africa: An Analysis of the 1993 SALDRU Survey in *Social Indicators Research*, 41.

KLASEN, S., 2000. Measuring Poverty and Deprivation in South Africa, *Review of Income and Wealth* 46:33-58

LEIBBRANDT, M & WOOLARD, I, 1999.* A comparison of poverty in South Africa's nine provinces, *Development Southern Africa,* 16:1.

LEIBBRANDT, M & WOOLARD, I, 2001. Household incomes, poverty & inequality in a multivariate framework in *Bhorat et al* , 2001.

LEIBBRANDT, M, WOOLARD, I, & BHORAT, H, 2000. Understanding Contemporary Household Inequality in South Africa in *Studies in Economics & Econometrics*, 24:3

LEIBBRANDT, M, WOOLARD, C, & WOOLARD, I, 2000. The contribution of income components to income inequality: a decomposable Gini analysis. *Journal of African Economies* 9

LESTER, A, 2000. "Historical geography". Ch. 3 in *The Geography of South Africa in a changing world.* FOX, R & ROWNTREE, K, (Eds), 2000. Cape Town, South Africa Oxford University Press Southern Africa.

LIPTON, M, DE KLERK, M & LIPTON, M (Eds), 1996. *Land, labour & livelihoods in rural South Africa.* Durban, South Africa: Indicator Press.

MAY, J, 1996. *Land, labour & rural livelihoods in KwaZulu-Natal*, Ch.2 in Lipton *et al*(1996).

MAY, J, (Ed), 2000. *Poverty & inequality in South Africa: meeting the challenge.* Cape Town, South Africa, David Philip Publishers.

MAY, J., CARTER, M., AND POSEL, D., 1995. "The composition and persistence of poverty in rural South Africa: an entitlements approach". *Land and Agriculture Policy Centre Policy Paper 15*. Land and Agriculture Policy Centre, Johannesburg, South Africa.

MØLLER, VALERIE (Ed.), 1997. Social Indicators Research 41:1 (journal issue devoted to South African quality of life).

NATTRASS, J, 1981. *The South African Economy: its growth and change*. Cape Town, South Africa: Oxford University Press.

POSEL, D, 2001. "How do households work? Migration, the household and remittance behaviour in South Africa". *Social Dynamics* 27.

REARDON, T, 1997. Using Evidence of Household Income Diversification to Inform Study of the Rural Nonfarm Labour Market in Africa World Development 25:5.

RIVERS, D & VOUNG, QH, 1988. Limited Information Estimators and Exogeneity Tests for Simulateous Probit Models. *Journal of Econometrics* 39.

SALDRU , 1994. *South Africans Rich & Poor: Baseline Household Statistics*. Cape Town, South Africa: Southern Africa labour & development research unit, University of Cape Town.

SCHULZ, T.P., 1983. Book review of "Theory of Fertility Decline" by J. Caldwell, *Population and Development Review 9*.

STARK, O, 1995. *Altruism & beyond: An Economic Analysis of Transfers and Exchanges within Families & Groups*., Cambridge, Cambridge University Press.

STATISTICS SOUTH AFRICA, 1996. *Living in South Africa. Selected findings of the 1995 October household survey.* Statistics South Africa, Pretoria.

STATISTICS SOUTH AFRICA, 1997a. *Earning and spending in South Africa Selected findings of the 1995 income & expenditure survey.* Statistics South Africa, Pretoria.

STATISTICS SOUTH AFRICA, 1997b. Documentation supplied with OHS & IES1995 raw data sets. Statistics South Africa, Pretoria.

VAN DER BERG, S, 1992. "Redressing economic inequality in South Africa" in *Redistribution and Affirmative action: Working on the South African political economy*. Hugo, P. (Ed.) Halfway House, South Africa: Southern Book Publishers.

VAN DER BERG, S, 2000. *Social policies to address poverty* in Bhorat *et al* (2001*)*.

WILSON, F. & RAMPHELE, M., 1989, *Uprooting poverty: the South African challenge.* Cape Town, South Africa: David Philip Publishers

WOOLDRIDGE, J.M., 1999. *Introductory Econometrics – A Modern Approach.* Stamford, CT; Thomson Learning.

WOOLDRIDGE, J.M., 2002. *Econometric Analysis of Cross Section and Panel Data.* Cambridge, MA: The MIT Press.

WORLD BANK, 1995. Key Indicators of Poverty in South Africa. Pretoria, South Africa: South African Communication Service.

Appendix 1

Individuals that were captured with any amount of income in the IES module were matched to the OHS data according to households' unique code, age, gender and race, by which 97.5 percent of the utilized sample of earners were matched. The sample was then increased by allowing for miscaptured data by either (i) race and gender to match perfectly but age to mismatch by up to two years or (ii) age and gender to match perfectly but race to have been mistakenly captured. These two procedures yielded in total 32 537 matched earners. Further, out of the matched earners, 871 wage earners or self-employed lacked information about the economic sector in which they were active. In addition to the above delimitations, four households without adult members and two households with indirect income exceeding R 900 000 were

dropped from the analyses. The sample delimitation process is illustrated in Table A1.1.[15]

Table A1.1: Sample delimitation process

Sample	Number of households	Share of total revisited sample	Share of African and Coloured revisited households
Total OHS/IES sample	28 585	100.0	
African and Coloured OHS/IES sample	22 366	78.2	100.0
Above sample with all earners identified in both data sets	20 572	72.0	92.0
Above sample with sector information captured for all wage earners and self-employed	19 920	69.7	89.1
Final sample	19 914	69.7	89.0
Unweighted figures			

Appendix 2

Table A2.1 shows the distribution of the number of additional, *non*-main direct income sources in the final sample and an immediate impression is that *less than 15 percent of households with a direct main income source have any other source of direct income*. Table A2.2 displays the distribution of the number of contributors to the main income among households with a direct income source. The figures in the column to the far right show that in 70 percent of the households the main income is earned by one individual and in 95 percent of households by two earners or less.

Table A2.1: Percentage of additional, regular sources of income, by main income source.

Number of additional direct sources of income	Main income source						All direct sources
	Core sectors	Mining and Quarrying	Primary sectors	Domestic services	Public transfers	Private transfers	
0	81	92	92	87	92	97	86
1	17	7	8	12	8	3	12
2	3	1	0	0	0	0	1
Sum	100	100	100	100	100	100	100
Unweighted figures. n= 15 023							

[15] Before dropping the two households with Indirect income larger than R900.000 the weighted mean of indirect income was R3709, the median R500 and the variance 485.000.000. The median value was unaffected by the exclusion of the two households but mean and variance respectively reduced to R3552 and 173.000.000

Table A2.2: Fractions of households with number of contributors to main income, by main income source category.[16]

Number of earners	Main income source category						All
	Core sector	Mining and Quarrying	Primary sectors	Domestic services	Public transfers	Private transfers	
1	65	97	70	82	68	92	71
2	28	2	24	16	29	7	24
3 or more	7	1	6	2	3	1	5
Total	100	100	100	100	100	100	100
Unweighted figures. n= 14 242							

Appendix 3

Table A3.1 shows the results from the regression based tests of endogeneity where estimates in bold are significant at 10 percent level or higher. Standard notation with three, two and and one stars (***/** / *) indicate significance at the one, five and ten percent level respectively. The two bottom rows of the table contain the F-values and implied probability that the null hypothesis applies, i.e. that the coefficient estimates for the first-stage residuals' are all equal to zero. In neither of the two approaches is the null hypothesis supported at any pertinent level. Hence, the *exogeneity of the income source variables in either format is not supported*. Tables A3.2 and A3.3 contain the results from the first-stage regression analyses from which the residuals and predicted main income sources are computed. The selection-correction procedure was applicable neither to the main income source categories Domestic services and Public transfers, nor to income fractions from Indirect income.

[16] The data structure does not allow attribution of "Indirect income" to individual household members and main income earners are not defined for the "Diversifying" category

Table A3.1: Regression based tests of endogeneity.

Dependent variable: log per adult-equivalent income		
VARIABLE	**SPECIFICATION 2**	**SPECIFICATION 3**
	Coeff. (Std.err)	**Coeff. (Std.err)**
African	**-0.216*****(0.024)	**-0.185*****(0.021)
Number of children 0-7 in household	**-0.076*****(0.006)	**-0.102*****(0.006)
Number of children 8-15 in household	**-0.071***** (0.005)	**-0.096*****(0.005)
Number of female adults in household	-0.002 (0.009)	**-0.049*****(0.007)
Number of male adults in household	0.005 (0.009)	**-0.038*****(0.007)
Number of elderly in household	**0.122*****(0.022)	-0.009 (0.021)
Share of adults with primary education	0.000 (0.000)	0.000 (0.000)
Share of adults with secondary education	**0.002*****(0.000)	**0.002*****(0.000)
Share of adults with matriculation	**0.006*****(0.000)	**0.006*****(0.000)
Share of adults with tertiary education	**0.012*****(0.001)	**0.012*****(0.000)
Share of adults unemployed	**-0.002*****(0.000)	**-0.002*****(0.000)
Share of adult non-participants	**-0.009*****(0.000)	**-0.008*****(0.000)
Rural location	**-0.080*****(0.024)	**-0.082*****(0.023)
Diversifying (No main income source)	**-1.950*****(0.168)	
Mining and Quarrying sectors income	**0.552*****(0.087)	**0.005*** (0.001)
Primary sectors income	**-0.638*****(0.104)	**-0.012*****(0.001)
Domestic services income	**-1.111*****(0.277)	**-0.016*****(0.002)
Public transfers income	**-0.576*****(0.057)	**-0.008*****(0.001)
Private transfers income	-0.063 (0.081)	-0.002 (0.001)
Indirect income	0.237 (0.521)	**-0.011*****(0.002)
1st stage residual Diversifying (No main income source)	**1.644*****(0.172)	
1st stage residual Mining and Quarrying sectors income	**-0.428*****(0.095)	**-0.004*****(0.001)
1st stage residual Primary sectors	0.044 (0.106)	**0.003*****(0.001)
1st stage residual Domestic services	0.366 (0.274)	**0.006*****(0.002)
1st stage residual Public transfers	**-0.185*****(0.059)	-0.000 (0.001)
1st stage residual Private transfers	**-0.546*****(0.092)	**-0.004*****(0.001)
1st stage residual Indirect income	-0.145 (0.509)	**0.012*****(0.002)
Intercept	**9.764*****(0.051)	**9.795*****(0.051)
F-test	F-value = 37.12	F-value =21.84
Prob > F	0.000	0.000

Significance levels are based on standard errors that take the clustered and stratified nature of the sample into account. Estimates in **bold** are significant. Symbols: ***/** / * Estimate significant at 5%/ 10% level. Weighted data, n= 19 914.

Table A3.2a: First-stage regressions for main income source dummy variables; Diversifying households, Mining and Quarrying Sectors, Primary sectors, Domestic services.

F – values	38.55	9.44		46.63		34.50	
Probability H0 true	0.0000	0.0000		0.0000		0.0000	
	Diversifying	**Mining and Quarrying**		**Primary sectors**		**Domestic services**	
		Outcome	**Selection**	**Outcome**	**Selection**	**Outcome**	**Selection**
African	**-0.096**** (0.048)	0.087 (0.269)	**0.291***(0.110)	-0.037 (0.125)	**-0.347***(0.099)	0.030 (0.148)	**0.219***(0.062)
Number of children 0-7 in household	**0.035***(0.012)	-0.065 (0.062)	**-0.097***(0.031)	**-0.071***(0.027)	**-0.032*** (0.019)	**-0.124***(0.044)	**0.072***(0.017)
Number of children 8-15 in household	**0.031***(0.011)	-0.089 (0.062)	**-0.123***(0.028)	**-0.086***(0.027)	**-0.089***(0.018)	**-0.091**** (0.039)	**0.112***(0.016)
Number of female adults in household	**0.145***(0.012)	**-0.370***(0.101)	**0.225***(0.054)	**-0.083*** (0.042)	**0.203***(0.022)	**-0.375***(0.097)	**0.283***(0.020)
Number of male adults in household	**0.080***(0.012)	**-0.428***(0.077)	0.013 (0.032)	-0.026 (0.046)	**0.213***(0.020)	**-0.516***(0.052)	**0.227***(0.019)
Number of elderly in household	**0.190***(0.021)	**-0.283*** (0.167)	-0.152 (0.073)	**-0.846***(0.077)	**-0.540***(0.042)	**-0.646***(0.144)	**-0.268***(0.034)
Share of adults with primary education	-0.001 (0.000)	0.000 (0.003)	**0.002*** (0.001)	**0.002**** (0.001)	**-0.002***(0.001)	0.001 (0.001)	0.000 (0.000)
Share of adults with secondary education	**-0.004***(0.000)	-0.003 (0.003)	**0.004***(0.001)	0.001 (0.002)	**-0.012***(0.001)	0.002 (0.002)	**-0.004***(0.001)
Share of adults with matriculation	**-0.007***(0.001)	-0.001 (0.003)	0.001 (0.001)	-0.005 (0.003)	**-0.020***(0.002)	0.005 (0.003)	**-0.011***(0.001)
Share of adults with tertiary education	**-0.010***(0.001)	0.002 (0.003)	-0.002 (0.002)	**-0.024*** (0.010)	**-0.032***(0.003)	0.008 (0.007)	**-0.024***(0.002)
Share of adults unemployed	0.000 (0.000)	0.003 (0.004)	**-0.004***(0.001)	**-0.003*** (0.002)	**-0.002***(0.001)	**0.005***(0.002)	0.001 (0.001)
Share of adult non-participants	**-0.004***(0.000)	**0.023***(0.004)	**-0.006***(0.001)	-0.002 (0.002)	**-0.017***(0.001)	**0.018***(0.004)	**-0.019***(0.001)
Rural location	**0.318***(0.034)	**0.277*** (0.156)	**-0.193*** (0.108)	**0.601***(0.210)	**1.478***(0.092)	**-0.208**** (0.081)	0.002 (0.042)
W Cape	-0.080 (0.065)	1.201 (0.494)	-0.201 (0.214)	**0.441**** (0.179)	**0.556***(0.139)	-0.222 (0.186)	0.010 (0.085)
E Cape	**-0.171***(0.047)	0.430 (0.380)	**0.074***(0.160)	-0.007 (0.152)	0.022 (0.093)	**-0.212*** (0.128)	0.045 (0.062)
N Cape	0.038 (0.078)	0.970 (0.519)	**0.640***(0.216)	**0.495***(0.172)	**0.679***(0.135)	-0.281 (0.184)	**0.291***(0.086)
Free State	**0.332***(0.057)	0.100 (0.433)	**0.812***(0.166)	**-0.542***(0.173)	**0.736***(0.114)	**-0.604***(0.138)	**0.519***(0.064)
North-West Province	0.048 (0.066)	0.331 (0.428)	**0.865***(0.155)	-0.203 (0.213)	**0.323***(0.120)	**-0.663***(0.165)	-0.006 (0.078)
Gauteng	**-0.193***(0.064)	0.049 (0.462)	**0.741***(0.179)	0.168 (0.193)	**-0.291**** (0.145)	-0.075 (0.140)	0.025 (0.082)
Mpumalanga	**-0.238***(0.056)	0.212 (0.393)	**0.881***(0.171)	**0.802***(0.141)	**0.539***(0.119)	0.209 (0.142)	**0.131*** (0.072)
Limpopo	**-0.230***(0.063)	0.582 (0.439)	**0.374*** (0.180)	0.229 (0.197)	-0.065 (0.116)	-0.119 (0.200)	**-0.184**** (0.088)
Share of females in the household			**-0.027***(0.004)		**-0.004***(0.001)		**0.014***(0.001)
Migrant head			**0.718***(0.097)		**-0.401***(0.085)		**-0.156**** (0.073)
Intercept	**-0.721***(0.080)	**2.092***(0.574)	**-1.762***(0.235)	**-1.133***(0.332)	**-0.802***(0.167)	**0.234***(0.4016)	**-1.761***(0.105)
Athrho			**-1.081***(0.171)		**0.751***(0.246)		-0.096 (0.242)
Rho			-0.793 (0.064)		0.636 (0.146)		-0.096 (0.240)

Significance levels are based on standard errors that take the clustered and stratified nature of the sample into account.
Estimates in **bold** are significantl. Symbols: ***/** / * Estimate significant at 1%/ 5%/ 10% level. Weighted data.

Table A3.2b: First-stage regressions for main income source dummy variables; Public transfers, Private transfers, Indirect income.

