

**Technical Paper Series**



**Technical Paper 2005:2**

**Forming Representative  
Household and Factor Groups  
for a South African SAM**

*Eisenburg  
March 2005*

# PROVIDE PROJECT

The Provincial Decision-making Enabling Project

## Overview


The Provincial Decision-Making Enabling (PROVIDE) Project aims to facilitate policy design by supplying policymakers with provincial and national level quantitative policy information. The project entails the development of a series of databases (in the format of Social Accounting Matrices) for use in Computable General Equilibrium models.


The National and Provincial Departments of Agriculture are the stakeholders and funders of the PROVIDE Project. The research team is located at Elsenburg in the Western Cape.

## PROVIDE Research Team

Project Leader:	Cecilia Punt
Senior Researchers:	Kalie Pauw Melt van Schoor
Junior Researchers:	Benedict Gilimani Lillian Rantho
Technical Expert:	Scott McDonald
Associate Researchers:	Lindsay Chant Christine Valente

## PROVIDE Contact Details

 Private Bag X1  
Elsenburg, 7607  
South Africa

 [ceciliap@elsenburg.com](mailto:ceciliap@elsenburg.com)  
 +27-21-8085191  
 +27-21-8085210

For the original project proposal and a more detailed description of the project, please visit [www.elsenburg.com/provide](http://www.elsenburg.com/provide)

# Forming Representative Household and Factor Groups for a South African SAM<sup>1</sup>

## Abstract

*This Technical Paper explains the formation of representative households and factor groups for inclusion in the PROVIDE Project Social Accounting Matrix. A general guideline for forming household and/or factor groups is that they should reproduce the socio-economic stratification within the society as accurately as possible. Such groups should also be made up of 'relatively homogenous' groups that are easily recognisable for policy purposes. Here the Income and Expenditure Survey (2000) and Labour Force Survey (September 2000) are used to find demographic statistics, income and expenditure patterns, and employment data used for the creation of representative household and factor groups. The current PROVIDE Social Accounting Matrix groups households according to a range of characteristics, first disaggregating by province and race, and thereafter factors such as agricultural employment, geographical location of the household (homeland or non-homeland), gender and education status of the head of the household, and income level of the household come into play. Provincial factor groups are disaggregated by race and skill level/occupation of the factor. Some alternative household and factor groupings, some of which were used in previous versions of the PROVIDE Social Accounting Matrix, are also discussed, as are a range of other characteristics that may be used to form interesting household groups for future Social Accounting Matrices.*

---

<sup>1</sup> The main author of this paper is Kalie Pauw, Senior Researcher of the PROVIDE Project.

**Table of Contents**

<b>1. Introduction .....</b>	<b>4</b>
<b>2. Classification and disaggregation of accounts .....</b>	<b>5</b>
2.1. Guidelines.....	5
2.2. Some examples of Southern African SAMs .....	7
2.2.1. South Africa.....	7
2.2.2. Other Southern African SAMs.....	8
<b>3. Household groups for a South African SAM.....</b>	<b>10</b>
<b>4. Factor groups for a South Africa SAM.....</b>	<b>31</b>
<b>5. Alternative household and factor groupings .....</b>	<b>40</b>
5.1. Household groups for a previous version of the National SAM.....	40
5.2. Household groups for a previous version of the series of provincial SAMs .....	45
5.2.1. Provincial-level household groups .....	45
5.2.2. Provincial-level factor groups .....	55
5.3. Household groups with a geographical- and agricultural focus.....	58
5.3.1. Former homelands areas .....	58
5.3.2. Agricultural households .....	60
5.3.3. Metropolitan Households.....	62
5.3.4. Nodal areas .....	64
<b>6. Concluding remarks.....</b>	<b>65</b>
<b>7. References .....</b>	<b>66</b>
<b>8. Appendix:.....</b>	<b>67</b>
8.1. Hypothesis testing (Hotelling's T <sup>2</sup> -test).....	67

