Homing in on the Core: Households Incomes, Income Sources and Geography in South Africa

Sten Dieden
University of Gothenburg
sten.dieden@handels.gu.se

Development Policy Research Unit
Working Paper 04/90

December 2004
ISBN 1-920055-07-X

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 utilised 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.

Acknowledgements

While any defects or shortcomings in this work are entirely my own responsibility, I am 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.

This publication was sponsored by the Secretariat for Institutional Support through Economic Research in Africa (SISERA).

Development Policy
Research Unit
Tel: +27 21 650 5705
Fax: +27 21 650 5711

Table of Contents

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 to 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 utilised 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 utilises the perception that geographical location may affect household income levels via variations in the accessibility of different income sources across locations.

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 recognise 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 categorisation of South African households by their income sources provides a composite appreciation of some key facets of deficient household incomes in the country.

The empirical analysis involves a comparison of the results from three reduced form Weighted Least Squares (WLS) regression specifications. All specifications have a set of household characteristics as explanatory variables in common. Two of the specifications are

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))

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 utilised 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 utilisation of province dummies as instruments come at a cost of lost explanatory power in the second-stage regression, the third specification utilises 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 utilisation of 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 categorisation is based on the extent to which 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-mentioned 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", which is explained in more detail below, where the income sources in each category are listed and described in as close approximation as possible of the wording in the IES95 questionnaire. The composition of the categories is as follows:

> **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 the Workmen's Compensation, Unemployment Insurance, or 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 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.

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.

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. Such attributes include low levels of education, low or high age, and female-headed households. In addition large household sizes and/or many dependants as well as location in rural areas are associated with low incomes. 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 the 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 is even more relevant 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. Each household is supplied with a weight 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 these 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* the sampled households by the IES95 in per-adult-equivalent income deciles by population group.[5]

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").

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			
	African	Coloured	Asian	White
1	*96.4*	3.6	0.0	0.0
2	*94.2*	5.5	0.2	0.2
3	*90.0*	9.1	0.4	0.5
4	*86.6*	11.6	1.1	0.7
5	*81.1*	13.9	2.2	2.8
6	76.3	13.8	3.6	6.3
7	67.0	12.5	6.0	14.5
8	49.7	9.2	6.7	34.4
9	24.8	5.0	4.8	65.4
10	29.2	3.8	2.6	64.5
Total	69.5	8.8	2.8	18.9

This study uses a sub-sample consisting of 19 914 of the revisited households, the selection of which was 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 headed by individuals who belong to either 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 base. Households in the two data sets are easily matched, since they were equipped with matching identifiers in both data sets, whereas individuals were not. Individuals that were captured as income earners 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 do 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. Furthermore, irrespective of which definition is applied, 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 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 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 above that of 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 of diversification. The sole exception is "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.

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 (see. 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.

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 INCOME SOURCE	% AGE SHARE OF HOUSEHOLDS DERIVING INCOME	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						

Among 19 percent of the households that access Indirect income, the source's contribution falls in the interval one-third-to two-thirds, classifying the households into the Diversifying category. Consequently, Indirect income contributes more than one-third of the income in the latter category. In the column with the one-third-to two-thirds contributions it can also be seen that substantial fractions of the Diversifying households access Core or Primary sectors income 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.

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 of 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 overall distribution of households in the sample.

The last column of Table 4 lists 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, in that the mean incomes of households with Core sector or Mining and Quarrying main income sources are found at R12 854 and R14 536 respectively, 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 that of the full sample, it is surprising to find the mean in the Indirect Income at 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 transfers as their 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.

6.1 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 apply it to log per-capita income in the same data set and motivate 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. Following is a list of variables common to all specifications in this study briefly motivated along the lines of 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 been 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 (16 years old or older) that are unemployed or non-active by the expanded definition for unemployment.[8] The left-out labour market status category is thus 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 by their different regional economies discussed in the next section. With respect to the explanatory variables that have been listed thus far, expectations are that the signs of their coefficient estimates would match closely to those estimated by Leibbrandt and Woolard (2001). Hence, belonging to African population group is expected to have a negative impact on income as is higher numbers of household members of all age categories and genders 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

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.

dummies, the analyses by Leibbrandt and Woolard (2001) returned no significant difference in income levels between the Western Cape (W Cape), KwaZulu-Natal (KZN) 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 (N Cape) and the Free State, followed in rank by the Eastern Cape (E Cape), the North West Province (NW Province), and Limpopo.