F – values	84.93		48.44		26.36	
Probability H0 true	0.0000		0.0000		0.0000	
	Public transfers		**Private transfers**		**Indirect income**	
	Outcome	**Selection**	**Outcome**	**Selection**	**Outcome**	**Selection**
African	0.093 (0.080)	**-0.320*****(0.056)	**0.357*** (0.148)	**0.244*****(0.072)	**0.234**** (0.115)	0.090 (0.058)
Number of children 0-7 in household	**-0.095*****(0.020)	0.023 (0.015)	**0.004**** (0.026)	**0.072*****(0.014)	-0.027 (0.018)	0.005 (0.013)
Number of children 8-15 in household	**-0.087*****(0.017)	**0.031**** (0.014)	**0.065**** (0.024)	**0.075*****(0.014)	0.022 (0.019)	-0.017 (0.012)
Number of female adults in household	**-0.247*****(0.020)	**-0.043**** (0.018)	**-0.226*****(0.029)	**-0.142*****(0.019)	**0.058**** (0.021)	**0.076*****(0.016)
Number of male adults in household	**-0.244*****(0.019)	**0.030*** (0.016)	**-0.273*****(0.034)	**-0.276*****(0.019)	-0.071 (0.020)	0.020 (0.014)
Number of elderly in household	0.082 (0.080)	**1.739*****(0.045)	**-1.321*****(0.089)	**-0.292*****(0.031)	-0.207 (0.045)	**-0.180*****(0.020)
Share of adults with primary education	**-0.002*****(0.001)	**-0.003*****(0.000)	-0.000 (0.001)	**0.001*** (0.001)	**-0.002*****(0.001)	**-0.002*****(0.000)
Share of adults with secondary education	**-0.007*****(0.001)	**-0.007*****(0.001)	0.000 (0.001)	**0.002*****(0.001)	**-0.005*****(0.001)	**-0.005*****(0.001)
Share of adults with matriculation	**-0.009*****(0.002)	**-0.007*****(0.001)	0.002 (0.002)	-0.001 (0.001)	**-0.004*****(0.001)	**-0.005*****(0.001)
Share of adults with tertiary education	**-0.012*****(0.002)	**-0.008*****(0.001)	-0.006 (0.004)	-0.000 (0.001)	**-0.004*****(0.002)	**-0.004*****(0.001)
Share of adults unemployed	0.000 (0.001)	**-0.003*****(0.000)	**0.003*****(0.001)	**0.003*****(0.000)	**0.002*****(0.001)	-0.000 (0.001)
Share of adult non-participants	**0.013*****(0.001)	**0.017*****(0.001)	**0.026*****(0.002)	**0.017*****(0.001)	**0.003*****(0.001)	**-0.005*****(0.000)
Rural location	**0.106*****(0.053)	**-0.251*****(0.034)	**0.281*****(0.076)	0.041 (0.040)	0.026 (0.066)	**0.745*****(0.042)
W Cape	0.099 (0.107)	-0.038 (0.076)	-0.329 (0.220)	-0.019 (0.094)	-0.155 (0.200)	**-0.282*****(0.086)
E Cape	**0.456*****(0.068)	0.066 (0.049)	**0.435*****(0.093)	**0.115**** (0.054)	-0.043 (0.091)	**-0.319*****(0.062)
N Cape	**0.205*** (0.121)	**-0.177**** (0.085)	0.150 (0.176)	-0.088 (0.105)	**0.263*****(0.136)	-0.010 (0.095)
Free State	**0.333** (0.084)	**-0.320*****(0.060)	**0.247*** (0.131)	**-0.154**** (0.068)	0.360 (0.093)	-0.117 (0.073)
North-West Province	0.061 (0.097)	**-0.230*****(0.066)	**0.283**** (0.119)	0.030 (0.075)	0.022 (0.109)	**-0.173**** (0.078)
Gauteng	0.108 (0.117)	**-0.528*****(0.079)	-0.248 (0.287)	**-0.601*****(0.092)	0.057 (0.137)	0.031 (0.080)
Mpumalanga	0.156 (0.097)	**-0.401*****(0.061)	0.052 (0.130)	**-0.315*****(0.094)	-0.541 (0.142)	**-0.310*****(0.075)
Limpopo	**0.259*****(0.094)	**-0.114*****(0.066)	**0.281**** (0.119)	0.006 (0.071)	0.290 (0.102)	**-0.221*****(0.079)
Share of females in the household		**0.006*****(0.001)		0.002 (0.001)		**-0.002*****(0.001)
Migrant head		**-0.481*****(0.080)		**-0.179*****(0.062)		**0.229*****(0.062)
Intercept	-0.218 (0.174)	**-1.288*****(0.094)	**-2.724*****(0.337)	**-1.945*****(0.120)	**-1.943*****(0.173)	**0.707*****(0.100)
Athrho		-0.048 (0.112)		**0.207**** (0.091)		**1.400*****(0.214)
Rho		-0.048 (0.112)		0.204 (0.087)		0.885 (0.046)

Significance levels are based on standard errors that take the clustered and stratified nature of the sample into account.
Estimates in **bold** are significant at the 1% level.
Symbols: ** / * Estimate significant at 5%/ 10% level.
Weighted data.

Table A3.3a: First-stage regressions for fractions of total income from income sources; Mining and Quarrying Sectors, Primary sectors, Domestic services.

F – values	13.58		52.69		44.27	
Probability H0 true	0.0000		0.0000		0.0000	
	Mining and Quarrying		**Primary sectors**		**Domestic services**	
	Outcome	**Selection**	**Outcome**	**Selection**	**Outcome**	**Selection**
African	2.598 (5.623)	**0.321*****(0.111)	1.925 (2.550)	**-0.338*****(0.099)	1.906 (2.278)	0.219 (0.062)
Number of children 0-7 in household	-1.226 (1.195)	**-0.094*****(0.032)	**-1.769*****(0.488)	**-0.032**** (0.019)	**-1.924*****(0.545)	**0.073*****(0.016)
Number of children 8-15 in household	-0.441 (1.019)	**-0.123*****(0.028)	**-1.648*****(0.536)	**-0.092*****(0.018)	**-1.391*****(0.509)	**0.113*****(0.016)
Number of female adults in household	**-6.521***** (1.913)	**0.204*****(0.068)	**-2.618*****(0.760)	**0.205*****(0.022)	**-7.048*****(0.871)	**0.281*****(0.020)
Number of male adults in household	**-10.037***** (1.551)	0.026 (0.036)	**-2.325*****(0.741)	**0.211*****(0.020)	**-9.338*****(0.623)	**0.228*****(0.019)
Number of elderly in household	**-6.402**** (3.144)	**-0.171**** (0.078)	**-14.137*****(1.369)	**-0.541*****(0.042)	**-6.966*****(1.306)	**-0.270*****(0.034)
Share of adults with primary education	-0.010 (0.031)	**0.002*** (0.001)	**0.045*****(0.016)	**-0.002*****(0.001)	**0.053*****(0.019)	0.000 (0.000)
Share of adults with secondary education	-0.053 (0.042)	**0.004*****(0.001)	**0.096*****(0.027)	**-0.012*****(0.001)	**0.091*****(0.026)	**-0.003*****(0.001)
Share of adults with matriculation	-0.005 (0.036)	0.001 (0.001)	0.024 (0.047)	**-0.0120*****(0.002)	**0.076*****(0.044)	**-0.011*****(0.001)
Share of adults with tertiary education	0.072 (0.053)	-0.003 (0.002)	-0.201 (0.151)	**-0.0320*****(0.003)	-0.023 (0.123)	**-0.024*****(0.002)
Share of adults unemployed	0.061 (0.058)	**-0.004*****(0.001)	**-0.056*** (0.032)	**-0.002*****(0.001)	**0.103*****(0.037)	0.001 (0.001)
Share of adult non-participants	**0.605***** (0.073)	**-0.006*****(0.001)	**0.097*****(0.031)	**-0.017*****(0.001)	**0.348*****(0.038)	**-0.019*****(0.001)
Rural location	4.652 (3.496)	**-0.193*** (0.107)	**6.696*** (3.159)	**1.473*****(0.092)	**-5.372*****(1.312)	0.002 (0.042)
W Cape	**30.200*****(11.284)	-0.203 (0.216)	**8.842**** (3.530)	**0.556*****(0.140)	-4.677 (3.020)	0.011 (0.085)
E Cape	**11.122***** (7.968)	0.043 (0.166)	-2.843 (3.418)	0.022 (0.093)	**-5.731*****(2.224)	0.045 (0.062)
N Cape	**16.283*****(10.665)	**0.592*****(0.218)	**6.184*** (3.312)	**0.677*****(0.136)	**-7.224*****(2.982)	**0.292*****(0.086)
Free State	**3.094***** (9.603)	**0.743*****(0.176)	**-11.938*****(2.589)	**0.736*****(0.115)	**-15.133*****(1.897)	**0.520*****(0.064)
North-West Province	**4.240***** (9.280)	**0.824*****(0.168)	**-6.869*** (3.651)	**0.328*****(0.121)	**-10.991*****(2.184)	-0.005 (0.078)
Gauteng	**4.441***** (9.842)	**0.702*****(0.189)	4.593 (3.201)	**-0.294*** (0.145)	-0.277 (2.498)	0.026 (0.082)
Mpumalanga	**3.168***** (9.672)	**0.847*****(0.174)	**11.413*****(2.951)	**0.543*****(0.120)	**5.400**** (2.440)	**0.132*** (0.072)
Limpopo	**12.422***** (9.269)	**0.355**** (0.180)	3.599 (3.875)	-0.064 (0.117)	**5.650*** (3.086)	**-0.184**** (0.088)
Share of females in the household		**-0.026*****(0.005)		**-0.005*****(0.001)		**0.014*****(0.001)
Migrant head		**0.546*****(0.171)		**-0.418*****(0.085)		**-0.148**** (0.074)
Intercept	**108.878***** (17.919)	**-1.726*****(0.238)	**51.674*****(5.366)	**-0.796*****(0.168)	**60.481*****(4.037)	**-1.768*****(0.106)
Athrho		**-1.283*****(0.431)		**0.196*****(0.074)		**-0.151**** (0.061)
Lnsigma		**3.140*****(0.174)		**3.117*****(0.019)		**3.167*****(0.018)
Rho		-0.857 (0.114)		0.194 (0.071)		-0.149 (0.060)
Sigma		23.102 (4.034)		22.572 (0.428)		23.742 (0.431)
Lambda		-19.804 (6.006)		4.3704 (1.645)		-3.547 (1.440)

Significance levels are based on standard errors that take the clustered and stratified nature of the sample into account. Estimates in **bold** are significant at the 1% level.
Symbols: ** / * Estimate individually significant at 5%/ 10% level. Weighted data.

Table A3.3b: First-stage regressions for fractions of total income from income sources; Public transfers, Private transfers, Indirect income.

F – values	119.98		77.20		27.53
Probability H0 true	0.0000		0.0000		0.000
	Public transfers		**Private transfers**		**Indirect income**
	Outcome	**Selection**	**Outcome**	**Selection**	**WLS**
African	2.376 (1.754)	**-0.320*****(0.056)	3.416 (2.890)	**0.245*****(0.072)	0.340 (0.800)
Number of children 0-7 in household	**-2.030*****(0.365)	0.024 (0.015)	0.138 (0.480)	**0.071*****(0.014)	0.064 (0.178)
Number of children 8-15 in household	**-1.689*****(0.346)	**0.031**** (0.014)	**1.704*****(0.455)	**0.074*****(0.014)	-0.109 (0.168)
Number of female adults in household	**-5.824*****(0.390)	**-0.044**** (0.018)	**-5.834*****(0.514)	**-0.141*****(0.018)	**0.584*****(0.201)
Number of male adults in household	**-5.822*****(0.349)	**0.030*** (0.016)	**-8.793*****(0.641)	**-0.277*****(0.019)	**-0.403**** (0.179)
Number of elderly in household	**2.504*****(1.008)	**1.736*****(0.045)	**-22.65*****(0.946)	**-0.300*****(0.030)	**-4.265*****(0.331)
Share of adults with primary education	**-0.054*****(0.014)	**-0.003*****(0.000)	0.005 (0.023)	**0.001**** (0.001)	**-0.030*****(0.008)
Share of adults with secondary education	**-0.186*****(0.019)	**-0.007*****(0.001)	0.036 (0.025)	**0.002*****(0.001)	**-0.084*****(0.010)
Share of adults with matriculation	**-0.268*****(0.029)	**-0.007*****(0.001)	0.020 (0.032)	-0.001 (0.001)	**-0.087*****(0.011)
Share of adults with tertiary education	**-0.410*****(0.050)	**-0.008*****(0.001)	**-0.159**** (0.064)	-0.000 (0.001)	**-0.088*****(0.017)
Share of adults unemployed	0.010 (0.013)	**-0.003*****(0.000)	**0.087*****(0.017)	**0.003*****(0.000)	**0.014*****(0.008)
Share of adult non-participants	**0.336*****(0.015)	**0.017*****(0.001)	**0.575*****(0.028)	**0.017*****(0.001)	**-0.028*****(0.009)
Rural location	**2.683**** (1.082)	**-0.250*****(0.034)	**8.371*****(1.593)	0.040 (0.040)	**6.384*****(0.613)
W Cape	2.732 (2.284)	-0.038 (0.076)	**-6.622*** (3.990)	-0.022 (0.094)	**-2.811**** (1.253)
E Cape	**11.734*****(1.368)	0.066 (0.049)	**10.192*****(2.020)	**0.116**** (0.054)	**-4.083*****(0.898)
N Cape	**8.612*****(2.533)	**-0.177**** (0.085)	3.500 (4.056)	-0.087 (0.105)	1.635 (1.430)
Free State	**8.745*****(1.718)	**-0.320*****(0.059)	3.710 (2.666)	**-0.153**** (0.068)	**7.425*****(1.160)
North-West Province	2.463 (1.912)	**-0.230*****(0.066)	2.289 (2.521)	0.029 (0.075)	1.744 (1.206)
Gauteng	-0.245 (2.672)	**-0.528*****(0.079)	**-10.628**** (4.408)	**-0.603*****(0.092)	0.290 (1.170)
Mpumalanga	**5.006**** (1.981)	**-0.401*****(0.061)	-1.359 (2.651)	**-0.315*****(0.094)	**-4.873*****(0.937)
Limpopo	**6.290*****(2.125)	**-0.114*** (0.066)	4.0287 (2.697)	0.009 (0.071)	0.459 (1.316)
Share of females in the household		**0.006*****(0.001)		0.002 (0.001)	
Migrant head		**-0.484*****(0.080)		**-0.174*****(0.063)	
Intercept	**56.580*****(2.925)	**-1.292*****(0.094)	-1.008 (5.480)	**-1.945*****(0.119)	**18.221*****(1.397)
Athrho		**-0.089*** (0.048)		**0.415*****(0.056)	
Lnsigma		**3.304*****(0.009)		**3.390*****(0.020)	
Rho		-0.089 (0.048)		0.393 (0.047)	
Sigma		27.212 (0.246)		29.667 (0.598)	
Lambda		-2.411 (1.307)		11.650 (1.598)	

Significance levels are based on standard errors that take the clustered and stratified nature of the sample into account. Estimates in **bold** are significant at the 1% level. Symbols: ** / * Estimate individually significant at 5%/ 10% level. Weighted data.

Appendix 4

Table A4.1 contains predicted partial estimates for the shares of income derived from the various categories in Specification 3. For the predictions it is assumed that otherwise identical households in each Main income category derive the average share of income from the Main income sources in that category (depicted in Table10). For instance, households in the Mining and Quarrying Main income source category on average derive 90 percent of their income from the Mining and Quarrying sectors, while households with their main income from the Primary sectors category on average derive 84 percent from *their* main income source.

Table A4.1: Differences in observed mean income, predicted partial percentage effect on income for the different Main and sole complementary income source categories as compared to Core sectors main income

Main income source	Average fraction from main source (%)	Difference in mean income (%)	Sole complementary income source category						
			Core sectors	M & Q sectors	Primary sectors	Domestic services	Public transfers	Private transfers	Indirect income
Core	92	0	0	6	-9	-12	-3	-2	-8
M & Q sectors	90	13	**108**	**114**	**87**	**84**	**94**	**95**	**88**
Primary sectors	84	-65	**-45**	**-48**	**-67**	**-82**	**-68**	**-66**	**-77**
Domestic services	86	-65	**-60**	**-62**	**-87**	**-78**	**-79**	**-77**	**-87**
Public transfers	93	-76	**-26**	**-27**	**-40**	**-43**	**-35**	**-35**	**-40**
Private transfers	94	-75	**-20**	**-21**	**-32**	**-34**	**-28**	**-27**	**-31**
Indirect income	77	-11	**-34**	**-38**	**-79**	**-86**	**-66**	**-64**	**-66**

As can be seen, the predictions suggest that income levels are considerably affected by the source of complementary income. For all main income sources, the impact on income is higher (or less negative) when the complimentary income source is from either the Core or the Mining and Quarrying category. For the current purposes it suffices to notice that the following observations can be made:

(i) In the Mining and Quarrying sectors category the *absolute* magnitude of the joint impacts are higher than the observed difference in mean income, irrespective of which complementary income sources is utilized.

(ii) In both transfer categories and in the Indirect income the *absolute* magnitude of the joint impacts are higher than the observed differences in mean incomes irrespective of which complementary income sources is utilized.

(iii) In the Primary sectors and Domestic services categories the *absolute* magnitude of the joint impacts are lower than the observed differences in mean incomes, only when the complementary income source is either from the Core or the Mining and Quarrying sectors category. The net impacts from the latter two sets of combinations are very similar to the dummy variable impacts based on the dummies specification in to Table 12.

Hence, the implications with respect to the regularities in endowments of other characteristics would be largely in congruence with those discussed in connection with the dummies specification.

Chapter III:

Integration into the South African core economy – household level covariates

Sten Dieden*

Abstract

This paper intends to further improve the understanding of income generation among the formerly underprivileged households in South Africa. The study utilizes household survey data in order to analyse the households' integration into the core economy. The emerging picture of household income generation disputes common perceptions of the multitude of means by which African households are assumed to generate their income. The majority of households rely to a large extent on one income source and one income earner. Contextual information and descriptive statistics justify the estimation of separate multinomial logit models for urban and non-urban households, which return probabilities for a household having a main income source from either one of five main income source categories. Results from the regression analyses indicate that prominent covariates of low core economy integration are earners of income with low levels of education, of female gender, and of either old or young working-age. A non-urban household's location in either a former "homeland" or in an agriculturally or commercially developed area also yields disparate implications for the main income source probabilities. The results from the study suggest associations between main income sources and households' demographic compositions which are compatible with findings in previous research on endogenous household formation in South Africa.

* Department of Economics, University of Gothenburg, PO Box 640, SE 405 GOTHENBURG. Sweden. Ph. +46 31 773 41 98 Fax 0946 31 773 2648. sten.dieden@handels.gu.se.

While any defects or shortcomings in this work are entirely my own responsibility, I would like to thank Arne Bigsten, Lennart Flood, Björn Gustafsson, Murray Leibbrandt, Laura Poswell, Donald Storrie, and Ali Tasiran for most valuable comments to previous versions of this work. The financial provision for this research by the Swedish International Development Cooperation Agency (Sida) is thankfully acknowledged. Special thanks also to the Centre for Social Science Research at the University of Cape Town for hosting me while I conducted the research for this work.