**List of Tables**

Table 1: Household groups in the South African SAM .....	8
Table 2: Household groups in the Tanzania SAM .....	8
Table 3: Household groups in the Malawian SAM.....	9
Table 4: Household groups in the Botswana SAM.....	9
Table 5: Household groups in the Bangladesh SAM.....	10
Table 6: Western Cape household groups – summary statistics .....	22
Table 7: Eastern Cape household groups – summary statistics .....	23
Table 8: Northern Cape household groups – summary statistics.....	24
Table 9: Free State household groups – summary statistics .....	24
Table 10: KwaZulu-Natal household groups – summary statistics .....	25
Table 11: North West household groups – summary statistics.....	27
Table 12: Gauteng household groups – summary statistics .....	28
Table 13: Mpumalanga household groups – summary statistics .....	29
Table 14: Limpopo household groups – summary statistics .....	30
Table 15: Occupation codes .....	31
Table 16: Western Cape factor groups – summary statistics ( <i>inclabp_old</i> ) .....	34
Table 17: Eastern Cape factor groups – summary statistics ( <i>inclabp_old</i> ).....	35
Table 18: Northern Cape factor groups – summary statistics ( <i>inclabp_old</i> ) .....	35
Table 19: Free State factor groups – summary statistics ( <i>inclabp_old</i> ).....	36
Table 20: KwaZulu-Natal factor groups – summary statistics ( <i>inclabp_old</i> ).....	36
Table 21: North West factor groups – summary statistics ( <i>inclabp_old</i> ) .....	37
Table 22: Gauteng factor groups – summary statistics ( <i>inclabp_old</i> ) .....	37
Table 23: Mpumalanga factor groups – summary statistics ( <i>inclabp_old</i> ).....	38

Table 24: Limpopo factor groups – summary statistics ( <i>inclabp_old</i> ) .....	38
Table 25: Urban-rural and racial household composition in South Africa .....	41
Table 26: Average income (adult equivalent) and household size measures by race .....	42
Table 27: Adjusted household size ( <i>E</i> ) by race and location .....	45
Table 28: Number of survey-level households per province, by race and location .....	48
Table 29: Merging urban-rural households within certain race groups .....	48
Table 30: Mean income of households per province, by race and location (weighted).....	49
Table 31: Final sub-groups before splitting by income .....	49
Table 32: Composition of sub-groups, number of observations and suggested number of income groups within each sub-group .....	50
Table 33: Number of wage-earning workers per race, province and occupation type .....	56
Table 34: Workers by race and province (summary table) .....	58

### List of Figures

Figure 1: Western Cape household groups .....	13
Figure 2: Eastern Cape household groups.....	14
Figure 3: Northern Cape household groups .....	15
Figure 4: Free State household groups.....	16
Figure 5: KwaZulu-Natal household groups.....	17
Figure 6: North West household groups .....	18
Figure 7: Gauteng household groups .....	19
Figure 8: Mpumalanga household groups.....	20
Figure 9: Limpopo household groups .....	21
Figure 10: Average wage and number of observations per factor group.....	32
Figure 11: Former homelands in South Africa.....	59
Figure 12: Magisterial districts and homeland areas in KwaZulu-Natal .....	60
Figure 13: Municipalities/Cities and population size (number of households) .....	64
Figure 14: Nodal areas .....	65

## 1. Introduction

The household accounts of Social Accounting Matrices (SAMs) capture all economic transactions between households and other agents in the economy (see PROVIDE, 2003 for a detailed description of a SAM and the SAM approach to modelling). A single household account in a macro-SAM can be disaggregated into many sub-accounts depending on the requirements of the modellers. Traditionally Computable General Equilibrium (CGE) models have followed the representative household group (RHG) approach, which means that each household group, which is represented by an account in the SAM, is treated as a single entity or actor in the economy. The assumption is that each individual household that forms part of a specific RHG is affected in the same average manner by a policy shock. When forming household groups it is important to try and form groups of households that have similar preferences and characteristics. This ensures that the assumption underlying the RHG approach is as realistic as possible, since similar households are more likely to be affected similarly by economic shocks in reality.

In terms of income distribution analysis the RHG approach has its drawbacks. Although average incomes can be compared across RHGs, giving some indication of the between-group inequality, the evaluation of changes in intra-household group income distributions is not possible. Bourguignon *et al.* (2002) maintain that between group distributions add as much to overall inequality as within-group distributions, and therefore it is important to at least have some level of understanding of how within-group income distributions may be affected by policy shocks.<sup>2</sup> In reality it is very likely that individual households within RHGs are affected in different ways. Economic policy shocks can affect a whole range of economic variables, including relative prices of commodities, wages or employment levels of different types of labour, expansion or recession of different industries or savings- and investment plans of institutions. Furthermore, not only is each RHG made up of individual households, but each individual household is also made up of a number of individuals. Thus, even household members within households may be affected in different ways by economic shocks. The logic therefore is looking at the average impact on a group of households, not the individual impact.