The variables representing households' utilisation of income sources are included in the remaining two specifications. The inclusion of these variables is an attempt to investigate whether partial impacts on income levels exist, that originate in the utilisation 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 characteristics that may also determine households' allocations to main income source categories or the returns from income sources. The 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. 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, the last specification contains six variables representing the continuous fractions of total income derived from each source. With respect to the expected 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 and relative magnitudes of 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, 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 $1 \text{x } k$ vector 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.

6.2 Analytical concerns

This subsection discusses two complications that arise from the utilisation 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, the current values of a number of the explanatory variables – such as labour force participation and income sources utilised –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 recognises 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 apply, for example, to the 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 in itself be determined by household income levels. This study utilises 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 composition of the provinces, 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, KwaZulu-Natal, Gauteng and Limpopo – contain nearly 65 percent of the working-age population[10], but with very dissimilar distributions across rural and urban areas. In the Eastern Cape, KwaZulu-Natal, the North West Province, Mpumanlanga, and Limpopo, most of the population is rural, although the Durban metropole is situated in KwaZulu-Natal, which is the third largest city in South Africa. At the other extreme are found the largely urbanised provinces of the Western 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 many of South Africa's gold mines.

The Northern Cape is scarcely populated but highly urbanised. The province contains largely desert and savannah areas, but also some of the country's vast diamond findings near its capital, Kimberley. From there the bushy highland landscape, the "Karoo", extends into the largely agricultural, but also relatively urbanised Free State, with Bloemfontein as its capital. It is also the country's judicial capital. Other fertile farming areas are found south and east of the coastal mountain ranges in the E and Western Cape and in KwaZulu-Natal, which in turn also host the prosperous and industrial coastal cities of Port Elizabeth, Cape Town and Durban, all of which are among the largest ports on the African continent.

The conditions in the four most populous provinces are likely to have a large impact on the extent to which provinces covary with Main income source categories. Table 6 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 also high and the expanded unemployment rate is among the lowest, while its official unemployment rate 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 Eastern 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,

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.

noteworthy that the official unemployment rate at 27 percent in the Eastern Cape is almost 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. In both cases half of the Core Sector employment is found in Public service which leaves the provinces ranked as number one and two in this respect.

In the remaining most populous province, KwaZulu-Natal, the rural dwellers constitute 70 percent of the working- age population. The unemployment rates are high and the employed are underrepresented among the working-aged, but not by as much as in Limpopo or the Eastern 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 Eastern 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 its 28 percent fraction of Diversifying households ranks as the third largest among the provinces.

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		9
E Cape	67	33		17
N Cape	32	68		2
Free State	46	54		7
KZN	70	30		19
NW Prov.	70	30	100	9
Gauteng	7	93		14
Mpumalanga	79	21		8
Limpopo	92	8		13
All	57	43		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 Western 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 Western Cape have Core sector main incomes, while 9 percent rely on Primary sector income. The North West Province hosts a fraction of the working-aged which is similar to that of the Western Cape 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 North West 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 the Mining and Quarrying sectors, with a similar fraction of households' Main income sources.

Almost one-quarter of the employees in the North West 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 Northern 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 Northern 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 Western Cape, at R14 035 and R10 090 respectively, where main incomes from Core sector are most common. At the opposite end one finds the Eastern Cape with high dependence on transfers and the average income at R 5 846. In the Northern 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 third and fifth highest fractions of Public transfers dependency respectively. 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 include province dummy variables and serves as a benchmark (henceforth "the geography specification"), whereas the other two-stage specifications 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 (among which Specifications 2 and 3 include income sources) is postulated. The general relationship is modelled as:

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

where Y_i represents log annual per-adult-equivalent income for household i and X_i is a vector common to all specifications that contains variables reflecting household characteristics. 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 specifications and income share specifications respectively, and zero elsewhere. The province dummies are symbolised 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 deliberate selection of only African and coloured households for this analysis renders the sample no longer representative of the entire South African population (including also the Indian/Asian and white households). The need to identify all income earners in both data sets led to a further loss of observations. Consequently, the applied sample differs from that for which the original scaling factors were computed to emulate the population size in each strata and cluster. However, these weights still contain

information about the *relative* representativity of the observations. While not returning a representative sample, the application here of the original weights renormalised 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]