1. Introduction

This paper analyses variation in South African households' income sources as a phenomenon associated with households' location, income earner characteristics and demographic composition. Two findings that often recur in research on poverty and inequality in South Africa illustrate the relevance of income sources to household welfare in the country. Firstly, it is widely recognised that, compared to less destitute households, poor households derive larger shares of their income from transfer incomes (Van der Berg (2000)). Secondly, a powerful catalyst in avoiding poverty and a crucial explanatory factor in income inequality is households' access to *wage* income (Carter and May (1999), Leibbrandt, Woolard, and Woolard (2000), Leibbrandt and Woolard (2001a), Leibbrandt, Woolard, and Bhorat (2000), Van der Berg (2000)).

Perhaps the most common route to the microeconomic analysis of household income generation in less developed countries is through some version of the "Agricultural household model" (Nakajima (1970), Singh, Squire and Strauss (1986)). In the most basic version of the model, prices and household endowments of land and labour enter as givens. The model's output encompasses *inter alia* the household's optimal allocation of its labour-time between farming activities and freely available off-farm employment opportunities. This approach is however not applicable to South Africa due to the country's legacies of racially discriminatory policies. Through the 1913 *Natives Land Act* the African population group was denied rights of landownership outside of "reserves". In conjunction with the former act, a battery of apartheid era regulations reinforced spatially uneven economic development and led to institutionalised labour migration in the African population. Consequently, peasant agriculture in the rural non-White population has also become virtually absent. (Wilson and Ramphele (1989), Bundy (1988)).

As places of work and permissible permanent residence for the African population were often separated by large distances, many African rural dwellers were also barred from labour market participation (Klasen (2000)). In addition,

unemployment rates, especially among women, are very high and increasing in rural South Africa (Kingdon and Knight (1999), Casale (2004)). Hence, the agricultural household model's assumptions regarding unconstrained allocations of labour endowments into a variety of activities do not adequately match the conditions of current South Africa.[1]

This paper does not seek to articulate a complete model for the allocation of income sources in the South African setting, rather the ambition is to identify characteristics of households that can explain variation in their modes of income generation. In a country where dependence on, for example, transfer or wage income appears to vary with income levels (Leibbrandt, Woolard, and Bhorat (2000), Van der Berg (2000)), the identification of the aforementioned characteristics would facilitate the targeting of public transfers for poverty alleviation, as well as provide empirical guidance for the development of theories that attempt to explain why some households are constrained in income generation.

While the study is not entirely dissimilar in scope from other investigations (May *et al* (1995), Lipton, de Klerk, and Lipton (1996), Carter and May (1999), Posel, (2001), Leibbrandt and Woolard (2001b)), it augments previous research in several ways. Firstly, the quantitative analyses draw on the finding that a considerable fraction of households derive the bulk of their income from one income source. The historical legacies referred to previously also justify this study's core-periphery approach to household income generation, in which households' income sources are classified according to their association with the South African core economy. Secondly, the concentration and classification of income sources warrants the analysis of households' allocation into income source categories, rather than the more common investigation of households' average shares of income from various origins (see Ellis (2000)).

[1] Versions of the agricultural household model that incorporate circumstances specific to agricultural households in Southern Africa have been developed by Low (1986).

The final means by which this analysis contributes to the literature on South Africa is through the applied methodology. The allocation to income source categories is analysed through the estimation of the probabilities of households holding main income sources from specific categories, as associated with a group of household level explanatory variables. The probability models are estimated separately for households in rural and urban areas through two multinomial logit regression frameworks.

The paper proceeds in the following manner. The next section provides information about the context in which household characteristics would have assumed their impact on income source variation in South Africa. Section 3 explains the classification of households' income source categories. There exist several historical reasons to expect households' access to wage income or dependence on transfer incomes to be related to microeconomic factors. Section 4 therefore discusses explanatory variables and analytical considerations suggested by previous research on South African households' access to income sources. Section 5 introduces the data from Statistics South Africa's 1995 October Household Survey and explains how this study's sample is constructed based on the applied definition of a main income source. In Section 6 two issues are discussed based on descriptive analyses. Firstly, an informal assessment of the extent to which main income source concept is representative households' income generation is provided. Thereafter the concept is linked to household income levels and the labour market. Section 7 introduces the empirical model and the explanatory variables. The results from the regression analyses are presented in Section 8. Finally, conclusions are drawn in Section 9.

2. Contextual information and income sources

Two historically entrenched features of income generation among African and coloured households in South Africa are divergent from what is common on the rest of the continent. These features are the generally very small contributions to household income from small-scale agriculture and the widespread dependence on transfer incomes among both rural and urban families (Reardon (1997), Jooma (1991)). This section introduces briefly, the historical origin of the complex

interlinkages between institutional legacies and households' current geographical locations, their assets and their demographic endowments, all of which would relate strongly to modes of income generation.

The migration labour system and land policy

Income generation among large parts of South Africa's non-white population cannot be explained outside of an historical context. The racial segregation, dispossession of land rights, and forced removals were all cementing factors in what came to constitute the "migrant labour system" under the apartheid era (Nattrass (1981)), Wilson and Ramphele (1989)). At the heart of the system was a predominating 'closed-compound system', the roots of which extend back to the vast mineral discoveries in the 1860s. By this arrangement mine workers were required to live in closed and guarded barracks on the mining premises, without their families and with few opportunities for leave. Similar practices soon spread to other migrant-receiving sectors and as a consequence, cash remittances from migrant workers have become an historically entrenched and important source of income for rural African families (Jooma (1991)).

The components of the migrant labour system were however complemented by a battery of laws that further inhibited the landownership and settlement rights of Africans. In consequence of the *1913 Natives Land Act*, the bulk of South African land was reserved for white ownership only. By the same act, Africans were denied rights of residence except during work contracts outside designated "reserves". The latter amounted to 13 percent of the total land area and were the only areas where Africans were allowed to farm their own land. Massive forced relocation of Africans took place to these mostly non-developed "reserves", where initially agricultural conditions were often absent or soon deteriorated due to high population densities (Wilson and Ramphele (1989)).[2]

[2] In the early twentieth century commercial forms of labour tenancy and sharecropping still prevailed in some instances. However, in the course of the first half of the century, commercialisation of white farming and increasing land segregation led to the demise of these practices (Lester (2000).

As time passed, the exclusionary land practices became an all-encompassing system in both rural and urban areas. When the apartheid programme was instituted by the *Afrikaner National Party* in 1948, "influx control" into the urban areas of "white" South Africa became even tighter and from the 1960s, Africans were officially considered citizens of the "reserves", by that time relabelled "tribal areas", "homelands" or "Bantustans". The general economic and environmental degradation of the former "homelands" ensured that households there became even more dependent on remittances and continued to send members to provide cheap labour for the major employers elsewhere in South Africa (Lester (2000), Bundy (1988)).

The migration of many of the working-age men from the "tribal areas" also resulted in a "peculiar (and quite unnatural) household structure", where children, the elderly and women were vastly overrepresented.[3] Interlinked with many of the migrants not only earning but also *spending* most of their income in the economy's core areas – or on goods produced there – the process was one of increasing spatially uneven economic development. The distribution of employment opportunities thus became increasingly more inequitable to the disadvantage of the "reserves "(Wilson and Ramphele (1989)).

The migration regulations were, however, abolished in 1986. In light of the abolishment, the persistence of oscillating migration and sustained residence by large fractions of the African population group in the formerly designated areas appears puzzling. A variety of explanations have however been offered, among which one finds lack of employment opportunities in rural areas, high costs of relocation, poor access to the urban labour markets, and shortage of housing in the "black" urban residential areas (Murray (1987), Jooma (1991)).

[3] Wilson and Ramphele (1989) refer to a study of a migration-wise not very intensive area in KwaZulu-Natal where 81 percent of the residents aged 20-50 were women.

Economic growth, labour market performance, and earnings towards the end of the apartheid era

Coupled with the migration and settlement regulations, a battery of laws under apartheid also undermined the African and coloured population's access to education, ability to increase their wages, and their upward mobility in the labour market.[4] These obstacles rendered the population group confined to poorly paid, low-skilled employment. However, the same practice also created bottlenecks in the labour market (Lundahl and Moritz (1996), Bhorat *et al* (2001))

The first oil shock in 1973 marked the beginning of a period of economic decline, from which the South African economy is still trying to recover. Among the internal factors that contributed to this downturn was the aforementioned barring of African labour from all but unskilled or low-skilled occupations and from higher education. Following a considerable mechanisation and expansion of the manufacturing sector after WWII, the lack of educated labour eventually became a severe constraint to economic growth in the early 1970s. Subsequent to a series of strikes at the same point of time, an official recognition of African labour rights and improved access to education came to signify an important shift in economic power, in the wake of which also followed improvements in the wages for African workers (Lundahl and Moritz (1996), Bhorat *et al* (2001)).

In addition to these internal factors must be recognized the severe constraint on petroleum imports posed by Iran's embargo as of 1979. Prior to this embargo the international reaction to apartheid had been growing since the 1960s, especially on the African continent. (The pressure from abroad increased up to a point in the late 1980s when virtually all Western countries to varying extents had put sanctions of South African products into practice.) In the 1970s the world

[4] Two factors that contributed to the rising of a race barrier in the South African labour market were, firstly, the governments' provided provision of military troops for the protection of the mine-owners interests and, secondly, the victory in the 1924 elections of the (white) farmers and workers coalition. In both of the latter groups' interest rested poor labour rights of Africans. During and after the second world access to education for the non-white population had progressed, but also an , but also an educational race barrier was raised by the 1953 Bantu Education Act and the 1956 Extension of University Education Act. Through the former the education facilities available to Africans were pegged to the population groups' contribution to tax income and through the latter the same population group was banned from white universities (Lundahl and Mortitz (1996)).

economy was in a slump and the international resistance to apartheid damaged South African exports, which had adverse effects on the country's foreign savings. These latter circumstances became even more detrimental by a considerable capital flight taking place in the same era, possibly due to the growing political turbulence in many parts of Southern Africa. A general uncertainty also prevailed with respect to South Africa's own political and economic future, especially after the 1976 Soweto protest (Lundahl and Moritz (1996)).

Together all these factors led to a slow-down in capital accumulation and technological development. Economic growth rates fell below population growth rates and consequently the *per capita* income declined by 15 percent from 1974 to 1993. Due to the economic stagnation, unemployment rates increased and were further augmented by distorted relative costs of (often subsided) capital and labour that encouraged further mechanisation. The consequences of the mechanisation were particularly grave for rural African wage employment in agriculture which dropped by approximately one million jobs between 1971 and 1993 (Bhorat, Hodge and Dieden (1998), Bhorat *et al* (2001))

Apartheid to a large extent also barred the non-white population from opportunities to raise non- wage income from land, capital and entrepreneurship. Accordingly, analyses of microdata from the early 1990s attest to high poverty and unemployment rates as well as widespread dependence on transfer incomes in both urban and rural areas (SALDRU (1994*),* World Bank (1995)).While both private and public transfers were and are common, one need consider in this context, that private remittances are directly related to economic growth and public transfers depend on governments' position on redistribution (Bhorat *et al* (2001)).

3. South African households' sources of income

The South African literature distinguishes, by one set of labels or another, between at least four broad groups of household income sources: private transfers, public transfers, self-employment, and wage income (May *et al* (1995),

Carter and May (1999), Van der Berg (2000)).

In a study of poverty and labour market participation, van der Berg (1992) decomposes the sectors of employment for the South African labour force into three groups. The categorization is based on the extent to which to workers and dependants "participate in the modern consumer economy", i.e. those employed in the core consumer economy, the marginal modern sectors and the peripheral labour force. The first category consists of "the dominant high-wage modern sectors of manufacturing, government services, and other industries and services" excluding mining. The "marginal modern sector" includes "the low wage sectors of commercial agriculture and domestic services, as well as mining", although the author acknowledges that mining is no longer a low-wage sector. The "peripheral labour force" encompasses those raising their livelihoods from subsistence agriculture, the informal sector and the unemployed. Through the application of a classification similar to the above, households' income source categories are here going to be classified by origin either in the "core" sectors or in the "marginal" sectors or as being of a "peripheral" nature.[5]

Conceptual deviations from the classification by van der Berg (1992) are that wage income from the mining sector is considered "core", as are "capital income" and "self-employment income". "Peripheral income" is raised by earners that are not employed, of which the unemployed are a subset. The two non-core origins are furthermore each divided into two subcategories. Thus, here, the "core" sectors include all sectors *except the primary sectors and domestic services*, each of which constitute separate subcategories under "marginal sectors", whereas "private transfers" and "public transfers" are the two subcategories of "peripheral" income sources. The details on the income source

[5] For analyses of the relationship between rural South African households' entitlements and choices of resource allocations, May *et al* (1995) use a different household data set and apply nine different "livelihood strategy classes", some of which resemble the main income source categories applied here.

categories follow in as close approximation as possible to the wording in the IES95 questionnaire:[6]

Income originating from the core economic sectors (henceforth "Core sectors income")**:** Salaries and wages[7] from secondary sectors, including mining and quarrying, private services, public services, and residual "other" sectors.[8] *Self-employment income* in the form of net profit from business or professional practice/activities conducted on a full time basis. *Capital income* from the letting of fixed property, royalties, interests, dividends and annuities is also included.

Primary sectors income: salaries and wages as above from agriculture, fishing, and forestry.

Domestic services income: salaries and wages from private households.

Private transfers: alimony, maintenance and similar allowances from divorced spouses or family members living elsewhere and regular allowances from family members living elsewhere.

Pensions and public transfers: pensions resulting from own employment, old age and war pensions, social pensions or allowances in terms of

---

[6] As will be discussed in more detail in Section 5, the category of households *without* a main income are defined by none of their income sources contributing 66.7 percent or more to their total household incomes. The same category also includes households relying on "indirect income" which by Statistics South Africa's definition is derived from [i] hobbies, side-lines, part-time activities, or the sales of vehicles, property etc; [ii] payments received from boarders and other members of the household; [iii] the pecuniary value of goods and services received by virtue of occupation; [iv] gratuities and lump sum payments from pension, provident and other insurance or from private persons; [v] 'other income' withdrawals, bursaries, benefits, donations and gifts, bridal payment or dowries and all 'other income'.

[7] The "salaries and wages" concept includes bonuses and income from over time, commissions and directors fees, part-time work and cash allowances in respect of transport, housing and clothing.

[8] The secondary sectors encompass the Statistics South Africa (1997b) "Major sector divisions": Manufacturing; Electricity, gas and water; and Construction. "Private services" is made up of the major divisions: Wholesale and retail trade, repair of motor vehicles, motor cycles and personal and household goods, hotels and restaurants; Transport, storage and communication; and Financial intermediation, insurance, real estate and business services.

disability grants, family and other allowances, or from funds such as e.g. the Workmen's Compensation, Unemployment Insurance, Pneumoconioses and Silicosis funds.

4. Determinants and considerations from previous research

In the labour economics literature, the microeconomic determinants of an *individual's* allocation into employment and economic sectors is usually attributed to individual characteristics (see Willis (1986)). Eligibility for transfer incomes is also determined by individual characteristics such as age, disability, or parenthood. Henceforth it is implicitly assumed that the *characteristics of households' individual income earners* may serve as explanatory factors for variations in households' dominating income sources. (Descriptive statistics in Section 6 illustrate the extent to which the latter assumption is reasonable.)

It will also be shown in this section that a search for explanatory factors may be warranted in *the composition of households with respect to members' age and labour force status*. A growing body of literature further suggests that the living arrangements and demographic characteristics of South African households alter in response to the economic circumstances of individual members, such as access to certain sources of income (Klasen and Woolard (2001), Edmonds, Mammen, and Miller (2003), Keller (2003)). While no attempts are made here to draw inference as to the nature of such intra-household processes, analytical complications arise if explanatory household size and composition variables are not *statistically* exogenous.

For the above reasons, the remainder of this section first reviews some of the relevant determinants of access to certain income sources that have been recognised in previous research. Following that discussion, some findings from research on South African household formation will provide the background for a short discussion of the relevance of statistical endogeneity ("simultaneity") and areas of this study in which it may be expected.

Employment, core and peripheral sector wage income

Several studies of labour force participation, employment, and earnings have been conducted on South African data. These studies attest to the determinants of employment being found among age, experience, gender, education, marital status, and race. (Mwabu and Shultz (2000), Naudé and Serumaga-Zake (2001)). The allocation of employed individuals into core and marginal economic sectors is a process of high relevance to this study. The channels through which individual characteristics would influence this allocation occur through individual expected earnings and reservation earnings (Wambugu (2003)). The former would differ across sectors by for example skills requirements.

Economic activities may differ across regions and thereby affect the economic sectors accessible to the household, due to factors such as varying search or commuting costs. Thus in addition to all of the aforementioned determinants and given the spatial discrimination legacies discussed in Section 2, one would also expect a household's geographic location to explain variation in income sources. This study will apply the nine provinces of South Africa[9], urban and non-urban areas, as well as two non-urban "subregions" as explanatory variables. The definitions of the two subregions follow official definitions from Statistics South Africa (1997b) and refer to "tribal areas" - which would overlap with the former "homelands" – and to "agricultural or amenities areas".[10]

Unemployment, non-participation, and peripheral income sources

This section discusses the study's approach to the relationship between *dependence* on transfer income sources and unemployment and/or economically non-active status among household members. The use of household members'

[9] The empirical analyses use dummy variables for all provinces except KwaZulu-Natal which serves as the baseline.

[10] The label for this sub-region is not official but is intended to abbreviate the Statistics South Africa (1997b) definition "area with farms, agricultural holdings, holiday resorts, agricultural schools and colleges and other rural areas".

unemployment status as an explanatory factor for households' access to income sources implicitly suggests that unemployment is considered involuntary.[11]

As opposed to what will be assumed about unemployment, a healthy working-age adult's *non-participation* in the labour force is assumed to be an outcome of premeditated choice. As pointed out by Sahn and Alderman (1988), an estimated probability of labour force participation is often interpreted as the probability that a wage offer exceeds an individual's reservation wage. That reservation wage may be subject to influence from *inter alia* household composition variables, such as the number of children in the household. Presumably the relationship between the number of children in the household and the amount of non-employed household labour, that is available to assume e.g. child care responsibilities, would also affect the participation decision. Hence, the fractions of children in the household will enter as explanatory variables, juxtaposed to the fractions of unemployed and inactive, adult, non-income earners to the total number of adults.[12]

Remittances

Throughout the developing world income remitted between relatives and friends is known to be common. Economic theory encompasses a variety of motives for transfer behaviour (Cox and Jimenez (1990), Stark (1995)) and much of the theory builds on Becker's (1965, 1973, 1974) seminal representations of the "unitary household". Altruism is a fundamental driving force in such a household and both income and resources are allocated so as to maximise the combined welfare of the household. [13]

---

[11] Kingdon and Knight (1999) analyse the extent to which rural unemployment in South Africa can be considered voluntary).