Various researchers have attempted to overcome the limitations of the RHG approach. For a given sample, if more household groups are formed, the groups are likely to become more homogenous. In the extreme case each individual household can

---

<sup>2</sup> Grouping households with similar income levels can of course reduce the within-group inequality.

have a separate SAM account. Models based on such SAMs have been developed, with success, in the past (see Cockburn, 2001).<sup>3</sup> However, this approach becomes infeasible when the underlying survey data contains large numbers of observations as this dramatically increases the time it takes to solve the CGE model. Given powerful modern-day computers Bourguignon *et al.* (2002:3) feel that the issue is not so much computational, but rather the “*observability of heterogeneous factors or preferences*” at the individual household level. It simply becomes too onerous to analyse results for hundreds of households. Perhaps a greater concern is data reliability. Cockburn’s approach requires good quality data since each individual observation is used on its own rather than grouping households and using averages. In many instances average income and expenditure data are more reliable, especially when the data contains many outliers. The real problem is finding the ideal number of household groups given the quality of the data.

The aim of this paper is to look into the process of disaggregating SAM accounts, focusing specifically on the household accounts and to a lesser extent on the factor accounts. Section 2 looks at some of the guidelines that can be followed when forming SAM accounts in general. These general guidelines can be applied to the formation of household groups as well. This section also presents the household groupings of various southern African SAMs as examples of possible disaggregations. Section 8 then explains how the household groups of the current version of the PROVIDE South African National SAM were formed by disaggregating households by race, location and income group. Section 5.1 extends the national-level disaggregation to the provincial level. These provincial-level household groups may be used in provincial SAMs, regional SAMs or an integrated South African SAM. Finally, section 6 concludes the discussion and points out some key issues for future research.

## **2. Classification and disaggregation of accounts**

### **2.1. Guidelines**

The classification and disaggregation of accounts is the most important step in the development of a SAM. Decaluwé *et al.* (1999) highlights two key issues in deciding on the SAM classification scheme:

- The level and extent of disaggregation has to be decided on. Although, in many instances, a fairly aggregated SAM is sufficient, Decaluwé *et al* suggests it is

---

<sup>3</sup> Technically speaking Cockburn’s approach is still that of a RHG, since the household information is typically drawn from a household survey rather than a census, and hence each household in the

always possible to “*consolidate and aggregate sub-accounts – but not the other way around* – [and therefore] *it may be better to start at a level of aggregation which is as detailed as data reliability allows*” (1999:10).

- Especially in the case of household and factor accounts, homogeneity is important. The classification of RHGs requires that households with similar income sources and expenditure levels (for example) be grouped together.

Rivero *et al.* (1986, cited in Decaluwé *et al.*, 1999) argue that certain requirements should be met if a certain classification is to be used in a SAM. Adapting these conditions to household groups in particular, he argues that the classification should

- correctly reproduce the socio-economic stratification within the society and the economy;
- distinguish relatively homogenous household groups and categories;
- be composed of socio-economic groups that are recognisable for policy purposes, i.e. they should be distinguishable as target groups for policy experiments;
- be based on comparatively stable characteristics that are reliable and easily measured; and
- be derivable from (a combination of) existing data sources.

Decaluwé *et al.* (1999) conclude that there is no unique or standard classification scheme. The level and extent of disaggregation should depend on the country or region on which that SAM is based, as well as (importantly) the objectives of the studies for which the SAM is being developed. Therefore, it is not surprising to find a vast array of classification schemes that have been employed all over the world. Household disaggregation is usually based on one or a combination of the following: *household location* (rural or urban), *asset ownership* (land or capital), or *various characteristics of the head of the household* (employment status, industry of employment, educational attainment, gender, language, race etc.). A brief survey of some of the classification schemes used recently in southern African countries is provided below.

---

survey already represents a segment of the population. This approach also places a bigger burden on the survey.

## 2.2. Some examples of Southern African SAMs

### 2.2.1. *South Africa*

A 1998 South African SAM compiled by Thurlow and van Seventer (2002) follows the 1997 SAM used by Lewis (2001) in the formation of household groups by simply forming fourteen groups based on household income alone. It is unclear whether household income was adjusted to take into account differences in size and composition or structure (relative number of children under the age of ten) of households. Households are first divided into income deciles (ten groups) and then the top income decile is further disaggregated into five smaller groups to provide a detailed breakdown at this end of the income distribution.

The authors admit that this simple disaggregation is perhaps insufficient as “*policy makers are often interested in a richer household picture*” (Thurlow and Van Seventer, 2002:42). This is especially true in a developing country like South Africa where current development issues are complex. Given the country’s history of discrimination a racial breakdown of households may be required to enable the analysis of, for example, redistribution policies. Inequality between provinces and rural/urban households may require a breakdown by location or province. Other possible classification criteria for households include the education, skill level or gender of the head of the household, as some of these factors are key in determining the socio-economic circumstances of households in South Africa.