8.1 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. $\mathrm{E}(x_k\varepsilon) \neq 0$). Endogeneity usually arises in applied 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:

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

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

where Z_1 is a vector of explanatory variables, the δ and γ vectors and the scalar u 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 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 Z_1-vector are a subset of Z which in turn contains at least one element not in Z_1 and that this element is partially correlated with y_2, but not simultaneously determined with y. The maintained exogeneity of Z implies crucially that $E(Z'u) = 0$ while the concern here is with the

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)

validity of $E(\gamma_2 u) = 0$. Since $E(Z'u) = 0$ and assuming $E(\gamma_2 v_2) = 0$ 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 as:

$$u = v_2 \rho_1 + e_1 \tag{4}$$

where $\rho_{1=E(v_2 u)/E(v_2^2)}$ and it can be shown that $E(v_2 e_1) = 0$ and $E(Z' e_1) = 0$

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

$$y = \mathbf{Z_1}\delta + \lambda y_2 + \rho_1 v_2 + e_1 \tag{5}$$

Following an OLS regression of equation (5) 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 $\mathbf{Z}$ not in $\mathbf{Z}_1$. 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 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 access to the income source categories constitutes a typical sample selection problem, if the same variables that explain the *magnitude of the fraction* of total income derived from a specific income source also explain a households' *utilisation* of the source. Similarly, the fact that a household accesses a certain income source does not necessarily 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 utilise 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		
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		
toNumber 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	Values if main income source	
Indirect income main income source	0.041	0.198	Mean	Std.dev
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. It will be followed by a summary presentation of the results for the variables common to all three specifications, after which the impacts from the variables unique to each specification will be discussed.

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.) Hence, 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 out 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 these authors discuss their results in detail the interested reader is referred to this particular study.[13] The following two paragraphs briefly summarise the results with respect to the estimates for the variables common to all specifications. Thereafter the focus will rest mainly with the impact of income source categories on income levels and what may be learnt from these impacts by comparing the results with the geography specification.

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 from the dummies specification; increasing *numbers* of working-age adults of either gender

13 As mentioned the geography specification in this study is similar – although not identical - to that devised by Leibbrandt and Woolard (2001). The signs and significance of 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.

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 the 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.

The analysis now turns to the impacts on income levels from the Main income source categories. An immediate result in that respect is that 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 mean incomes. The estimated partial coefficients are of lower absolute magnitudes than the difference in mean incomes both for the two transfer categories and for the households in the Primary sector category. These results imply that the households in the transfers and Primary sector categories *systematically hold more unfavourable endowments of controlled-for characteristics* (henceforth "other characteristics"*)* than do the Core sector households. In the absence of such regularities in endowments, the former three categories of households would have been less worse-off compared to households in the Core sector households.

Along the same line of reasoning the results indicate that the Mining and Quarrying sectors 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 not 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 Western Cape, the North West 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 Northern 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 Eastern 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 Northern Cape, Free State and North West Province would have been *even worse off*, as compared to the left out province. However, the households in the Eastern Cape and Mpumalanga would relatively speaking have *been better off* if it were not for their systematically more unfavourable endowments.

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)).

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 Western Cape and Gauteng have the largest fractions of households in the Core sectors category, while the Eastern 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 upon 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 utilised 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' 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. Indications from the comparison of the different model specifications are that households' location in the rural-urban dimension, education levels among adults and provincial location are 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, measures that facilitate and increase both public transfers, conditional on their appropriate funding, are 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 possible simultaneous determination of income levels, income sources, and household characteristics among 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., 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 utilised 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.[15] The sample delimitation process is illustrated in Table A1.1.[16]

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			

15 Adults are defined as individuals who are 16 years or older.

16 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

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							

Table A2.2: Fractions of households with number of contributors to main income, by main income source category.[17]

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							

17 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

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 one stars (***/** / *) indicate significance at the one, five and ten percent levels 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