[12] Sahn and Alderman also point to the fact that the more productive assets (often landholdings) a household possesses, the less likely are household members to engage in wage labour. As mentioned previously, little evidence exists in the data for households' involvement in agricultural production. Variation in access to productive assets is therefore assumed absent across households.

[13] Neither intra-household sharing of resources nor the question of the exact nature of the decision process that underlies the optimal allocation of household labour to various activities is trivial. Objections to Gary Becker's (1965) representation of the "unitary" household have been raised by Varley (1996) and seminal work on intra-household resource allocation has been conducted by Haddad and Kanbur (1990) and Thomas (1990).

Posel (2001) highlights however, that over the past two decades a large number of studies have compiled evidence that household relations and allocations are not driven purely by altruism and that household members differ both in interests and powers to implement ambitions. One informative classification of motives for remitting may be divided into "altruism" vs. "trade in an exchange of service with the receivers" (de la Brie *et al* (2002)), but several other reasons for remitting that would fall into either one or both of these categories jointly have also been put forward.[14]

Further to South African remittance behaviour, two studies have found a negative impact on private transfers from access to public pensions (Jensen (2002), Case and Deaton (1998)). Posel (2001) tests several hypotheses about remittances and estimates the impact on remitted amounts in sole migrant households. The author uses explanatory factors such as the resource base of the household (including access to pensions), the composition of the receiving household according to migrant kinships, as well as characteristics of the sender, that reflect the migrant's earnings potential and attachment to the household. The results indicate that a variety of motives spur the sending of remittances. While acknowledging that remittances are outcomes of highly complex processes, the explanatory variables in this study are restricted to the fraction of children out of household size and age of income earners which proxy strong covariates in Posel's study.[15]

Public transfers[16]

The South African social security system is quite unique to the continent. While there are social support programmes to cover other circumstances, the *Old Age*

---

[14] De la Brie *et al* (2002) provide a dense review of analyses on various remittance motives such as: insurance motives; the desire to refund the household's past expenditures; investments for the future in prospective inheritance, status or social capital; and the social security motive to remit which is largely driven by parent's age and income. Work discussed by the same authors also show that motives to remit vary between the genders.

[15] Posel (2001) finds that the presence of children and grandchildren of the migrant have positive impacts, while the presence of parents has a negative impact on remitted amounts.

[16] Details regarding the historical background, institutional characteristics and practical implementation of the South African OAP can be found in Lund (1992), Van der Berg (1994) and Case and Deaton (1998).

Pensions (OAP) system encompass some 60 percent of the total social security budget (Budlender (2000)). While a means test for eligibility does apply in practice, it is claimed to have little effect or not be binding to African households, and the receivers usually collect the maximum amount (Case and Deaton (1998), Jensen (2002)).

Endogenous household formation

In analyses of the relationship between household formation and income, common practise has until recently been to chiefly perceive living arrangements as exogenous. Several recent findings however, suggest that two-way causality may apply to South African household formation and income sources. Edmonds *et al* (2003) provide a number of findings that suggest impacts from income sources on household structures. The first relates to migration, where absent members constitute a defining characteristic of households that rely on private transfers (Wilson and Mamphele (1989)). If younger members are also encouraged to migrate due to successful outcomes of the households' previous migration histories, their migration may in turn activate a process that transplants and expands the existing demographic characteristics among consecutive generations in the household. Secondly, it has also been shown that the income from a OAPs-eligible person in the household may serve to finance other members' migration (Edmonds *et al* (2003), Posel, Fairburn, and Lund (2004)).

Keller (2003) reports higher prevalence of multi-generational demographic household structures among the poorest forty percent of households, as measured by *per capita* income. On the same note, Edmonds *et al* (2003) find that female, pensions-eligible household heads are more likely to reside with their adult children than with certain other relations. Thus, these findings may jointly suggest that OAPs could instigate multi-generational household formations, especially when pensioners are female.

A growing international literature exists on unemployment and patterns of household formation[17], most studies of which assume that household formation is exogenous. Klasen and Woolard (2001) use two-stage least squares regression techniques in order to control for causality running from unemployment to household formation around a non-labour income source. The authors find that access to state transfers increases the likelihood of attracting unemployed persons to a household and that unemployed adults reside with their parents longer than do the employed. Consistent with findings also by Bertrand *et al* (2000), the same authors find that households' collection of remittance income, pensions and other non-wage private income is correlated with lower shares of working age adults in labour force participation and employment.[18]

If living arrangements are endogenous to income generation, a reasonable assumption is that the number of members the household is capable to support is determined by the household's income level. As will be further discussed in Section 6, the case may be that *income levels differ with income sources.* If so and if income levels determine household size and/or composition, the empirical investigation must involve tests of the simultaneity between income sources and household size. Taking into account the suggested generational relationships between private transfers and pensions (public transfers) respectively, there may exist reasons to believe that the *fractions of children* and *unemployed* are endogenous to income sources, as well as the fractions of *inactive* members that would contain *inter alia* the caretakers of young children.

[17] For example, Atkinson and Micklewright (1991) as well as Arulampulam and Stewart (1995) focus on issues such as the effects on reservation wages among unemployed from the availability of other household resources. Gregg and Wadsworth (1996) and OECD (1998) provide analyses of polarisation of employment and unemployment as a consequence of concentration of unemployed individuals in households with few or no members in employment.

[18] The authors do not apply the main income source concept, but find that 60 percent of the unemployed in their study live in households where someone is employed and 20 percent live in households receiving remittances.

5. The data, sample delimitation and main income source definition

In October 1995, Statistics South Africa conducted questionnaire-based interviews on a wide range of living standard issues with almost 30 000 households, representing all households in the country and containing nearly 131 000 inhabitants. Two months later, almost 28 585 of the same households were revisited in a more detailed investigation of their incomes and expenditures. These two surveys are often referred to as the October Household Survey and Income and Expenditure Survey 1995 (henceforth "OHS/IES 95"). In the two surveys, a household is defined as "a person or a group of people dependent on a common pool of income who normally occupy a dwelling unit or a portion thereof and who provide themselves with food or the necessary supplies or arranged for such provision." A member resides four nights a week in the household. All analyses in this study, subsequent to Table 1, are furthermore conducted with the supplied IES95 household weights renormalised to sum to unity (see Deaton (1997)).[19]

For the multivariate analyses in this study, a subsample consisting of 15 441 households that met three criteria was selected. As a first criterion, only African and coloured households are examined, as these are over-represented among low-income households and would face similar historical legacies.[20]

Since the quality of the information on individuals' labour market characteristics were greater in the OHS module than in the IES module, it was deemed desirable to extract information about income earners from the former. The second

---

[19] The sample for the two surveys was stratified by province, urban and non-urban area and population group. Altogether, 3 000 enumerator areas (EAs) were drawn as primary sampling units, within each of which ten households were visited. The data concerning households were weighted by the estimated number of households in each stratum and, in accordance with instructions from Statistics South Africa, the set of weights with the Income and Expenditure Survey are applied here, as the two surveys are being linked. (Statistics South Africa (1996, 1997a, 1997b).

[20] Apartheid policies defined four "racial classifications"; African, coloured, Asian/ Indian and white. The discrimination by race ran through all aspects of life and had tremendous effects on everyone's living standards. For these reasons official statistics in South Africa still apply "racial" categories and the same approach will be followed here (referring to the same categories as "population groups").

criterion therefore requires that all income earners in a household must be identified in both surveys. Households were equipped with matching identifiers in both data sets, but individuals were not. Thus, while matching households in the two data sets was straightforward, individuals with income in the IES module had to be matched to the OHS data by means of households' unique identifiers, gender, relationships to the household head, age, and race.

In total l 32 066 earners were matched. The latter constituted 93 percent of the earners listed in the IES module and in total 6.9 percent of the households that met with the first criterion were lost from the impending analyses. Since the applied matching procedure *ceteris paribu*s becomes more tentative, the higher is a household's number of earners in the IES module, the selection into this sample could be biased towards households with few earners. More detail on the matching procedure is found in Appendix 1.

Finally, the households that have a "main income source" constitute the target group for this investigation. The magnitude of the fraction of households that do *not* have a main income source, but are *diversified* (in terms of pecuniary income sources) depends on how a main income source is defined.[21] The definition of a main income source can be gauged to the fraction of total income originating from that source. The extent to which the defining contribution of a main income source affect the fraction of households that has one, among those that met the first two criteria, is illustrated in Table 1. The second column shows the fraction of those households with an income source that meet various defining cut-off contributions. For instance, 89.2 percent of the households would have a main income source if the cut-off contribution was set to 50 percent. In this study the cut-off contribution is set at 66.7 percent, an appeal of which is that the main income source contributes at least twice as much as any other income source.

[21] One important typology in the literature revolves around whether diversification takes place out of necessity ("for survival") or arises out of opportunities for choice ("for accumulation"). The literature recognises a multitude of other motives for livelihood diversification, but the identification of such motives is not the objective of this paper (see Ellis (2000) for an extensive review of diversification-related research.).

The second row of Table 1 shows that almost 75 percent of the households with all earners identified in both data sets had an income source that meets this study's definition. From the right hand side of the table it can be seen that in more than 70 percent of the households - quite robust to which definition is applied - the main income is earned by one member, and in almost 25 percent of the households, two earners jointly raise the main income.

Table 1: Households with numbers main income earners by main income definitions, various cut-off contributions levels.

Main income's contribution to total household income	Share of households with a main income source	Number of contributors To main income			Total
		1	2	3 or more	
50%	89.2	72.7	22.8	4.5	100.0
66.7%	74.1	71.9	23.3	4.8	100.0
75%	67.1	71.5	23.6	4.9	100.0
90%	48.8	72.2	23.2	4.6	100.0
100%	27.1	74.6	21.7	3.7	100.0
Weighted figures. n = 20 834					

Two other observations are especially noteworthy. Firstly, the figures in the second column show that almost half the households raise 90 percent or more of their income from one source category and secondly, more than one-quarter of the households derive *all* their income from one source. Thus, almost regardless of the defining contribution of a main income source, households seem to rely to a large extent on *a single source of income* and on *one or very few earners*.

6. Main income sources in an earnings and labour market context

To provide rationale for the impending multivariate analyses, this section discusses four aspects of the social relevance of the main income concept based on descriptive statistics. Firstly it is shown how the distributions of main income sources differ in *urban and non-urban areas*. Secondly, the relationship between households' main income sources and *the income distribution* is discussed. Thereafter, there is a discussion of the extent to which the main income source is *representative of households' total income generation activities*. Finally,

individuals' labour market statuses are related to their households' main income source.

Urban and non-urban main income sources

For the historical reasons referred to in the Section 2, one would expect access and the distribution of households across main income sources to differ between the rural and urban samples. However, this geographical variation would not be due to a higher prevalence of agricultural activities in rural areas, as is the case elsewhere in sub-Saharan Africa. On that subject, it has been noted by Leibbrandt *et al* (2000) that the IES95 data do not adequately capture agricultural activities for own consumption. In this study's sample, 8.3 percent of all households were recorded with either slaughtered domestic animals or harvested crops in the year preceding the interview. Profit from agricultural activities should be registered in the IES questionnaire under "self-employment", but only 1.1 percent of the households that had slaughtered animals or harvested crops had records of *any* self-employment profits at all.

The above figures presumably understate the importance of agriculture, which according to May (1996), assumes several important functions as *inter alia* a supplementary source of nutrition and as a safety net for vulnerable households in South Africa. But left with little choice other than taking the data at face value, agricultural production is not treated as a separate source of income. Since the term "rural" has an intuitive connotation of agricultural activities, which is thus quite misleading in this context the term "non-urban" will henceforth be applied to areas not within municipal boundaries or that by other means fail to meet the Statistics South Africa definition of "urban".[22]

Table 2 shows the distribution of main income sources in the two subsamples. As can be seen, *Core sectors income is much more prevalent in the urban than in the non-urban sample*, with 75.8 percent and 41.6 percent of the households in each sample respectively. Further, urban main income sources are considerably more

[22]In addition, the sometimes very high population densities found in "rural" areas of South Africa raises doubts as to the appropriateness of the terminology. On this matter, Mabin (1989) defines "rural slums" as the many areas that were 'urban' in respect of their population densities but 'rural' in respect of [the absence of] proper urban infrastructure or service".

concentrated around either Core sectors or Public transfers main incomes, which together account for more than 90 percent of the households. Rural households are more reliant on public sector and private sector transfers than their urban counterparts; with the respective rural shares being 27.5 percent and 14.4 percent. These fractions are nearly twice and four times as large as their counterparts in urban areas. Clearly, location is a key factor in explaining Core sectors integration.

Table 2: Distribution of main income source categories in the sample, by location

Main income source category	Urban	Non-urban	Total
Core sectors	75.8	41.6	58.7
Primary sectors	1.4	13.6	7.5
Domestic services	3.6	2.9	3.3
Public transfers	15.2	27.5	21.3
Private transfers	4.0	14.4	9.2
Total	100.0	100.0	100.0
Spatial distribution	51.1	48.9	100.00
Weighted figures. n = 15 441			

Main income sources and the income distribution

The positions of this study's households in the income distribution are illustrated in Tables 3 and 4. These tables show the separate distributions of non-urban and urban households across ten household income brackets according to the households' main income sources. The brackets are defined by the cut-off income levels of the *full* IES95 sample's household income deciles. Accordingly, the figures in the tables can be read as e.g. 22.1 percent of this study's non-urban households that have a Primary sectors main income, fall into the first decile of the income distribution in the full population. Before turning to the analyses of the figures in these two tables, it should be noted that the fraction of households in the four lower deciles in the non-urban areas is nearly twice that of the urban.

A common trend in both areas is that roughly 65 percent of the households with Core sectors main income sources are found in the fifth to eighth deciles, whereas similar or larger fractions of households with other main income sources are found in the first to third deciles. Moreover, the concentration of non-core households in the two lower deciles is higher in non-urban areas and especially

dense for the peripheral main income sources. Clear links thus seem to exist between low household income levels and low core economy integration.

Table 3 :Distribution of main income source categories among non-urban households, by household income deciles

Non-urban households	Main income source category					All non-urban households
Income bracket	Marginal sources		Peripheral sources		Core sectors	
	Primary Sectors	Domestic Service	Public Transfers	Private Transfers		
1	22.1	33.7	24.4	34.3	3.7	17.2
2	19.1	18.9	35.5	21.1	3.6	17.4
3	23.9	17.4	13.3	20.0	7.1	13.2
4	13.6	13.1	18.8	10.3	10.2	13.1
5	11.0	9.4	5.1	6.8	17.8	11.6
6	5.8	4.1	1.6	4.9	17.7	9.4
7	3.3	1.6	0.8	2.2	15.9	7.6
8	1.0	1.6	0.2	0.2	13.0	5.7
9	0.2	0.0	0.3	0.2	7.5	3.3
10	0.0	0.4	0.0	0.0	3.3	1.4
Total	100.0	100.0	100.0	100.0	100.0	100.0
Weighted figures. n = 7 845						

Table 4: Distribution of main income source categories among urban households, by household income deciles

Urban households	Main income source category					All urban Households
Income bracket	Marginal sources		Peripheral sources		Core sectors	
	Primary Sectors	Domestic Service	Public Transfers	Private transfers		
1	20.6	27.6	26.8	39.2	1.7	8.2
2	18.2	15.1	24.8	17.3	2.3	7.0
3	14.8	20.7	14.6	18.0	5.0	7.7
4	14.5	17.0	19.7	10.4	7.5	9.9
5	12.6	9.3	7.0	7.0	12.7	11.5
6	6.7	7.2	2.8	3.4	16.1	13.1
7	10.4	2.8	2.0	2.8	17.7	14.1
8	0.2	0.4	0.9	0.5	17.5	13.4
9	1.0	0.0	1.0	1.0	13.5	10.4
10	1.0	0.0	0.3	0.3	6.3	4.8
Total	100.0	100.0	100.0	100.0	100.0	100.0
Weighted figures. n = 7 893						

Main income sources as representative of households' income generation

Tables 5 and 6 provide impressions of how representative the main income source is of a household's total income generation. The first table shows the distribution of the number of additional, *non*-main income sources in the final sample and Table 6 displays the distribution of the number of contributors to individual households' main incomes in the sample.

Table 5: Number of additional, regular sources of income, by main income source category

Number of additional regular sources of income	Main income source category					
	Core sectors	Marginal sources		Peripheral sources		Total
		Primary sectors	Domestic services	Public transfers	Private transfers	
0	84.2	92.8	90.3	92.5	96.4	88.0
1	13.7	7.0	9.5	7.0	3.5	10.7
2	1.8	0.2	0.0	0.4	0.1	1.2
3 or more	0.2	0.1	0.2	0.0	0.0	0.2
Total	100.0	100.0	100.0	100.0	100.0	100.0
Weighted figures. n = 15 442						

As can be seen in Table 5, *the vast majority of households do not have another source of regular income*. The only noteworthy deviations are found among households in the Core sectors and domestic service categories, where additional income are found in 13 percent and 10 percent of households respectively. As shown in Table 6, in approximately 70 percent of the households, the main income is earned by one individual, but deviations from the one-earner pattern are found in the Domestic services and Private transfers categories, where the corresponding figures are 84 percent and almost 93 percent respectively.