In reality this SAM will probably fail to serve as a useful database for a CGE model if the modeller wishes to address any of the issues that are currently high on the agenda of South African policymakers. In fact, as argued by Decaluwé *et al.* (1999:10), a “*household classification based on income or expenditure brackets does not satisfy any of [the] requirements*” listed by Rivero (1986). The poorest segment of the society can quite possibly include a household head classified as a landless agricultural worker and an urban informal sector worker. Policies aimed at improving these two households’ conditions are likely to be very different.

Another example of a South African SAM is the regional Western Cape SAM developed by McDonald and Punt (2001), which contains thirty household groups created by first disaggregating households by race, then location and finally income. Urban Coloured households are divided into “*triciles*” (three groups), while rural Coloured households are split into “*duociles*” (two groups). Asian households are also split into “*duociles*”. Finally, urban White households are disaggregated into quartiles (four groups), while rural White households are divided into “*duociles*”. The top

income group is split into two further groups. This further disaggregation is necessary because of the high degree of inequality within high-income groups. It is often the case that high-income households earn a large proportion of national income, and hence have different preferences even compared to other relatively well-off households.

The fact that household income is used directly without adjusting for the household size and structure is a shortcoming, especially if account is taken of the large differences in household size between various socio-economic sub-groups in the South African economy. Table 1 shows the thirty household accounts used in this Western Cape SAM.

Table 1: Household groups in the South African SAM

African		Coloured		Asian	White	
Urban	Rural	Urban	Rural	All	Urban	Rural
Quintile 1	Quintile 1	Tricile 1	Duocile 1	Duocile 1	Quartile 1	Duocile 1
Quintile 2	Quintile 2	Tricile 2	Duocile 2a	Duocile 2a	Quartile 2	Duocile 2a
Quintile 3	Quintile 3	Tricile 3a	Duocile 2b	Duocile 2b	Quartile 3	Duocile 2b
Quintile 4	Quintile 4	Tricile 3b	-	-	Quartile 4a	-
Quintile 5a	Quintile 5a	-	-	-	Quartile 4b	-
Quintile 6a	Quintile 6a	-	-	-	-	-

Source: McDonald and Punt (2001)

### 2.2.2. Other Southern African SAMs

More recently Thurlow and Wobst (2003) developed a poverty-focused SAM for **Tanzania**, which, due to the intended use of the SAM, gives more attention to the household classification. Households are firstly separated into rural and urban households. The remaining disaggregation is based on the adult-equivalent income level of the household and the education of the head of the household. Official poverty lines published in the Tanzanian Household Budget Survey for 2000/1 (HBS) are also used in the disaggregation. The HBS listed household location and education of the head as two factors that account the most for the incidence of household poverty, hence the reason for including these factors in the classification. Table 2 shows the formation of the twelve household groups in the Tanzanian SAM.

Table 2: Household groups in the Tanzania SAM

Rural	Urban
Below food poverty line	Below food poverty line
Between food and basic needs poverty lines	Between food and basic needs poverty lines
Non-poor – head with no education	Non-poor – head with no education
Non-poor – head not finished primary school	Non-poor – head not finished primary school
Non-poor – head not finished secondary school	Non-poor – head not finished secondary school
Non-poor – head finished secondary school	Non-poor – head finished secondary school

Source: Thurlow and Wobst (2003)

The 1998 SAM for **Malawi** (Chulu and Wobst, 2001) follows a slightly different approach. Due to the important role that agriculture plays in rural areas, land ownership is an important determinant of wealth in these areas. For households not involved in agriculture the level of education of the head of the household plays an important role in determining the head's employability and hence social class. Data is drawn from the Integrated Household Survey (IHS) of 1997/8. The fourteen households in the Malawian SAM are shown in Table 3.

Table 3: Household groups in the Malawian SAM

Rural		Urban	
Agriculture	Non-agriculture	Agriculture	Non-agriculture
Less than 0.5ha land	No education	Urban agriculture	No education
0.5ha to 1.0 ha land	Low education	-	Low education
1.0ha to 2.0 ha land	Medium education	-	Medium education
2.0ha to 5.0 ha land	High education	-	High education
More than 5ha land	-	-	-

Source: Chulu and Wobst (2001)

The 1991 **Zimbabwe** SAM (Thomas and Bautista, 1999) classifies households based more or less on the national Income and Consumption Expenditure Survey (ICES) classification for households. The classification takes into account household activities (farming or commercial), land ownership and location. The groups are (1) large-scale commercial farmers (owners/managers, high income), (2) large scale commercial farm workers (typically low income), (3) smallholders (a combination of smallholders, communal households and resettlement households), (4) urban high-income households and (5) urban low-income households.

The household account of the 1994/5 **Mozambique** SAM compiled by Arndt *et al.* (1998) only comprises of two households: rural and urban. More information will only become available once the 1997 household survey data is released. Finally, a SAM for **Botswana** (see McDonald, 2002) also follows the urban-rural split, but then divides households according to their main sources of income. Non-citizens are also included as a separate household group. The seven household groups are shown in Table 4.