Table A3.1: Regression based tests of endogeneity

Dependent variable: log per-adult-equivalent income		
Variale	**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 QuarryingSectors, 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	**0.03 5***** (0.012)	-0.065 (0.062)	**-0.097 ***** (0.03 1)	**-0. 07 1***** (0.027)	**-0.032 *** (0.019)	**-0.12 4***** (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)	0.00 1 (0.001)	-0.005 (0.003)	**-0.0 20***** (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 (0.00 2)	**-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*** (0.10 8)	**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 (0.06 4)		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 significant. 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)	**0.32 1*** (0.1 11)**	1.92 5 (2.5 50)	**-0.338 *** (0.099)**	1.906 (2.27 8)	0.21 9 (0.062)
Number of children 0-7 in household	-1.22 6 (1.19 5)	**-0.09 4*** (0.032)**	**-1.76 9*** (0.488)**	**-0.032 ** (0.019)**	**-1.92 4*** (0.545)**	**0.073 *** (0.016)**
Number of children 8-15 in household	-0.441 (1.01 9)	**-0.123 *** (0.028)**	**-1.648 *** (0.536)**	**-0.09 2*** (0.018)**	**-1.3 91*** (0.50 9)**	**0.11 3*** (0.01 6)**
Number of female adults in household	**-6.521 *** (1.913)**	**0.204 *** (0.068)**	**-2.61 8*** (0.760)**	**0.20 5*** (0.022)**	**-7.048 *** (0.87 1)**	**0.281 *** (0.0 20)**
Number of male adults in household	**-10.037 *** (1.551)**	0.02 6 (0.03 6)	**-2.325 *** (0.74 1)**	**0.211 *** (0.020)**	**-9.338 *** (0.623)**	**0.22 8*** (0.019)**
Number of elderly in household	**-6.402 ** (3.144)**	**-0.171 ** (0.07 8)**	**-14.13 7*** (1.369)**	**-0.54 1*** (0.042)**	**-6.96 6*** (1.306)**	**-0.2 70 *** (0.034)**
Share of adults with primary education	-0.010 (0.03 1)	**0.00 2* (0.00 1)**	**0.04 5*** (0.01 6)**	**-0.002 *** (0.00 1)**	**0.053 *** (0.019)**	0.000 (0.000)
Share of adults with secondary education	-0.05 3 (0.04 2)	**0.004 *** (0.001)**	**0.096 *** (0.02 7)**	**-0.01 2*** (0.00 1)**	**0.09 1*** (0.026)**	**-0.003 *** (0.00 1)**
Share of adults with matriculation	-0.00 5 (0.03 6)	0.00 1 (0.001)	0.02 4 (0.047)	**-0.01 20 *** (0.00 2)**	**0.076 *** (0.044)**	**-0.011 *** (0.001)**
Share of adults with tertiary education	0.07 2 (0.05 3)	-0.00 3 (0.00 2)	-0.201 (0.15 1)	**-0.03 20 *** (0.003)**	-0.02 3 (0.12 3)	**-0.02 4*** (0.00 2)**
Share of adults unemployed	0.061 (0.05 8)	**-0.00 4*** (0.001)**	**-0.056 * (0.03 2)**	**-0.002 *** (0.00 1)**	**0.10 3*** (0.03 7)**	0.00 1 (0.00 1)
Share of adult non- participants	**0.605 *** (0.073)**	**-0.00 6*** (0.00 1)**	**0.097 *** (0.031)**	**-0.01 7*** (0.00 1)**	**0.34 8*** (0.03 8)**	**-0.019 *** (0.00 1)**
Rural location	4.652 (3.49 6)	**-0.193 * (0.10 7)**	**6.69 6* (3.15 9)**	**1.473 *** (0.092)**	**-5.37 2*** (1.31 2)**	0.002 (0.042)
W Cape	**30.200 *** (11.284)**	-0.20 3 (0.216)	**8.842 ** (3.530)**	**0.55 6*** (0.1 40)**	-4.677 (3.0 20)	0.01 1 (0.085)
E Cape	**11.12 2*** (7.96 8)**	0.04 3 (0.166)	-2.84 3 (3.418)	0.022 (0.093)	**-5.73 1*** (2.224)**	0.045 (0.062)
N Cape	**16.28 3*** (10.665)**	**0.592 *** (0.21 8)**	**6.184 * (3.31 2)**	**0.677 *** (0.13 6)**	**-7.224 *** (2.98 2)**	**0.292 *** (0.08 6)**
Free State	**3.094 *** (9.603)**	**0.743 *** (0.176)**	**-11.9 38*** (2.589)**	**0.73 6*** (0.115)**	**-15.13 3*** (1.897)**	**0.520 *** (0.064)**
North -West Province	**4.2 40 *** (9.280)**	**0.824 *** (0.168)**	**-6.869 * (3.65 1)**	**0.32 8*** (0.12 1)**	**-10.991 *** (2.18 4)**	-0.00 5 (0.07 8)
Gauteng	**4.441 *** (9.84 2)**	**0.702 *** (0.189)**	4.593 (3.20 1)	**-0.2 94* (0.14 5)**	-0.27 7 (2.498)	0.026 (0.082)
Mpumalanga	**3.168 *** (9.672)**	**0.84 7*** (0.17 4)**	**11.41 3*** (2.95 1)**	**0.54 3*** (0.1 20)**	**5.400 ** (2.4 40)**	**0.13 2* (0.07 2)**
Limpopo	**12.42 2*** (9.269)**	**0.355 ** (0.180)**	3.59 9 (3.875)	-0.064 (0.11 7)	**5.6 50* (3.08 6)**	**-0.18 4** (0.088)**
Share of females in the household		**-0.026 *** (0.00 5)**		**-0.00 5*** (0.001)**		**0.014 *** (0.001)**
Migrant head		**0.546 *** (0.171)**		**-0.418 *** (0.08 5)**		**-0.148 ** (0.07 4)**
Intercept	**108.878 *** (17.919)**	**-1.72 6*** (0.238)**	**51.674 *** (5.36 6)**	**-0.796 *** (0.16 8)**	**60.481 *** (4.037)**	**-1.76 8*** (0.10 6)**
Athrho		**-1.28 3*** (0.43 1)**		**0.196 *** (0.074)**		**-0.15 1** (0.061)**
Lnsigma		**3.1 40*** (0.174)**		**3.117 *** (0.019)**		**3.167 *** (0.018)**
Rho		-0.857 (0.114)		0.19 4 (0.071)		-0.149 (0.0 60)
Sigma		23.102 (4.03 4)		22.572 (0.428)		23.74 2 (0.43 1)
Lambda		-19.804 (6.006)		4.3704 (1.64 5)		-3.547 (1.4 40)