Table 6 Number of contributors to main income, by main income source

Number of earners	Main income source category					
	Core sectors	Marginal sources		Peripheral sources		Total
		Primary sectors	Domestic services	Public transfers	Private transfers	
1	68.8	69.9	84.2	69.9	92.7	71.9
2	25.1	24.4	13.6	27.3	5.9	23.3
3 or more	6.1	5.7	2.2	2.8	1.4	4.8
Total	100.0	100.0	100.0	100.0	100.0	100.0
Weighted figures. n = 15 442						

Labour force participation and the main income source categories

Table 7 shows the distribution of adults across labour market statuses by the main income source of the households in which they live.[23] The left-hand side of the table focuses on the non-participants whereas the right-hand side shows the distribution of participants across the statuses "unemployed", "self-employed" and "employed". The concentration of adults to either core or public transfer households is high at just over 83 percent. With respect to labour force participation, very large fractions of adults in households with peripheral main income sources do not participate, but the reasons for non-participation differ between the two groups. More than one-third of the adults in the Private transfers category are enrolled compared to somewhat less than one-quarter for the Public transfers category. The fraction of retired members is of a similar size to that of the enrolled in the latter category, but very small in all other main income categories in this sample.

The right-hand side of the table shows evidence of very high unemployment rates in all households in the sample. The small fraction of labour force participating adults from peripheral income households displays dramatically higher unemployment rates – at 90-95 percent –than do the participants from other households. Also the relatively *low* unemployment rate in Primary sectors households is noteworthy.

In order to facilitate the interpretation of the later results from the multivariate analyses, Table 8 displays the distribution of children and adults in the various labour force statuses across their households' main income sources. The shares of children in the various household categories follow the all-over distribution of individuals in the sample closely. The only exception is in Private transfer households where children's share is one-and-half times their fraction of all

[23] This study follows the official Statistics South Africa (1997b) definitions of expanded unemployment (including "discouraged seekers") and economically non-active (henceforth "inactive"). A "child" is defined as 14 years old or younger and the definition of an "adult" follows. The term "working-age" refer to adults below the gender-specific retirement ages (see the section on "Public transfers" below). A "retired" individual is above working age and has been captured with labour force activity status "retired" in the OHS 1995 questionnaire.

individuals. The distribution of enrolled adults is similar to that of children, whereas the employed and self-employed are vastly over-represented among households with Core sectors main incomes.

Retired adults are under-represented everywhere except in households that rely on Public transfers, where their fraction is three times that of their share of all individuals, suggesting that households may form around pensions. The under-representation of retired adults in households that depend on private transfers could imply that such dependence arises in households with neither employed nor elderly among their members. The "Other" category of non-participating adults and unemployed are slightly under-represented among Core sectors households, whereas the non-participants are over-represented at fractions of one-and-half to two times the fractions of all adults in households with peripheral income sources. Finally, among households with private transfers the fraction of unemployed is twice as large as that of adults, but the unemployment rate is only slightly larger higher than among Public transfers households.

In conclusion this far, the use of households' main income sources as indicators of integration yields a picture of integration as a partly spatially driven phenomenon, where low integration is associated with low household incomes, low labour force participation and high unemployment rates. Further, while one-quarter of the households that met the first two criteria (population group and identification) did *not* have a main income source, the main income source is of considerable relevance to income generation among the approximately 75 percent of households that *do* have one. Few households in the latter category have other income sources or other members that derive regular income, making the households also extremely vulnerable to the loss of main income earners or incomes.[24]

[24] In a dynamic perspective, Ardington and Lund (1996) raise the valid objection to the use of a "dominant source of income" for the analysis of rural livelihoods that such sources may be of a temporary nature.

Table 7: Adults' labour force status, by households' main income source

Main income source	Non-participants in labour force				Labour force status among participants				Share of all adults
	Enrolled	Retired	Other	Total	Unemployed	Employed	Self-employed	Sum	
Core	22.7	3.3	13.2	39.2	22.8	72.2	4.9	100.0	60.0
Primary	15.4	0.6	15.9	31.9	15.7	83.9	0.4	100.0	6.2
Domestic	23.5	1.4	8.7	33.7	28.5	71.3	0.2	100.0	2.8
Public tr	23.1	27.5	24.9	75.5	90.7	7.5	1.9	100.0	23.3
Private tr	37.5	1.8	24.7	63.9	94.8	3.6	1.6	100.0	7.6
Total	23.5	8.6	16.9	48.9	33.9	62.2	3.9	100.0	100.0
Weighted figures. n = 46 514									

Table 8: Distribution of individuals across households' main income source category, by labour force status

Main income source category	Children	Enrolled	Retired	Other non-participating	Unemployed	Employed	Self-Employed	All
Core	56.7	58.1	22.8	47.1	48.1	83.0	91.3	59.5
Primary	5.4	4.1	0.4	5.9	3.8	11.2	0.9	6.1
Domestic	2.8	2.8	0.5	1.4	3.0	4.2	0.2	2.8
Public tr	22.6	22.9	74.7	34.4	29.9	1.4	5.4	23.2
Private tr	12.5	12.2	1.6	11.2	15.1	0.3	2.2	8.4
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Weighted figures. n = 55 464								

7. Empirical modelling and explanatory variables

The identification of characteristics of households with different main income sources proceeds through the use of two, five-way multinomial logistic models. It is thus assumed that the probability of a given household, *i*, holding a specific income source, *m*, is a function of its endowment vector of *S* explanatory variables, X_i, and a vector of *income-source specific* parameters, β_m, according to:[25]

$$(1) \qquad \Pr(y_i \cong main \quad income \quad category \quad m) = \frac{\exp(X_i \beta_m)}{\sum_{m=1}^{4} \exp(X_i \beta_m) + 1}$$

$$i = 1,2,...,n \qquad m = 1,...,4$$

where *n* is the sample size. In order for the expression to be uniquely defined, one set of β's (for the Core sectors category in this case) is normalised to zero. By the vector of explanatory variables, the ensuing probabilities are thus functions of the characteristics that influence a household's access to various types of income.[26] It follows from equation (1) that the marginal effect of explanatory variables on the probability that household *i* has main income source *m* is given by

$$2) \qquad \frac{\partial \Pr(y_i = m \mid X_i)}{\partial x_s} = \Pr(y_i = m \mid X_i)\left[\beta_{sm} - \sum_{k=1}^{4} \beta_{sk} \Pr(y_i = k \mid X_i)\right]$$

Thus, the marginal impact depends not only on the change in the variable and the coefficient for that variable, but on the values of all other variables as well all the other slope parameters. Consequentially, marginal effects will vary with the variable values at which they are estimated. Further, the *sign* of the marginal effect of from an increase in a given variable need not match that of the slope parameter for the same variable. Hence, the individual slope parameters convey little information *per se*. The regression results are therefore presented in marginal effects format for a hypothetical household. The output is thereafter

---

[25] Long (1997) shows that this model may be derived either as a probability model or a discrete choice model.

[26] Long (1997) refers to Amemiya (1985) who has shown that "under conditions which are likely to apply in practice the implied likelihood function is globally concave, ensuring the uniqueness of ML estimates".

complemented with simulation exercises that illustrate the impacts from variables on estimated probabilities for holding the various main income sources.

Based on the discussion of previous findings in Section 4 and the support of the latter in the exploratory analyses in Section 6, the explanatory variables encompass three sets of variables. The included *earner characteristics* are age, gender and education, where households with several individuals contributing to the main income are incorporated by the use of fractions of earners in each age, gender and education category. In order to capture pensioners as prospective income earners, the cut off-lines for earner age categories have been drawn to capture the gender specific thresholds for OAP eligibility at 60 years for women and 65 years for men. *Household characteristics* include race, the number of household members and the fractions of children, unemployed and non-active in labour force out of total household size. Finally, two sets of *geographical variables* capture the provincial differences in economic endowments and the possible impact from residence in the two non-urban "subregions". Summary statistics of the explanatory variables are found in Table 9.

8. Empirical results and simulations

This section consists of two subsections. The first discusses the regression output and focuses on the general fit of the two estimated models. It discusses the significance of the estimated marginal effects on the probabilities for holding main incomes from the various categories. However, the marginal effects do not illustrate very well the discrete changes in predicted probabilities, which would follow authentic changes in earner or household composition characteristics.[27] In the second section, the results from the regression analyses are therefore illustrated by three simulation exercises. The simulations show the direction and strength of impact, as well as the absolute predicted probabilities, associated with changes in the explanatory variables.

[27] For instance, if the number of children in a household of five changes from one to two, the fraction of children changes from 20 percent to 40 percent, which is more than a marginal change. The change also affects other household composition fractions. Hence, the total effect is not readily accessible from the marginal effects output.

Table 9: Summary statistics of explanatory variables.

VARIABLE	NON-URBAN SUBSAMPLE						URBAN SUBSAMPLE					
	N	Mean	Std.dev	Median	Min	Max	N	Mean	Std.dev	Median	Min	Max
African	7227	0.93787	0.24140	1	0	1	7394	0.71788	0.45006	1	0	1
Number of non-earners	7227	3.54103	2.65037	3	0	28	7394	2.89099	2.27835	3	0	24
W Cape	7227	0.03999	0.19595	0	0	1	7394	0.15810	0.36486	0	0	1
E Cape	7227	0.26346	0.44054	0	0	1	7394	0.17758	0.38218	0	0	1
N Cape	7227	0.02767	0.16405	0	0	1	7394	0.05937	0.23634	0	0	1
Free State	7227	0.05618	0.23028	0	0	1	7394	0.12794	0.33405	0	0	1
KwaZulu-Natal	7227	0.18915	0.39166	0	0	1	7394	0.11901	0.32383	0	0	1
NW Province	7227	0.09852	0.29804	0	0	1	7394	0.08480	0.27860	0	0	1
Gauteng	7227	0.02242	0.14804	0	0	1	7394	0.17568	0.38058	0	0	1
Mpumalanga	7227	0.13754	0.34444	0	0	1	7394	0.05613	0.23018	0	0	1
Limpopo	7227	0.16508	0.37127	0	0	1	7394	0.04138	0.19919	0	0	1
Subregion 1 N-U: "semi-town" / U: "informal dwellings"	7227	0.05189	0.22182	0	0	1	7394	0.08791	0.28318	0	0	1
Subregion 2 N-U: "town w/o local authority / U: "hostels"	7227	0.05936	0.23632	0	0	1	7394	0.04274	0.20228	0	0	1
Subregion 3 N-U: "Tribal area"	7227	0.56441	0.49587	1	0	1	Not defined for urban sample.					
Subregion 4 N-U: agricultural or amenities area	7227	0.22707	0.41896	0	0	1						
Share of earners female	7227	0.42777	0.44530	0.5	0	1	7394	0.42041	0.41892	0.5	0	1
Share of earners in education category: None	7227	0.27615	0.43380	0	0	1	7394	0.11458	0.30608	0	0	1
Share of earners in education category: Primary	7227	0.45702	0.47909	0	0	1	7394	0.38386	0.45893	0	0	1
Share of earners in education category: Secondary	7227	0.14667	0.33760	0	0	1	7394	0.23388	0.39031	0	0	1
Share of earners in education category: Matriculated	7227	0.06111	0.22508	0	0	1	*					
Share of earners in education category: Tertiary	*						7394	0.11248	0.29892	0	0	1
Share of earners in age category: ≤19 yrs	7227	0.00830	0.09074	0	0	1	*					
Share of earners in age category: ≤24 yrs	*						7394	0.02637	0.16025	0	0	1
Share of earners in age category: 20-24 yrs	7227	0.03003	0.17067	0	0	1	*					
Share of earners in age category: 25-34 yrs	7227	0.21295	0.38875	0	0	1	7394	0.27130	0.41311	0	0	1
Share of earners in age category: 35-59 yrs	7227	0.48871	0.47703	0.5	0	1	7394	0.52334	0.465214	0.5	0	1
Share of earners in age category: 60-64 yrs	7227	0.06741	0.23292	0	0	1	7394	0.04761	0.19823	0	0	1
Share of earners in age category: ≥ 65 yrs	7227	0.15533	0.34845	0	0	1	7394	0.09157	0.27676	0	0	1
Share of non-earners children (≤ 14 years)	7227	0.41124	0.33924	0.44444	0	1.5	7394	0.38215	0.34897	0.4	0	3
Share of adult non-earners unemployed	7227	0.17754	0.32347	0	0	1	7394	0.21071	0.35339	0	0	1
Share of adult non-earners inactive	7227	0.51112	0.45124	0.5	0	1	7394	0.43263	0.44312	0.33333	0	1
Share of adult non-earners retired	7227	0.02861	0.12838	0	0	1	7394	0.03974	0.16016	0	0	1
Weighted figures. n = 55 464												

*) The two specifications differ in category variables left out for the avoidance of multi-collinearity. In both samples the earner age category 60-64 years old and the fraction of non-earners in employment were left out. But in the non-urban sample the matriculated education category was left out whereas the urban-sample it was tertiary education. For similar reasons the two youngest earner age categories in the urban subsample were collapsed into one category.

Regression results and marginal effects

Tables 10 and 11 display the estimated marginal effects on the probabilities for holding main incomes from the various categories. The marginal effects are computed for a household of six, with two children, one unemployed and two inactive members (in which case the sixth member could be an employed or retired non-main income earner). The household is assumed to reside in KwaZulu-Natal, in a "tribal area" for the non-urban household, with a male earner in the age category 35-69 assumed to have primary education as his highest educational achievement. Strictly speaking the marginal effects of the two models are not directly comparable, since in the rural specification two dummy variables are included for subregions which are not defined for the urban model.

Equation 2) showed that the marginal effects are partly based on estimated slope parameters and on the values of households' explanatory variables. With respect to the estimation of the slope coefficients (see Tables A2.1 and A2.2 in Appendix 2 for detail), Hausman–tests support the assumption of *Independence of Irrelevant Alternatives* for both models and Wald-tests do furthermore not support that combining any two of the defined outcome categories will improve the fit of either model. The pseudo-coefficient of determination in the estimation of the slope parameters take on a larger value for the non-urban subsample at almost 0.46 compared to 0.40 for urban areas.[28]

A further impression of the model's general fit can be derived by studying the estimates in either **bold** or *italics* in the first four columns of Tables 10 and 11 (with at least one addition symbol or asterisk). The *parameter estimates* underlying those estimated marginal effects were significant at the 10 percent level or higher. Just over half the parameter estimates were significant in both subsamples by the aforementioned measure. In both cases, significant estimates were more concentrated in the peripheral main income source categories and in the non-urban Core sectors households.

[28] The R^2 values may to some extent exaggerate explanatory powers, since the null hypothesis that the variable may have no effect on the outcome cannot be rejected for four variables at the 10 percent level in each model.

Table 10: Multinomial logit marginal effects on estimated probabilities for having main income from categories, non-urban sample

Explanatory variables	Primary Sectors	Domestic Services	Public Transfers	Private Transfers	Core Sectors
Predicted probability	1.5%	0.7%	16.1%	12.0%	69.7%
Geography					
W Cape (d) (x)	0.0097 (0.013)	-0.0039 (0.003)	-0.0418 (0.052)	-0.0744 (0.039)	0.1103 (0.069)
E Cape (d)	*0.0056* +++ (0.008)	-0.0033 (0.002)	**0.1100** *** (0.037)	**0.1902** *** (0.053)	**-0.3025***** (0.040)
N Cape (d) (x)	0.0112 (0.012)	-0.0032 (0.003)	-0.0747 (0.042)	-0.0239 (0.061)	0.0906 (0.074)
Free State (d) (xx)	-0.0036 (0.005)	-0.0012 (0.002)	**-0.0694** ** (0.033)	0.0221 (0.048)	0.0521 (0.061)
NW Province (d)	0.0059 (0.007)	**-0.0050** * (0.003)	**-0.0930** *** (0.032)	0.0424 (0.028)	0.0497 (0.042)
Gauteng (d)	**-0.0081** * (0.005)	0.0023 (0.005)	-0.0666 (0.046)	**-0.1197** *** (0.036)	**0.1922** *** (0.054)
Mpumalanga (d)	**0.0514** ** (0.025)	0.0041 (0.003)	**-0.0924** *** (0.030)	**-0.0577** ** (0.026)	**0.0945** ** (0.046)
Limpopo (d) (xx)	*0.0217* ++ (0.020)	-0.0032 (0.002)	0.0114+ (0.030)	**0.1030** ** (0.045)	**-0.1330** *** (0.048)
Tribal area (d)	**-0.0406** ** (0.018)	-0.0017 (0.003)	**0.0513** * (0.027)	**0.0794** ** (0.032)	**-0.0884** * (0.049)
Agr/Amnts area (d)	*0.1015* +++ (0.084)	*0.0138* +++ (0.013)	*0.0776* ++ (0.065)	-0.0332 (0.030)	**-0.1598** ** (0.078)
Earner characteristics					
Share female	-0.0001 (0.000)	**0.0002** ** (0.000)	**0.0013** *** (0.000)	**0.0015** *** (0.000)	-0.0029 (0.000)
Share w/no educ	**0.0003** ** (0.000)	**0.0001** ** (0.000)	**0.0039** *** (0.001)	**0.0013** *** (0.000)	**-0.0056** *** (0.001)
Share w/prim educ	**0.0003** ** (0.000)	**0.0001** ** (0.000)	**0.0030** *** (0.001)	**0.0008** *** (0.000)	**-0.0042***** (0.001)
Share w/ sec educ	**0.0001** * (0.000)	**0.0001** * (0.000)	**0.0019** *** (0.001)	**0.0007** ** (0.000)	**-0.0029** *** (0.001)
Share w/ trtry educ	*-0.0003* ++ (0.000)	*-0.0001* + (0.000)	*-0.0007* + (0.001)	**-0.0019** *** (0.001)	**0.0030** *** (0.001)
Share ≤ 24 years	**0.0001** * (0.000)	0.0000 (0.000)	**-0.0028** *** (0.001)	**0.0016** *** (0.001)	0.0010 (0.001)
Share 25-34 years	**0.0002** ** (0.000)	0.0000 (0.000)	**-0.0060** *** (0.001)	*-0.0001* +++ (0.000)	0.0059 (0.001)
Share 35-59 years	*0.0000* + (0.000)	0.0000 (0.000)	**-0.0037** *** (0.001)	*-0.0001*+++ (0.000)	**0.0039** *** (0.001)
Share ≥ 65 years	-0.0002 (0.000)	-0.0001 (0.000)	**0.0021** *** (0.000)	**-0.0014** * (0.001)	-0.0005 (0.001)
Household characteristics					
African (d) (xx)	-0.0063 (0.009)	-0.0059 (0.007)	-0.0646 (0.067)	0.0555 (0.040)	0.0212 (0.078)
(∧) Household size	*-0.0008* + (0.001)	*-0.0011*+++ (0.001)	-0.0032 (0.005)	**-0.0209** *** (0.005)	**0.0260** *** (0.006)
(∧) Share children	0.0000 (0.000)	0.0000 (0.000)	*0.0004* + (0.001)	**0.0026** *** (0.001)	-0.0030 (0.001)
(∧) Share unmp (x)	**-0.0018** *** (0.001)	-0.0002 (0.000)	*0.0120* ++ 0.008)	*0.0113* ++ (0.008)	-0.0213 (0.011)
(∧) Share inac (xx)	**0.0005** * (0.000)	0.0001 (0.000)	*-0.0038* + (0.003)	*-0.0038* + (0.003)	0.0070 (0.004)

Significance levels are based on standard errors that take the clustered nature of the sample into account.