Table 4: Household groups in the Botswana SAM

Rural	Urban
Wage income	Wage income
Self-employed	Self-employed
Transfers	Transfers
Non-citizen households (not location specific)	

Source: McDonald (2002)

Household account disaggregation in southern African countries clearly places a high degree of importance on the rural-urban split. However, each individual country

should be evaluated separately when deciding on the classification. A good example is the racial split in the South African SAM, which is necessary due to the history of the country and subsequent policies that aim to redress past racial inequalities. Another good (non-African) example is the *gendered* 1993/4 SAM for **Bangladesh**, which was developed by Fontana and Wobst (2001). Households are split in a similar way as in the Malawian SAM in terms of land ownership and education of the head of the household, but the gender of the head of the household also comes into play. Table 5 shows how this was done.

Table 5: Household groups in the Bangladesh SAM

Group	Description
1. Agricultural landless	Owns no land
2. Agricultural marginal	Owns up to 0.49 acres
3. Agricultural small	Owns between 0.5 and 2.49 acres
4. Agricultural large	Owns more than 2.49 acres
5. Non-agricultural poor female-headed	Female-headed household, owns less than 0.5 acres but not involved in agriculture
6. Non-agricultural poor male-headed	Male-headed household, owns less than 0.5 acres but not involved in agriculture
7. Non-agricultural rich female-headed	Female-headed household, owns more than 0.5 acres but not involved in agriculture
8. Non-agricultural rich male-headed	Male-headed household, owns more than 0.5 acres but not involved in agriculture
9. Urban illiterate	Head has no schooling
10. Urban low educated	Head's education level is I-IV (school levels)
11. Urban medium educated	Head's education level is VI-VIII or IX-X
12. Urban highly educated	Head's education level is graduate or above

Source: Fontana and Wobst (2001)

### 3. Household groups for a South African SAM

The household accounts of the PROVIDE SAM are disaggregated by province, race, farming/non-farming, homelands/non-homelands, gender of the head of the household, education level of the head of the household, and income of the household. The four regions that have been identified for the purpose of regional analyses within the PROVIDE Project are made up of between two and three provinces each. These are the West Coast (Western Cape, Eastern Cape), East Coast (Eastern Cape, KwaZulu-Natal), Central (Free State, North West, Gauteng) and Border (Mpumalanga, Limpopo) regions. Given this regional focus a distinction along provincial lines as a first 'cut' is necessary. Furthermore, since the Project operates within the various National and Provincial Departments of Agriculture, an agricultural household distinction is highly relevant for certain provinces. The Income and Expenditure Survey of 2000 (IES 2000) (SSA, 2002a) was used throughout as the source of income data, while most of the demographic data was sourced from the Labour Force Survey of September 2000 (LFS

2000:2) (SSA, 2002b). The latter is merged with the IES 2000 to form a combined dataset with comprehensive data on income and expenditures, demographics and employment information for households and household members (see PROVIDE, 2005b for more).

The racial group and homeland/non-homeland distinction is driven by the history of South Africa. There are large differences in income levels and sources of income, expenditure patterns, and other characteristics between households of different racial groups. African households living in homeland areas are also typically more impoverished and isolated from the formal economy than non-homeland African households, hence this distinction in certain provinces. A further consequence of South Africa's past is the presence of 'fractured families'. The country has a very large share of female-headed households who are left to run the household while their husbands search for work on mines and in the cities. This has had important social implications, especially in rural areas, with husbands never returning or failing to fulfil their commitments in terms of supporting their families at home.

Education levels capture a skill dimension and improve the relationship between factor and household accounts.<sup>4</sup> Evidence in South Africa suggests a high correlation between education levels and employment status (Bhorat and Leibbrandt, 1996), and also between unemployment and poverty (see for example May, 1998). Only very large groups were further split into low-income and high-income groups, with the cut-off point around the median of income within the larger group.

A total of 162 household groups were formed. Obviously not all provinces or races can be disaggregated fully using these criteria due to limited numbers of observations for some of the criteria. For example, there are no homelands in the Western Cape, and too few farming households in the Northern Cape to justify having separate household groups for these. In most provinces Coloured and Asian households are also grouped together due to limited number of either or both of these households.<sup>5</sup>

---

<sup>4</sup> Education levels are (1) none- or pre-primary, (2) primary, (3) lower secondary (or grade 10), (4) upper secondary (or grade 12), (5) tertiary, and (6) don't know or missing value.