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	0.000 0		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)	**0.07 1*** (0.014)	0.064 (0.178)
Number of children 8- 15 in household	**-1.689** *** (0.346)	**0.031**** (0.014)	**1.70 4***** (0.455)	**0.07 4***** (0.014)	-0.109 (0.168)
Number of female adults in household	**-5.824** *** (0.390)	**-0.044**** (0.018)	**-5.83 4***** (0.514)	**-0.141** *** (0.018)	**0.58 4***** (0.201)
Number of male adults in household	**-5.82 2***** (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 ad ults with primary education	**-0.05 4***** (0.014)	**-0.003** *** (0.000)	0.005 (0.023)	**0.001**** (0.001)	**-0.0 30***** (0.008)
Share of adults with secondary education	**-0.186** *** (0.019)	**-0.00 7***** (0.001)	0.036 (0.025)	**0.002** *** (0.001)	**-0.084** *** (0.010)
Share of adults	**-0.268** *** (0.029)	**-0.007** *** (0.001)	0.020 (0.032)	-0.001 (0.001)	**-0.08 7***** (0.011)
Share of adults with tertiary education	**-0.410** *** (0.050)	**-0.00 8***** (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)	**0.01 7***** (0.001)	**-0.02 8***** (0.009)
Rural location	**2.683**** (1.082)	**-0.250** *** (0.034)	**8.3 71***** (1.593)	0.040 (0.040)	**6.38 4***** (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	**11.73 4***** (1.368)	0.066 (0.049)	**10.19 2***** (2.020)	**0.116**** (0.054)	**-4.083** *** (0.898)
N Cape	**8.61 2***** (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)	**-0.2 30***** (0.066)	2.289 (2.521)	0.029 (0 .075)	1.744 (1.206)
Gauteng	-0.245 (2.672)	**-0.52 8***** (0.079)	**-10.628**** (4.408)	**-0.60 3***** (0.092)	0.290 (1.170)
Mpumalanga	**5.006**** (1.981)	**-0.40 1***** (0.061)	-1.359 (2.651)	**-0.31 5***** (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)		**-0.17 4***** (0.063)	
Intercept	**56.5 80***** (2.925)	**-1.29 2***** (0.094)	-1.008 (5.480)	**-1.94 5***** (0.119)	**18.22 1***** (1.397)
Athrho		**-0.089*** (0.048)		**0.41 5***** (0.056)	
Lnsigma		**3.30 4***** (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)	

Sig nificance 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 note 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 utilised.

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 utilised.

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.