Symbols:

(d) Marginal effect for dummy variable represented by discrete change 0 -> 1

***/**/* Marginal effect significant at 1%/5%/10% level and coefficient estimate significant at 10% or higher

+ + + / + + / + Coefficient estimate significant at 1% / 5% / 10% level

(x) / (xx) Wald test H(0): Variable's all coefficients = 0, not rejected at 5 / 10%

(∧) Endogenous, observed variable replaced by prediction.

Weighted data.

Table 11: Multinomial logit marginal effects on estimated probabilities for having main income from categories, urban sample

Explanatory variables	Primary Sectors	Domestic Services	Public Transfers	Private Transfers	Core Sectors
Predicted probability	0.2%	2.7%	6.5%	0.7%	89%
			Geography		
W Cape (d) (x)	*0.0353* +++ (0.0283)	-0.0019 (0.0034)	-0.0042 (0.0111)	-0.0076 (0.0065)	-0.0215+++ (0.0311)
E Cape (d)	0.0060 (0.0053)	0.0017 (0.0035)	**0.0540** *** (0.0140)	**0.0276** *** (0.0085)	-0.0893 (0.0184)
N Cape (d)	**0.0291** * (0.0173)	**0.0110** * (0.0062)	**0.0759** *** (0.0240)	**0.0454** *** (0.0168)	**-0.1614** *** (0.0360)
Free State (d) (xx)	0.0003 (0.0046)	*0.0064* + (0.0042)	**0.0413** *** (0.0132)	**0.0341** *** (0.0092)	-0.0821 (0.0184)
NW Province (d)	0.0144 (0.0128)	-0.0025 (0.0036)	0.0023 (0.0137)	-0.0007 (0.0089)	-0.0135 (0.0227)
Gauteng (d)	-0.0037 (0.0037)	-0.0013 (0.0036)	**-0.0314** *** (0.0092)	-0.0147 (0.0059)	0.0510 (0.0131)
Mpumalanga (d) (xx)	*0.0287*++ (0.0185)	-0.0002 (0.0048)	-0.0161 (0.0127)	-0.0100 (0.0073)	*-0.0024*++ (0.0252)
Limpopo (d)	**0.1456*** (0.0823)	-0.0054 (0.0035)	-0.0084 (0.0168)	0.0002 (0.0093)	*-0.1320*+++ (0.0811)
			Earner characteristics		
Share female	0.0000 (0.0322)	**0.0003** *** (0.0001)	**0.0006** *** (0.0001)	**0.0004** *** (0.0001)	-0.0012 (0.0002)
Share w/ no educ	*0.0001* +++ (0.0001)	**0.0003** ** (0.0001)	**0.0016** *** (0.0004)	**0.0003** *** (0.0001)	**-0.0023** *** (0.0004)
Share w/ prim educ	*0.0001* +++ (0.0001)	**0.0002** ** (0.0001)	**0.0011** *** (0.0003)	**0.0002** *** (0.0001)	**-0.0017** *** (0.0003)
Share w/ sec educ	*0.0001* ++ (0.0001)	**0.0001** * (0.0001)	**0.0007** *** (0.0002)	**0.0001** ** (0.0001)	**-0.0011** *** (0.0003)
Share w/ tertry educ	*-0.0058* +++ (0.0040)	-0.0001 (0.0001)	0.0005 (0.0003)	*-0.0003* +++ (0.0002)	*0.0055* +++ (0.0037)
Share ≤ 24 years	*0.0000* + (0.0000)	*0.0001* ++ (0.0001)	**-0.0007** *** (0.0003)	**0.0003** ** (0.0001)	*0.0002* + (0.2687)
Share 25·34 years	0.0000 (0.0015)	0.0001 (0.0001)	**-0.0015** *** (0.0003)	No effect	0.0014 (0.0003)
Share 35·59 years	0.0000 (0.0001)	**0.0001** * (0.0001)	**-0.0010** *** (0.0002)	0.0000 (0.0001)	0.0009 (0.0002)
Share ≥ 65 years	0.0000 (0.0001)	0.0000 (0.0001)	**0.0013** *** (0.0003)	0.0001 (0.0001)	-0.0013 (0.0004)
			Household characteristics		
African (d)	0.0016 (0.0046)	0.0008 (0.0027)	-0.0241 (0.0163)	**0.0137** *** (0.0047)	0.0081 (0.0189)
(∧)Household size	*-0.0015* + (0.0016)	-0.0008 (0.0008)	-0.0017 (0.0026)	**-0.0071** (0.0024)	**0.0110** *** (0.0042)
(∧)Share children	0.0001 (0.0001)	*-0.0001* + (0.0001)	-0.0002 (0.0003)	**0.0004** ** (0.0002)	-0.0001 (0.0004)
(∧)Share unempl (xx)	-0.0001 (0.0007)	0.0007 (0.0005)	0.0016 (0.0015)	0.0004 (0.0008)	-0.0026 (0.0022)
(∧)Share inact (xx)	0.0001 (0.0006)	*-0.0005* + (0.0003)	-0.0010 (0.0012)	0.0001 (0.0006)	0.0012 (0.0017)

Significance levels are based on standard errors that take the clustered nature of the sample into account.
Symbols:
(d) Marginal effect for dummy variable represented by discrete change 0 ·> 1
***/**/* Marginal effect significant at 1%/5%/10% level and coefficient estimate significant at 10% or higher
+ + +/ + + / + Coefficient estimate significant at 1%/5%/10% level
(x)/(xx) Wald test H(0): Variable's all coefficients = 0, not rejected at 5 / 10%
(∧) Endogenous, observed variable replaced by prediction.
Weighted data.

In the vertical dimension, the variables in the output are divided into three sections. The middle section of each table contains the earner characteristics and displays the highest prevalence of significant parameter estimates in both cases. This finding attests to the particular relevance of earner gender, age and education levels. Based on the hypothetical characteristic endowments of these households, having female earners is associated with a positive marginal effects on all non-core main income source categories, except for Primary sectors income in both subsamples. The effects are strongest for the non-urban peripheral sources.

Attesting to a paramount importance of adults' education, the variables reflecting such endowments display the highest prevalence of significant estimates across outcome categories. This applies to both the depicted marginal effects and to the underlying parameters. As compared to having matriculated earners, the probabilities for having non-core main income sources increase with lower levels of education in earners. The marginal effect of tertiary education is significant only in the non-urban private transfer category, where the effect is negative.

The marginal effects of earners in age categories is strongest in the Public transfers category, where its effect is consistent with most public transfers being received by elderly earners. In non-urban areas the elderly earners are less likely to receive *private* transfer income. The youngest earners are however more likely to receive private transfers and there are also positive, significant marginal effects for young age categories on the probability for non-urban Primary sectors main income.

In both subsamples, the results in the upper third of each table indicate that main income sources are subject to substantial geographical variations. Starting with the non-urban subsample, positive and significant marginal effects on the probabilities for having either transfer income are found from residence in a former "tribal area". Significant impacts are the opposite on the probabilities for having Primary or Core sectors main income. The only significant effect found for residence in an "agricultural or amenities area" is negative and relates to the Core sectors category.

In the same subsample, the strongest of all significant geographical effects is the negative impact on the probability for having Core sectors main income from

residence in the Eastern Cape. The sign of the marginal effect from residence in Limpopo is also negative albeit only one-third of the strength of the former. Residence in Mpumalanga or Gauteng however has a positive impact on the probability for having a Core sectors main income source. Further, the probability for having private transfers as a main income source is reduced by residence Gauteng. While it is highly questionable whether the urbanised metropolitan province - encompassing Johannesburg, the capital Pretoria and Witwatersrand - actually hosts non-urban households, the latter two results are consistent with the province being an industrialised sending area for remittances.

Further, residence in Limpopo and the Eastern Cape display strong, significant and positive marginal effects on the probabilities for Private transfers main income. There is also a positive, significant provincial effect on the probability for public transfers dependence from residence in the Eastern Cape. The provincial impacts for Mpumalanga are however, the opposite. Residence in that province has a positive marginal effect on the probabilities both for having Core and Primary sectors main income, the mirror images of which are significant negative effects for either transfer income.

In the urban areas, residence in the Northern Cape has a significant and relatively strong negative impact on the probability for having Core sectors income, while the impacts on all non-core main incomes are positive and significant. The province' marginal impact on the probability for private transfers reliance is the strongest significant impact in the category. Other significant, positive albeit relatively weak impacts are found from residence in the Eastern Cape and the Free State. The same three provinces also display positive marginal effects on the probability of relying public transfers. The opposite is true for Gauteng.

The lowest third of both regression tables contain the endogeneity suspected household size and composition variables. Test results show that exogeneity was not supported in any subsample for any of these variables (see Tables A3 and A4 for some detail). A future sophistication of this analysis is a log-likelihood estimator that can simultaneously compute the predicted values for the endogeneity suspects and incorporate these (as well as the residuals) in the multinomial logit estimation.

Currently, the predicted household composition variables are incorporated through a non-simultaneous two-step procedure, which leaves the second-stage standard errors smaller than would a simultaneous estimator. Hence, the test-statistics are not strictly valid (Wooldridge (1999)). In order to alert the reader of this caveat the relevant cells Tables 10 and 11 are shaded in grey (The same caveat and notation applies to Table A2.1-4.) [29] In the non-urban areas a little more than half of the twenty parameter estimates for these variables were significant, but less than half (five) of these are associated with significant marginal effects. In the urban sample, six of the parameter estimates were positive and three display positive marginal effects.

In both subsamples, household size had a positive marginal effect on the probability for having Core sectors income. The effect was the opposite on Private transfers, while increasing shares of children had positive significant marginal effect on the same probability. With respect to labour market statuses, significant marginal effects were found only with respect to the probability for having Primary sectors main income, where the fraction of unemployed household members display a slightly negative impact and the share of inactive has the opposite effect. However, in non-urban areas the parameter estimates for both unemployed and non-active are positive in the two transfer categories.

Finally, the only outcome where African population group has a significant marginal effect (and parameter estimate) is on the urban Private transfers category, where it increases the probability. This is quite surprising for the South African context where ethnical impacts are so often found in living standards related research. While the direction of causality from the household characteristics may be open to discussion, the high and relatively prevalent significance of parameter estimates are consistent

[29] The variables were tested for endogeneity by the method suggested by Rivers and Voung (1988). Additional exogenous variables in the first-stage regressions were; the number of adults in the household, fraction of adult males out of total household size, fraction of adult females out of total household size, and the fraction of adult earners out of household size. Under the assumption of normally distributed errors in the first stage regression, a two-step estimator can be used to correct for endogenous variables (Wooldridge (2002). Comparing Tables A1 and A2 in Appendix 2 to Tables A3 and A4 respectively, reveals that the magnitudes of coefficients for these variables do not differ significantly between the two regressions. Thus, different estimates are not generated when the first-step residuals "clear up" the endogeneity as compared to when the predictors are used (Wooldridge (1999).

with some interaction between household member composition and income sources, which will be illustrated at the end of the next subsection.

Simulations of predicted probabilities[30]

The first simulation in Table 12 illustrates the impact on the predicted probabilities for holding the various main income source categories from the gender of the main income earner and the household's location. The hypothetical household is identical to that for which the marginal effects in Tables 10 and 11 were computed. Thus, the household is assumed to reside in KwaZulu-Natal, in a "tribal area" for the non-urban household. It has six members; two of which are children, one unemployed and two inactive adults. The earner is aged 35-59 and has primary education. The hypothetical earner is a male in the first two rows and female in the two bottom rows. With a male earner an urban household's probability for having a Core sectors main income is close to 90 percent, while it is just below 70 percent for a non-urban household. With a female earner the corresponding probabilities are approximately one-half and one-third.

Table 12: Simulation of impact from location and main income earner's gender[31]

Location and Earner gender	Predicted probabilities for having main income from categories				
	Primary sector	Domestic services	Public transfers	Private Transfers	Core Sectors
Urban male	0.2	2.7	6.5	0.7	89.9
Non-urban male	1.5	0.7	16.1	12.0	69.7
Non-urban female	0.8	10.6	25.4	30.2	33.0
Urban female	0.1	32.2	13.7	2.4	51.5

Thus, the probabilities for holding a Core sectors main income source are roughly halved as compared to the case of the male earner. For either gender the probabilities are 20 percentage points lower in the non-urban location as compared to the urban. The relationship between the two sets of probabilities are the opposite regarding the joint probability for holding either transfer type of transfer main income. The female's probabilities are approximately twice as high in both cases and 20 percentage points higher in the non-urban case. The gender and location differences are even more

[30] The reader show be advised that not all parameter estimates upon which probabilities were calculated were significant at the 10 percent level or higher.

[31] Table A3.1 in the Appendix 3 illustrates the corresponding predicted probabilities for otherwise identical households in Mpumalanga.

extreme in the case of Domestic services. For that category the probability is ten times higher with a female earner and three times as high in urban areas compared to the non-urban.

The second simulation, starting in the upper half of Table 13, illustrates the impact of sub-regional location and the earner's education level. The default household is similar to the preceding; a non-urban household located in the Eastern Cape and has a sole female main income earner in the age category 35-59. The lower half of the table illustrates the impact of age differences, with the corresponding simulations for a household with a female earner aged 24-35. In each half of the table, the household resides in a "Tribal area" in the upper three rows and in an "Agricultural/-amenities area" in the three lower rows. The simulations illustrate the impact of three different earner education levels for each area.

Table 13: Simulation of impact from location and main income earner's education and non-urban sub-regional location[32]

Education	Location	Predicted probabilities for having Main income from categories				
		Primary Sector	Domestic Services	Public Transfers	Private transfers	Core Sector
Earner aged 35 – 59						
None	Tribal area	0.5	2.4	34.4	55.6	7.1
Some 2:ndry	Tribal area	0.3	2.0	18.1	60.8	18.8
Matriculated	Tribal area	0.2	0.8	8.1	49.1	41.8
None	Agr/amenities area	18.9	11.2	44.6	17.4	8.0
Some 2:ndry	Agr/amenities area	11.4	11.5	28.5	23.1	25.5
Matriculated	Agr/amenities area	6.8	4.8	12.8	18.7	56.8
Earner aged 25 – 34						
None	Tribal area	1.7	4.8	11.0	70.3	12.2
Some 2:ndry	Tribal area	0.7	3.4	4.8	64.2	26.9
Matriculated	Tribal area	0.4	1.2	1.9	44.9	51.7
None	Agr/amenities area	46.2	16.6	10.7	16.4	10.2
Some 2:ndry	Agr/amenities area	26.3	16.1	6.4	20.5	30.7
Matriculated	Agr/amenities area	14.3	6.0	2.6	15.1	62.0

If the female earner aged 35-59 has no education in a household residing in a tribal area, the probability that the household depends on private transfers is almost 56 percent, while that for having Core sectors main income it is just over 7 percent. If her

[32] Table A3.2 in the Appendix3 illustrates the corresponding predicted probabilities for otherwise identical households with male earners.

education included some secondary schooling, the probability for Core sector income is almost tripled and with a complete matriculation, it is almost six times as high as in the no-education case. The probabilities for the private transfers are not affected much, whereas those of the other three main income categories drop drastically with education.

Remaining with the older earner, the impacts of increasing education are very similar in the "Agricultural/amenities" area. However, the probabilities for a Private transfers main income are much lower, whereas those for Domestic services or Primary sector income are considerably higher, as seems to be the case also for the Public transfers category. The probabilities for all three latter categories decrease with education. The effects of education are very similar for the younger earner in either type of area, as are the impacts of the altered location. The probabilities for Public transfer main incomes are much lower with the young earner, while those for a holding a Core sectors are higher. In the "Tribal area" the probability for the Private transfers category is higher for the younger earner, as is that of Primary sectors main income in the "Agricultural/amenities" area. In summary, increased levels of education - and matriculation in particular - seem to vastly improve chances of households having Core sectors main income regardless of location. However, the same probability is also strongly affected by sub-regional location.

A final simulation in Table 14 illustrates the impacts from household size and composition with respect to children, unemployed and inactive adults. The initial household again resides in a non-urban, "tribal area" in KwaZulu-Natal and has a female earner aged 35-59 with primary education. In the first row, the household of four contains two children and an inactive adult. Thus, in the first three rows, the female earner may be either the inactive working aged member or an employed or retired member.

As can be seen, for a female earner in this age category, the constellation of two children and an inactive member is more likely to be supported by an employee than by a receiver of a private or public transfer. Increasing the number of children raises the probability of the earner being a transfer recipient. This finding is in line with those of Posel (2001), who shows that transfers behaviour is partially driven by the

number of children in the receiving household. However, the probability of the earner accessing Core sectors income is still the largest and virtually unaffected by the increased number of children.