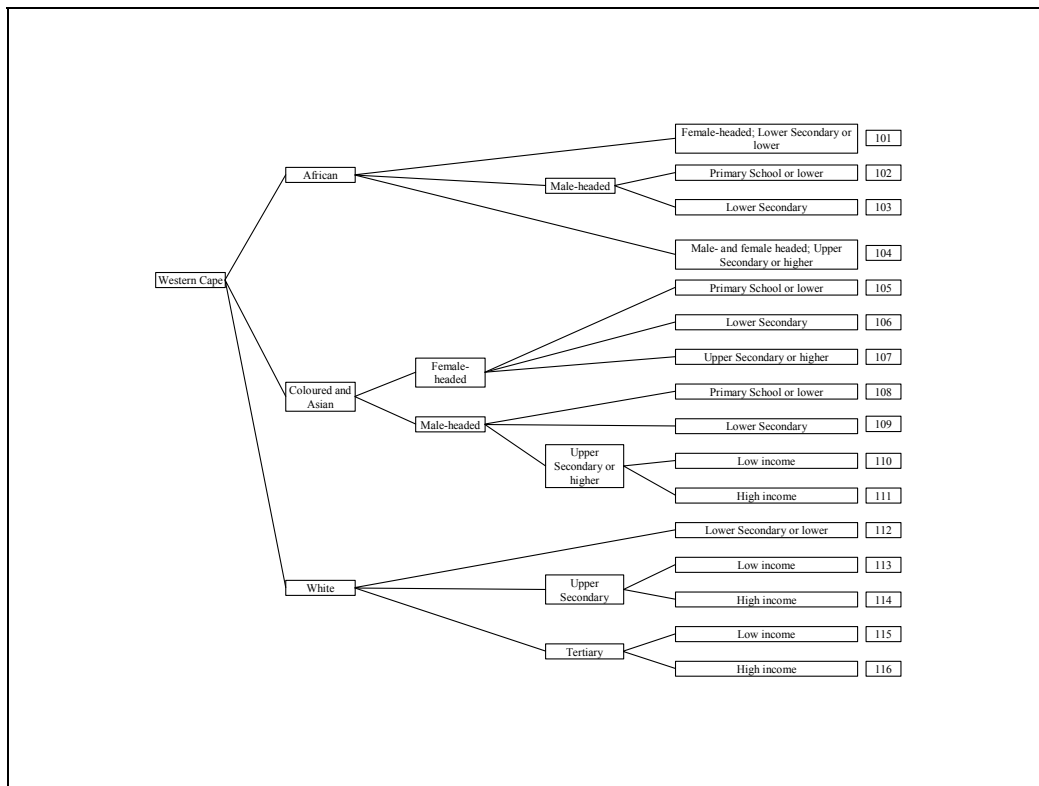
<sup>5</sup> In many of South Africa's provinces the limited numbers of Asian households makes it necessary to group Asian households with one of the other race groups. In section 5.2.1 it is argued that Asian and White households are probably more closely matched in terms of expenditure patterns and income levels, which in economic terms define to some extent their behavioural characteristics. However, politically speaking Asian households are classified as 'previously disadvantaged', and as such their present day circumstances have arguably been driven by their history of closer association with Coloured households. While there are arguments for both approaches to grouping racial groups, the racial classification here is driven by political considerations, and hence Asian and Coloured households are grouped rather than Asian and White households.

As a starting point the number of observations as well as the mean, standard deviation, inter-quartile range, skewness and kurtosis of total household income were listed for each sub-group. This data was examined to decide on natural aggregations within each province, given certain minimum requirements for the number of observations per household group, and similarities (or dissimilarities) in the distribution of income. In some cases household groups became fairly large during the aggregation process. All the very large household groups with 50,000 or more weighted observations were split into lower and upper income groups around the median household income of the group.

Below a brief description of the household groups of each province is given, together with a graphic showing clearly how the groups are made up. The numbers in the boxes on the right-hand side of each household group refers to the code (variable *newrhg*) given to each household group. The tables referenced contain more detailed summary statistics on the number of weighted and unweighted observations and the distribution of income within each group. Refer to footnote 4 for an interpretation of the education codes used in the tables of summary statistics.

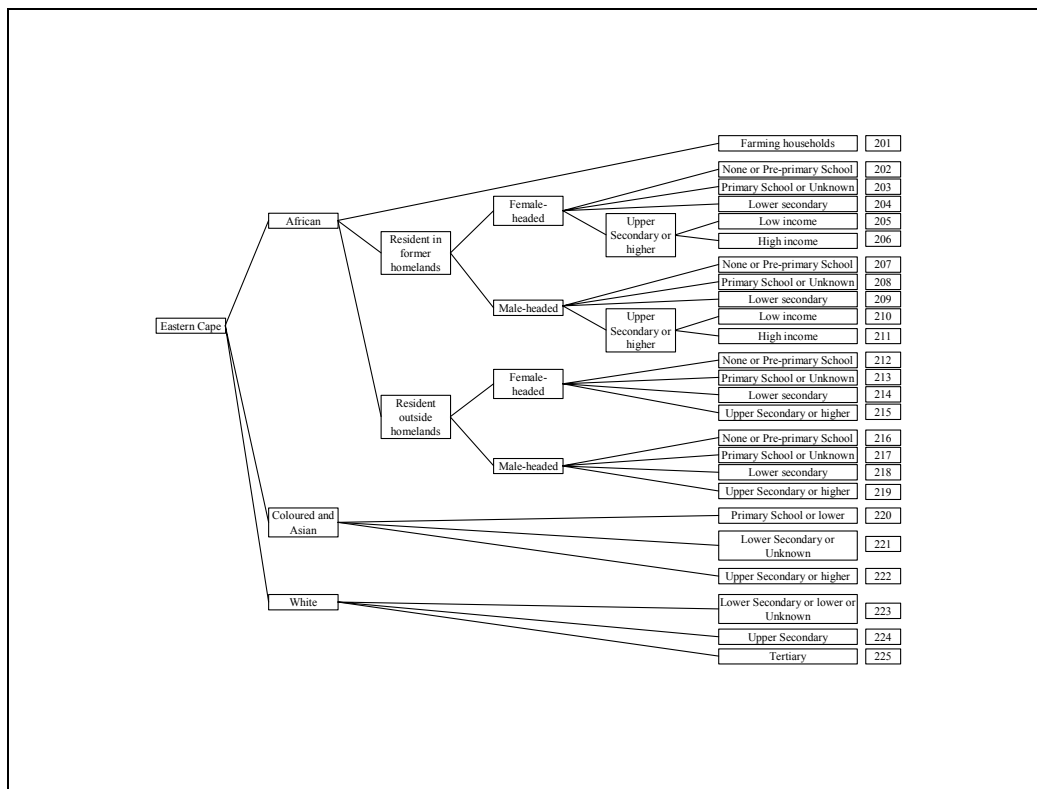
The Western Cape has 16 household groups (see Table 6). Of these groups four are African, eight are Coloured/Asian and five are White. Over 50% of the households in the province are Coloured, but since there are very few Asian households these two race groups were merged.

Figure 1: Western Cape household groups



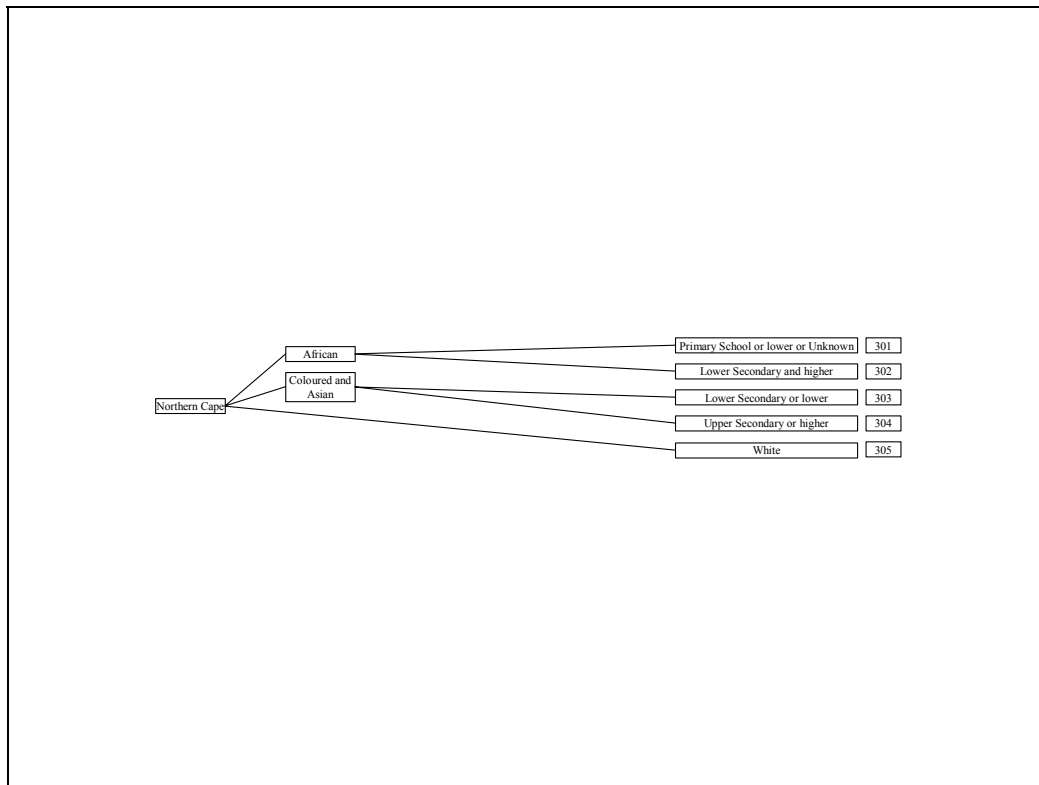
The slightly larger Eastern Cape has 25 household groups (see Table 7). Over 86% of the households in this province are African, and hence this race group is highly disaggregated. The Eastern Cape is also home to the largest former homeland, namely Transkei, while the former Ciskei also falls within its boundaries (see section 5.3.1). All African farming households are kept as a single household group. Selection into this group depends on the occupation of the head of the households, i.e. if the head is a skilled agricultural worker the household qualifies as an agricultural household. The remainder of the African households are non-farming households, and are disaggregated first into homeland and non-homeland households, and thereafter by gender of the head of the household. This distinction is also important, as about half of the African households are female-headed. Since there are relatively few non-African households, the remaining six household groups are made up of Coloured/Asian (three) and White (three) households.