Table 14: Simulated impact from household composition

Earner gender	House-hold size	Children	Adults	Predicted probabilities for having Main income from categories				
				Primary Sector	Domestic Services	Public Transfers	Private transfers	Core sector
F	4	2	1 inactive	4.9	21.3	10.2	17.6	46.1
F	5	3	1 inactive	4.0	16.8	11.9	22.2	45.0
F	5	2	1 inactive 1 unempl	0.2	5.5	27.7	52.0	14.6
M	4	2	1 inactive	8.8	23.9	2.3	11.5	53.4
M	5	3	1 inactive	7.8	20.0	2.7	14.4	55.1
M	5	2	1 inactive 1 unempl	1.0	12.3	8.4	45.1	33.3

As noted earlier, both unemployed and inactive members are over-represented in households with private transfers as the main source of income. In the third row, the replacement of one child with an unemployed member, more than doubles probabilities of having either transfer main income. The joint probability for having a transfer income is almost four-fifths. Hence a *joint presence* of adult members in such labour force statuses increases the probability that one of them is the earner, in which case the income must be a transfer and most likely a private one. Since females are more likely to be recipients of transfers, the latter result is consistent with a high prevalence of females among the rural unemployed and working-age inactive (Casale (2004)).

The increased probability for public transfers dependence with the additional unemployed member - despite the non-eligible age of the earner - may partially be driven by the previously suggested strong association between unemployed, children and female OAP earners. Such an explanation is consistent with the much lower probabilities for having a Public transfers main income if the earner were male, as shown in the table's last three rows. Thus, the impacts from the presence of children and non-employed adults on having the various main income sources illustrate considerable effects also from household composition.

9. Conclusions

This study has shown that among the majority of households in this subsample of coloured and African households captured by Statistics South Africa's 1995 *October Household Survey*, income generation largely revolves around one main income source. However, inference to the total South African population is restricted by the intentional selection only of households that meet the study's main income source definition. In this sample the main income is often earned by one single member, although in roughly one-fourth of the cases by two members. Descriptive analyses based on a classification of a household's main income sources according to core economy integration, showed that low levels of integration were associated with the lower end of the income distribution as well as with high non-participation and unemployment rates.

In the search for statistical regularities that can account for the variation in the sample's main income sources, it was found that integration is to a large extent driven by the characteristics of the household's earner or earners and much affected by the households location. Most prominently *the spread across main income source categories is much greater in non-urban areas than in the urban*, where core economy income sources account for over three-quarters of the households. Within the urban and the non-urban areas, variations in main income sources depend largely on differing characteristics of earners. Results imply that the *gender, education and age of main income earners* all have considerable impact on integration by main income sources. With small variation across non-core main income sources, the likelihood for low integration increases if the main income is earned by *individuals with low levels of education, often by women, by elderly and by earners of young working age*.

Within the urban and non-urban subsamples, main income sources are also subject to *inter-provincial variation* and in the non-urban case also by the households' *sub-regional location*. Of particular concern with low core-sector integration is the higher probabilities for transfers dependence in both urban and non-urban areas of the Eastern Cape, non-urban areas of Mpumalanga and Limpopo, as well as in the urban Northern Cape. Furthermore, non-urban household's residence in the former "tribal

areas" is associated with higher probabilities for transfer dependence and lower probabilities for accessing Core sectors income directly. Residence in agricultural or otherwise commercialised non-urban areas raise probabilities for Primary sectors or Domestic services main income sources.

Indications are furthermore that different household structures are associated with differing main income sources. After controlling for endogeneity and in line with previous findings, high fractions of unemployed persons are strongly and positively associated with public transfers (Klasen and Woolard (2001), Edmonds *et al* (2003), Betrand *et al* (2000)). Having large fractions of economically non-active members is positively associated with reliance on either type of transfer income source, but stronger for private transfers.

In non-urban areas, high fractions of children are positively associated with probabilities of transfers dependence. It has been noted by Keller (2003) that poor households differ from the non-poor in terms of generation structure.That finding is supported by the results here that public transfer are associated with high ages and strongly associated with low-income households, with unemployed and inactive members, as well as with young children. Elderly individuals receive pensions, while younger women often have young children. Multi-generation households would arise when receivers of public pensions support their children and grandchildren (Edmonds *et al (20*03), Klasen and Woolard (2001)). In support of that view, the results here were that the presence of unemployed household members and children increases the probabilities that households with female earners rely on transfers.

It may be questionable whether the patterns of living arrangements and income sources that are depicted through this 1995 data still prevail and whether derived policy implications apply. However, like many other studies, results from this investigation strongly endorse the need to stimulate employment creation. Such polices would be especially useful if they could be spatially targeted and if they could be assimilated to the very different patterns of non-integration that exist between both urban and non-urban areas as well as across sub-regions.

The considerable impact of education on Core sectors access suggests that adult literacy programmes may promote integration of marginalised or peripheral households. Finally, the empirical work in this paper highlights the importance of improving OAP and Child Support Grant take-up rates. For poverty alleviation purposes, transfers for children and young mothers would target the low-income, transfer dependent households and may also benefit the elderly. However, to the extent that household formation is endogenous to such transfers, household composition may reshape in response to increased take-up.

References

AMEMIYA, T. (1985). *Advanced Econometrics.* Cambridge, MA: Harvard University Press.

ARDINGTON, E. & LUND, F. (1995). Pensions and Development: Social Security as Complementary to Programmes of Reconstruction and Development. *Development Southern Africa*, 12(4).

ARDINGTON, E. & LUND, F. (1996). Questioning rural livelihoods, Ch. 2 in LIPTON, M., DE KLERK, M., & LIPTON, M. (Eds) (1996). *Land, labour & livelihoods in rural South Africa,* Durban, South Africa: Indicator Press.

ARULAMPULAM, W., & STEWART, M. (1995). The Determinants of Individual Unemployment Durations in an Era of High Unemployment. *Economic Journal*, 105.

ATKINSON, A. & MICKLEWRIGHT, J. (1991). Unemployment Compensation & Labor Market Transitions: A Critical Review. *Journal of Economic Literature*, 29

BECKER, G. (1965). "A Theory of the Allocation of Time." in *Economic Journal*, 175.

BECKER, G. (1973). "A theory of marriage: Part I." *Journal of Political economy*,81.

BECKER, G. (1974). "A theory of marriage: Part II." *Journal of Political economy*, 82.

BHORAT, H., HODGE, J., & DIEDEN, S. (1998). The Impact of Structural & Production Method Change on Employment Growth of Occupational Groups in South Africa. *The Trade & Industry Monitor*, 6(2). Development Policy Research Unit, University of Cape Town, South Africa.

BHORAT, H., LEIBBRANDT, M., MAZIYA, M., VAN DER BERG, S. & WOOLARD, I. (2001). *Fighting Poverty – Labour Markets & Inequality in South Africa*, Cape Town, South Africa: UCT Press.

BUDLENDER, D. (2000). *Human development*, Chapter 5 in May, J. (2000).

BUNDY, C. (1988). *The Rise and fall of the South African peasantry*. Cape Town, South Africa: David Philip Publishers.

CARTER, M. & MAY, J. (1999). Poverty, Livelihood & Class in Rural South Africa. *World Development*, 27(1).

CASALE, D. (2004) "What has the feminisation of the labour market 'bought' women in South Africa? Trends in labour force participation, employment and earnings 1995-2001", *Journal of Interdisciplinary Economics*,15."

CASE, A. & DEATON, A. (1998). Large Cash Transfers to the Elderly in South Africa. *Economic Journal*, 108(450).

COX, D., & JIMENEZ, E. (1990). "Achieving social objectives through private transfers: a review." *World Bank Research Observer,* 5(2).

DEATON, A. (1997). The analysis of household surveys: A microeconometric approach to development policy. Baltimore, MD: John Hopkins University Press.

DE LA BRIE`RE, B., SADOULET, E., & DE JANVRY, A. (2002). The roles of destination, gender & household composition in explaining remittances: an analysis for the Dominican Sierra. *Journal of Development Economics*, 68(2).

EDMONDS, E., MAMMEN, K. & MILLER, L. (2003). *Rearranging the Family? Income Support & Elderly Living Arrangements in a Low Income Country*. Working Paper, University of Dartmouth, Department of Economics.

ELLIS, F. (2000). *Rural Livelihoods and Diversity in Developing Countries*, Oxford, UK: Oxford University Press.

FOX, R. (2000). "Agriculture and rural development". Ch.9 in Fox and Rowntree (2000).

FOX, R. & ROWNTREE, K., (Eds) (2000). *The Geography of South Africa in a Changing World*, Cape Town, South Africa Oxford University Press Southern Africa.

GLEWWE, P. (1991). Investigating the determinants of household welfare in Côite d'Ivoire. *Journal of Development Economics*, 35(2).

GREGG, P., & WADSWORTH, J. (1996). It takes two: Employment Polarization in the OECD. *Centre for Economic Performance Discussion Paper* No. 304.

HADDAD, L., & KANBUR, R. (1990). How Serious is the Neglect of Intra-Household Inequality? *Economic Journal*, 100.

HODDINOTT, J. (1992). "Rotten kids or manipulative parents: are children old age security in western Kenya?" *Economic Development and Cultural Change*, 40(3).

JENSEN, R. (2001). "Public Transfers, Private Transfers and the 'Crowding Out' Hypothesis: Evidence from South Africa", *Journal of Public Economics,* 88(1-2).

JOOMA, A. (1991). *Migrancy After Influx Control*. South African Institute for Race Relations, Johannesburg, South Africa.

KELLER, S. (2003). *Household Formation, Poverty & Unemployment: the Case of Rural Households in South Africa*. Paper presented at Centre for Social Sciences Research, August 2003. University of Cape Town, South Africa.

KINGDON, G., & KNIGHT, J. (1999). "Unemployment and wages in South Africa: A spatial approach." Working Paper WPS/99.12, Centre for the Study of African Economies, Oxford University.

KLASEN, S. (2000, "Measuring poverty and deprivation in South Africa", *Review of Income and Wealth*, 46(1).

KLASEN, S., & WOOLARD, I. (2001). *Surviving Unemployment without State Support: Unemployment & Household Formation in South Africa*. Paper presented at the European Economic Association conference, August 2001, Lausanne, Switzerland.

LEIBBRANDT, M., & WOOLARD, I. (2001a). "Measuring Poverty in South Africa", in Bhorat *et al* (2001).

LEIBBRANDT, M., & WOOLARD, I. (2001b). "Household incomes, poverty & inequality in a multivariate framework" in Bhorat *et al* (2001).

LEIBBRANDT, M., WOOLARD, I., & BHORAT, H. (2000).* Understanding Contemporary Household Inequality in South Africa. *Studies in Economics & Econometrics*, 24(3).

LEIBBRANDT, M., WOOLARD, C., & WOOLARD, I. (2000). The contribution of income components to income inequality: a decomposable Gini analysis. *Journal of African Economies*, 9.

LESTER, A. (2000). Historical geography. Ch. 3 in Fox and Rowntree (2000).

LIPTON, M., DE KLERK, M., & LIPTON, M. (Eds) (1996). *Land, labour & livelihoods in rural South Africa*, Durban, South Africa: Indicator Press.

LONG, J.S. (1997). Regression models for categorical & limited dependent variables. Thousand Oaks, Sage Publications.

LOW, A. (1986). Agricultural development in southern Africa: farm-household economics and the food crisis. London: J. Currey Portsmouth.

LUND, F. (1992). "State Social Transfers in South Africa" *International Social Security Review*, 46.

LUNDAHL, M., & MORITZ, L (1996). *Det nya Sydafrika – ekonomi och politik efter apartheid*. (The new South Africa – Economics and politics after apartheid.) Stockholm, Sweden: SNS Publications.

MABIN, A. (1989). Struggle for the city: Urbanisation & political strategies of the South African State. *Social Dynamics*,15(1).

MAY, J. (1996). Land, labour & rural livelihoods in KwaZulu-Natal, Ch.2 in Lipton *et al* (1996).

MAY, J. (Ed) (2000). *Poverty & inequality in South Africa: meeting the challenge*, Cape Town, South Africa: David Philip Publishers.

MAY, J., CARTER, M., AND POSEL, D. (1995, "The composition and persistence of poverty in rural South Africa". *Data Research Africa research paper*.

MILLER, D., MULLAINATHAN, S., & MILLER, D. (2003). Public policy, extended families: evidence from pensions in South Africa. *The World Economic Review*, 17.

---

* Also published in Bhorat *et al* (2001).

MURRAY, C. (1987). "Displaced urbanization: South Africa's rural slums." *African Affairs*, 86. University of Cape Town, South Africa.

MWABU, G., & SCHULTZ, TP (2000). Wage premia for schooling & location, by gender & race in South Africa. *Economic Development and Cultural Change*, 48(2).

NAKAJIMA, C. (1970). "Subsistence and Commercial Family Farms: Some Theoretical Models of Subjective Equilibrium" in C.R. Wharton (ed.) *Subsistence Agriculture and Economic Development*, London and Portland, OR: Frank Cass.

NAUDÉ, W., & SERUMAGA-ZAKE, P. (2001). An analysis of the determinants of labour force participation & unemployment in South Africa's North-West province. *Development Southern Africa*, (18)3.

NATTRASS, J. (1981). *The South African Economy: its growth and change*. Cape Town, South Africa: Oxford University Press.

OECD (1998). *Employment Outlook June 1998*. Paris: OECD.

POSEL, D. (2001). "How do households work? migration, the household & remittance behaviour in South Africa". *Social Dynamics*, 27.

POSEL, D., FAIRBURN, J. A., & LUND, F. (2004) "Labour Migration and households: a reconsideration of the effects of the social pension on labour supply in South Africa", paper presented in October 2004 at the "*African Development and Poverty Reduction: The Macro-Micro Linkage*" conference hosted by Development Policy Research Unit (DPRU) and the Trade and Industrial Policy Strategies (TIPS) in association with Cornell University; Cape Town, South Africa.

REARDON, T. (1997). "Using evidence of household income diversification to inform study of the rural nonfarm labour market in Africa", *World Development*, (25)5.

RIVERS, D. & VOUNG, Q.H. (1988). "Limited information estimators and exogeneity tests for simultaneous probit models". *Journal of Econometrics*, 39.

SADOULET, E,. & DE JANVRY, A. (1995). *Quantitative development policy analysis*. Baltimore, MD: John Hopkins University Press.

SAHN, D. E., AND ALDERMAN, H. (1998). "The effects of human capital on wages, and the determinants of labor supply in a developing country". *Journal of Development Economics*, (29).

SALDRU (1994). *South Africans rich & poor: baseline household statistics*. Cape Town, South Africa: Southern Africa Labour & Development Research Unit, University of Cape Town.

SINGH, I., SQUIRE, L., & STRAUSS, J. (1986). *Agricultural household models: extensions & applications*. Baltimore, MD.: John Hopkins University Press.

STARK, O. (1995). *Altruism & beyond: an economic analysis of transfers and exchanges within families & groups*., Cambridge, Cambridge University Press.

STATISTICS SOUTH AFRICA. (1996). *Living in South Africa. Selected findings of the 1995 October household survey*. Statistics South Africa, Pretoria.

STATISTICS SOUTH AFRICA (1997a. *Earning and spending in South Africa Selected findings of the 1995 income & expenditure survey*, Statistics South Africa, Pretoria.

STATISTICS SOUTH AFRICA (1997b. Documentation supplied with OHS & IES1995 raw data sets. Statistics South Africa, Pretoria.

THOMAS, D. (1990). "Intra-household resource allocation. an inferential approach". *Journal of Human Resources*, XXV.

VAN DER BERG, S. (1992). "Redressing economic inequality in South Africa" in *Redistribution and affirmative action – working on the South African political economy*, Hugo, P. (Ed.). Halfway House, South Africa; Southern Book Publishers.

VAN DER BERG, S. (1994). "Issues in South African Social Security", mimeo, IBRD.

VAN DER BERG, S. (2000). "Social policies to address poverty" in Bhorat *et al* (2001).

VARLEY, A. (1996). "Women heading households: some more equal than others?" *World Development*, 24.

WAMBUGU, A. (2003). *Essays on earnings and human capital in Kenya*. Ph.D. thesis, School of Economics, Gothenburg University, Sweden.

WILLIS, R.J. (1986). Wage determinants: a survey & reinterpretation of human capital earnings functions. In Ashenfelter, O. & Layard, R. (Eds), *Handbook of labor economics*, Volume I, Elsevier Science.

WILSON, F., & RAMPHELE, M. (1989, *Uprooting poverty: the South African challenge.* Cape Town, South Africa: David Philip Publishers

WOOLDRIDGE, J.M. (2002). *Econometric Analysis of cross section and panel data.* Cambridge, MA: The MIT Press.

WOOLDRIDGE, J.M. (1999). *Introductory econometrics – a modern approach.* Stamford, CT; Thomson Learning.

WORLD BANK (1995). *Key indicators of poverty in South Africa*. Pretoria, South Africa: South African Communication Service.

Appendix 1

Individuals that were captured with any amount of income in the IES module were matched to the OHS data in three rounds;

In a first round, income earners who were listed as either the household head or the spouse of the household head were matched according to household and relationship to household head and gender. (Hence if two earners in a household were listed as a

female household head and her husband in the IES data and these details were identical in the OHS data, the earners were considered matched.) By this process 77 percent of the final sample of earners were matched.

In a second round the thitherto unmatched earners were matched according to household, age, gender and race. This procedure captured another 22 percent of the final sample. Finally, in the third round thitherto unmatched earners were matched by the same variables, but allowing for mismatch in three ways: (i) If household, gender and population groups were identical in both data sets, age was allowed to mismatch by up to three years, or by ten years; (ii) If gender and age were identical, population group was allowed to mismatch; (iii) If population group and age were identical, gender was allowed to mismatch. This last procedure captured the remaining one percent of the in all 32 066 earners that were matched. The latter constituted 93 percent of the earners listed in the IES module and in total 6.9 percent of the *households* that met with the first criterion were lost from the impending analyses.

The sample delimitation process and its final outcome is illustrated in Table A1.

Table A1: Sample delimitation process

Sample	Number of households	Share of total revisited sample	Share of revisited African and coloured subsample
Total OHS/IES sample	28 585	100.0	
African and coloured OHS/IES subsample	22 366	78.2	100.0
Above sample with all earners identified	20 834	72.8	93.1
Above sample with a main income source	15 441	54.0	69.0
Unweighted figures.			

Appendix 2

Table A2.1) Multinomial logit estimates for main income categories in non-urban sample

Number of obs/ Weighted	7548 / 10000416			
Walt chi2(108)	21075.75			
Prob > chi2	0.0000			
Pseudo R2	0.4581			
Log pseudo-likelihood	-7443486.4			
Explanatory variables	**Primary Sectors**	**Domestic services**	**Public Transfers**	**Private Transfers**
Geography				
W Cape (x)	0.351 (0.529)	-0.944 (0.633)	-0.448 (0.501)	-1.118 (0.825)
E Cape	**0.883*****(0.325)	-0.056 (0.280)	**1.090*****(0.176)	**1.520*****(0.200)
N Cape (x)	0.432 (0.416)	-0.726 (0.638)	-0.748 (0.484)	-0.344 (0.703)
Free State (xx)	-0.348 (0.373)	-0.255 (0.409)	**-0.638*** (0.348)	0.098 (0.409)
NW Province	0.260 (0.414)	**-1.267*****(0.453)	**-0.934*****(0.258)	0.234 (0.248)
Gauteng	**-1.017**** (0.418)	0.034 (0.581)	-0.779 (0.513)	**-36.143*****(0.436)
Mpumalanga	**1.355*****(0.328)	0.331 (0.284)	**0.983*****(0.220)	**-0.784*****(0.257)
Limpopo (xx)	**1.102**** (0.440)	-0.378 (0.359)	0.280 (0.220)	**0.833*****(0.227)
Tribal area	**-1.187*****(0.382)	-0.091 (0.360)	**0.504**** (0.226)	**1.209*****(0.316)
Agric./Amenit. Area	**2.304*****(0.441)	**1.340*****(0.428)	**0.654**** (0.326)	-0.064 (0.445)
Earner characteristics				
Share female	0.001 (0.003)	**0.035*****(0.003)	**0.012*****(0.002)	**0.017*****(0.002)
Share w/ no eductn	**0.030*****(0.005)	**0.028*****(0.005)	**0.032*****(0.004)	**0.019*****(0.003)
Share w/ prim eductn	**0.027*****(0.005)	**0.027*****(0.005)	**0.025*****(0.004)	**0.013*****(0.003)
Share w/ sec eductn	**0.013*****(0.004)	**0.017*****(0.005)	**0.016*****(0.004)	**0.010*****(0.002)
Share w/ tertiary eductn	**-0.022*****(0.008)	**-0.016*** (0.009)	**-0.009*** (0.005)	**-0.020*****(0.004)
Share ≤ 24 years	0.006 (0.004)	0.005 (0.005)	**-0.019*****(0.003)	**0.012*****(0.003)
Share 25-34 years	0.002 (0.005)	-0.003 0.005)	**0.046*****(0.003)	**-0.010*****(0.004)
Share 35-59 years	**-0.005*** (0.003)	-0.005 (0.004)	**0.029*****(0.002)	**-0.007*****(0.002)
Share ≥ 65 years	-0.010 (0.006)	-0.008 (0.008)	**0.014*****(0.004)	**-0.011*** (0.006)
				Household characteristics
African (xx)	-0.377 (0.367)	-0.635 (0.533)	-0.369 (0.384)	0.593 (0.635)
(∧) Household size	**-0.091*** (0.049)	**-0.190*****(0.067)	-0.057 (0.037)	**-0.212*****(0.044)
(∧) Share children	0.001 (0.005)	-0.002 (0.007)	**0.007*** (0.004)	**0.026*****(0.005)
(∧) Share unempl (x)	-0.090 (0.084)	-0.001 (0.065)	**0.105**** (0.047)	**0.125**** (0.057)
(∧) Share inact (xx)	0.022 (0.031)	0.001 (0.027)	**-0.034*** (0.019)	**-0.042*** (0.023)
Intercept	**-3.278** (0.680)	**-4.957*****(0.870)	**-1.711** (0.608)	**-4.474** (0.756)
χ-2- value Hausman test of IIA when category omitted	0.000 Supported	0.000 Supported	0.000 Supported	-5.7490 Supported

Significance levels are based on standard errors that take the clustered nature of the sample into account.
Symbols: (x) / (xx) Wald test H(0): Variable's all coefficients = 0, not rejected at 5 / 10%
(∧) Endogenous, observed variable replaced by prediction.
Weighted data.

Table A2.2) Multinomial logit estimates for main income categories in urban sample

Number of obs/ Weighted	7893/9999347			
Wald chi2(108)	5695.23			
Prob > chi2	0.0000			
Pseudo R2	0.3962			
Log pseudo-likelihood	-4849848.9			
Explanatory variables	**Primary sectors**	**Domestic services**	**Public Transfers**	**Private Transfers**
Geography				
W Cape (x)	**1.994*****(0.730)	-0.210 (0.427)	-0.064 (0.251)	-0.410 (0.385)
E Cape	0.817 (0.662)	0.276 (0.374)	**0.831*****(0.200)	**0.922*****(0.263)
N Cape	**2.001*****(0.723)	**1.000**** (0.406)	**1.113*****(0.262)	**1.322*****(0.339)
Free State (xx)	0.139 (0.792)	**0.639*** (0.365)	**0.693*****(0.213)	**1.037*****(0.268)
NW Province	1.271 (0.864)	-0.318 (0.484)	0.060 (0.283)	-0.018 (0.428)
Gauteng	-1.091 (1.026)	-0.211 (0.431)	**-1.026*****(0.280)	**-1.178*****(0.385)
Mpumalanga (xx)	**1.798**** (0.893)	-0.020 (0.555)	-0.381 (0.324)	-0.614 (0.478)
Limpopo	**3.432*****(1.060)	-0.766 (0.614)	-0.027 (0.408)	0.164 (0.437)
Earner characteristics				
Share female	0.001 (0.003)	**0.030*****(0.003)	**0.013*****(0.001)	**0.018*****(0.002)
Share w/ no eductn	**0.027*****(0.006)	**0.031*****(0.005)	**0.035*****(0.003)	**0.016*****(0.003)
Share w/ prim eductn	**0.022*****(0.006)	**0.023*****(0.005)	**0.024*****(0.003)	**0.013*****(0.002)
Share w/ sec eductn	**0.014**** (0.006)	**0.013*****(0.004)	**0.016*****(0.003)	**0.008*****(0.003)
Share w/ tertiary eductn	**-1.012*****(0.025)	-0.014 (0.013)	0.004 (0.005)	**-0.018*****(0.006)
Share ≤ 24 years	**0.007*** (0.004)	**0.013**** (0.007)	**-0.015*****(0.004)	**0.016*****(0.004)
Share 25-34 years	-0.001 (0.009)	0.005 (0.005)	**-0.031*****(0.003)	-0.002 (0.004)
Share 35-59 years	-0.009 (0.006)	**0.011**** (0.005)	**-0.020*****(0.002)	-0.002 (0.003)
Share ≥ 65 years	-0.007 (0.013)	-0.001 (0.009)	**0.026*****(0.003)	0.006 (0.007)
Household characteristics				
African	0.313 (0.971)	0.084 (0.337)	-0.400 (0.258)	**0.983*****(0.291)
(^) Household size	**-0.273*** (0.153)	-0.097 (0.080)	-0.046 (0.052)	**-0.338*****(0.072)
(^) Share children	0.014 (0.020)	**-0.014*** 0.008)	-0.005 (0.006)	**0.018*****(0.006)
(^) Share unemp (xx)	-0.012 (0.124)	0.077 (0.047)	0.034 (0.034)	0.022 (0.039)
(^) Share inacti (xx)	0.023 (0.096)	**-0.058*** (0.034)	-0.021 (0.027)	0.004 (0.028)
Intercept	**-6.098** (1.747)	**-6.409** (0.803)	**-2.312** (0.506)	**-4.942** (0.634)
χ-2- value Hausman test of IIA when category omitted	-0.381 Supported	0.000 Supported	-0.508 Supported	-0.000 Supported

Significance levels are based on standard errors that take the clustered nature of the sample into account.
Symbols: (x) / (xx) Wald test H(0): Variable's all coefficients = 0, not rejected at 5 / 10%
(∧) Endogenous, observed variable replaced by prediction.
Weighted data.

Table A2.3) Selected results from tests of endogeneity in non-urban sample

Number of obs/ Weighted	7548 / 10000416							
Wald chi2(108)	5679.29							
Prob > chi2	0.0000							
Pseudo R2	0.5967							
Log pseudo-likelihood	-7443486.4							
Explanatory variables	**Primary sectors**		**Domestic services**		**Public Transfers**		**Private Transfers**	
W Cape	0.655	(0.592)	-0.531	(0.691)	-1.373	(0.629)	-2.008	(0.706)
E Cape	0.520	(0.437)	-0.531	(0.384)	1.693	(0.204)	2.035	(0.238)
N Cape	0.495	(0.422)	-0.428	(0.634)	-0.975	(0.876)	-0.151	(1.086)
Free State	-0.381	(0.381)	-0.136	(0.415)	-0.890	(0.361)	-0.019	(0.584)
NW Province	0.275	(0.417)	-1.264	(0.462)	-0.906	(0.276)	0.260	(0.315)
Gauteng	-0.924	(0.439)	0.179	(0.599)	-1.467	(0.619)	-31.256	(0.969)
Mpumalanga	1.373	(0.326)	0.290	(0.296)	-1.397	(0.263)	-0.997	(0.290)
Limpopo	0.964	(0.465)	-0.758	(0.381)	0.645	(0.247)	1.242	(0.269)
"Tribal area"	-1.202	(0.389)	-0.126	(0.385)	1.107	(0.255)	1.880	(0.321)
"Agric./Amenities area"	2.412	(0.459)	1.440	(0.469)	0.320	(0.392)	0.041	(0.522)
Sh. MIE female	0.000	(0.003)	0.034	(0.004)	0.016	(0.002)	0.023	(0.002)
Sh. MIE No educ	0.052	(0.023)	0.051	(0.021)	0.035	(0.005)	0.022	(0.005)
Sh. MIE Pr educ	0.049	(0.022)	0.049	(0.021)	0.025	(0.005)	0.013	(0.004)
Sh. MIE Sec educ	0.036	(0.023)	0.039	(0.021)	0.018	(0.005)	0.014	(0.004)
Sh. MIE Matric educ	-0.002	(0.022)	0.004	(0.021)	-0.020	(0.007)	-0.020	(0.006)
Sh. MIE ≤ 24 yrs	0.010	(0.006)	0.012	(0.007)	-0.013	(0.004)	0.012	(0.004)
Sh. MIE 25-34 yrs	0.021	(0.018)	0.022	(0.017)	-0.058	(0.004)	-0.028	(0.004)
Sh. MIE 35-59 yrs	0.011	(0.015)	0.018	(0.014)	-0.036	(0.002)	-0.020	(0.003)
Sh. MIE ≥ 65 yrs	0.008	(0.017)	0.020	(0.016)	0.007	(0.005)	-0.007	(0.007)
African	-0.336	(0.373)	-0.389	(0.545)	-0.204	(0.506)	0.494	(0.898)
Household size	-0.157	(0.064)	-0.254	(0.081)	-0.151	(0.050)	-0.247	(0.061)
Sh. Children	-0.010	(0.013)	-0.009	(0.013)	0.010	(0.005)	0.093	(0.014)
Sh. Unemployed	-0.080	(0.085)	-0.002	(0.071)	0.191	(0.057)	0.271	(0.066)
Sh. Inactive	0.043	(0.037)	0.028	(0.032)	-0.075	(0.024)	-0.012	(0.029)
Resid. Househ size	-0.178	(0.120)	0.173	(0.151)	0.414***	(0.099)	0.501***	(0.120)
Resid. sh. Children	5.066	(4.921)	5.314	(4.640)	-0.321***	(0.095)	-0.340***	(0.083)
Resid. sh. unemployed	0.083	(0.085)	0.008	(0.071)	-0.021	(0.057)	-0.017	(0.065)
Resid. sh. Inactive	-0.041	(0.037)	-0.040	(0.033)	0.238***	(0.028)	0.242***	(0.030)
Intercept	-6.904	(3.717)	-9.555	(3.452)	-1.058	(0.788)	-10.019	(1.524)

χ – test of H_0 : residuals' parameters are all zero = 175.9 p-value = 0.0000
Significance levels are based on standard errors that take the clustered nature of the sample into account.

Weighted data.

Table A2.4) Selected results from tests of endogeneity in urban sample

Number of obs/ Weighted	7893/9999347							
Wald chi2(108)	7870.3							
Prob > chi2	0.0000							
Pseudo R2	0.5768							
Log pseudo-likelihood	-3398956.4							
Explanatory Variables	**Primary Sectors**		**Domestic Services**		**Public Transfers**		**Private Transfers**	
W Cape	2.000	(0.723)	-0.133	(0.432)	-0.151	(0.281)	-0.549	(0.445)
E Cape	0.804	(0.660)	0.409	(0.380)	1.227	(0.239)	1.191	(0.358)
N Cape	1.982	(0.724)	1.051	(0.423)	1.685	(0.318)	1.813	(0.452)
Free State	0.120	(0.791)	0.778	(0.370)	1.059	(0.264)	1.347	(0.369)
NW Province	1.262	(0.865)	-0.248	(0.491)	0.113	(0.331)	0.107	(0.557)
Gauteng	-1.088	(1.028)	-0.097	(0.434)	-1.142	(0.331)	-1.098	(0.446)
Mpumalanga	1.787	(0.884)	0.018	(0.551)	-0.363	(0.431)	-0.728	(0.567)
Limpopo	3.449	(1.046)	-0.723	(0.627)	-0.085	(0.417)	0.126	(0.517)
Sh. MIE female	0.001	(0.003)	0.031	(0.003)	0.017	(0.002)	0.023	(0.002)
Sh. MIE No educ	0.026	(0.005)	0.032	(0.005)	0.040	(0.004)	0.019	(0.005)
Sh. MIE Pr educ	0.022	(0.006)	0.024	(0.005)	0.025	(0.004)	0.013	(0.004)
Sh. MIE Sec educ	0.014	(0.006)	0.013	(0.004)	0.016	(0.004)	0.006	(0.004)
Sh. MIE Matric educ	-1.262	(0.026)	-0.015	(0.013)	0.002	(0.007)	-0.013	(0.008)
Sh. MIE ≤ 24 yrs	0.006	(0.005)	0.017	(0.007)	-0.018	(0.005)	0.007	(0.005)
Sh. MIE 25-34 yrs	-0.001	(0.009)	0.010	(0.006)	-0.043	(0.004)	-0.022	(0.005)
Sh. MIE 35-59 yrs	-0.008	(0.005)	0.015	(0.005)	-0.029	(0.002)	-0.017	(0.003)
Sh. MIE ≥ 65 yrs	-0.010	(0.012)	0.001	(0.009)	0.032	(0.005)	0.018	(0.009)
African	0.302	(0.945)	0.069	(0.331)	-0.312	(0.291)	1.016	(0.375)
Household size	-0.270	(0.152)	-0.102	(0.079)	-0.199	(0.071)	-0.379	(0.093)
Sh. Children	0.014	(0.019)	-0.011	(0.008)	0.013	(0.008)	0.078	(0.019)
Sh. Unemployed	-0.009	(0.120)	0.069	(0.047)	0.042	(0.039)	0.054	(0.053)
Sh. Inactive	0.023	(0.092)	-0.052	(0.034)	-0.008	(0.032)	0.070	(0.043)
Resid. househ size	0.554***	(0.210)	0.296*	(0.161)	0.513***	(0.139)	0.697***	(0.185)
Resid. sh. children	0.015	(0.039)	0.051	(0.038)	0.043	(0.033)	0.006	(0.038)
Resid. sh. unemployed	0.003	(0.124)	-0.068	(0.048)	0.100**	(0.041)	0.142***	(0.054)
Resid. sh. inactive	-0.025	(0.091)	0.048	(0.034)	0.150***	0.034)	0.110***	(0.041)
Intercept	-6.208	(1.688)	-7.097	(0.836)	-2.670	(0.616)	-9.060	(1.596)

χ – test of H_0 : residuals' parameters are all zero = 24.17 p-value = 0.0198
Significance levels are based on standard errors that take the clustered nature of the sample into account.
Weighted data.

Appendix 3

Table A 3.1.) Simulation of impact from location and main income earner's gender; Mpumalanga.

Location and earner gender	Estimated probabilities for holding main income categories; Mpumalanga				
	Primary sectors	Domestic services	Public transfers	Private transfers	Core sectors
Urban Male	1.3	**2.7**	**4.5**	**0.4**	91.2
Rural Male	6.7	**1.1**	**6.8**	**6.2**	79.2
Rural Female	4.1	**20.0**	**12.8**	**18.6**	44.5
Urban Female	0.8	**33.4**	**9.9**	**1.4**	54.5

Table A 3.2) Simulation of impact from location and main income earner's education and non-urban subregional location; Male earners.

Education	Location	Estimated probabilities for holding main income categories				
		Primary Sector	Domestic Services	Public Transfers	Private transfers	Core sector
Rural male earner aged 35 – 59 in the E Cape						
None	Tribal area	1,7	0,3	36,3	36,7	25,1
Some 2:ndry	Tribal area	0,7	0,2	15,1	31,7	52,4
Matriculated	Tribal area	0,3	0,0	4,5	17,2	78,0
None	Agr/amenities are	41,1	0,8	31,6	7,7	18,8
Some 2:ndry	Agr/amenities area	21,4	0,7	17,3	8,8	51,8
Matriculated	Agr/amenities area	8,9	0,2	5,4	5,0	80,5
Rural male earner aged 25 – 34 in the E Cape						
None	Tribal area	5,1	0,5	10,9	43,3	40,2
Some 2:ndry	Tribal area	1,6	0,3	3,5	29,2	65,4
Matriculated	Tribal area	0,5	0,1	0,9	13,8	84,7
None	Agr/amenities area	21,8	0,7	8,5	16,8	52,2
Some 2:ndry	Agr/amenities area	6,3	0,3	2,6	10,7	80,1
Matriculated	Agr/amenities area	2,0	0,1	0,6	4,5	92,8