Figure 2: Eastern Cape household groups



The Northern Cape is one of the smallest provinces in South Africa as measured by its population size and contains only five household groups (see Table 8). Coloured households make up the largest group, followed by Africans. The Coloured/Asian and African sub-groups are both disaggregated into two education groups, while White households are grouped together as a single household group.

Figure 3: Northern Cape household groups



The vast majority (84%) of households in the Free State (see Table 9) are African. As in the Eastern Cape the male-/female-headed distinction is important in the Free State, with almost 40% of African households headed by females. There are six African female-headed and six African male-headed households in total, while the remainder of the households are grouped into Coloured/Asian (one group) and White households (three education groups).

Figure 4: Free State household groups

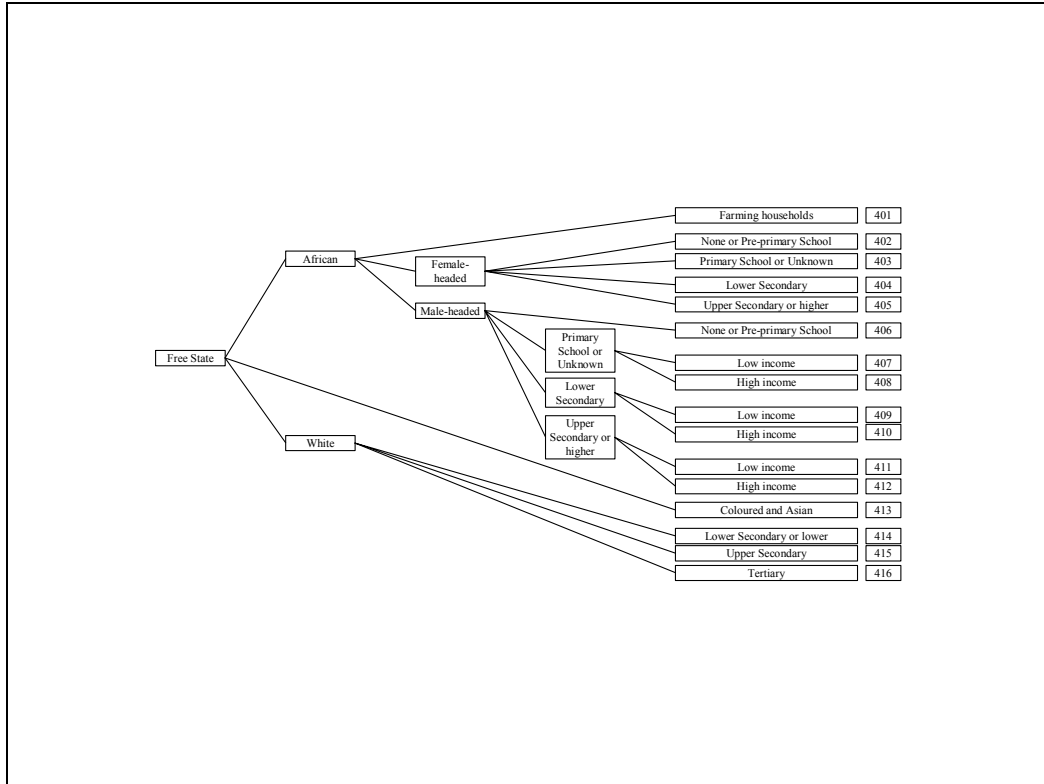
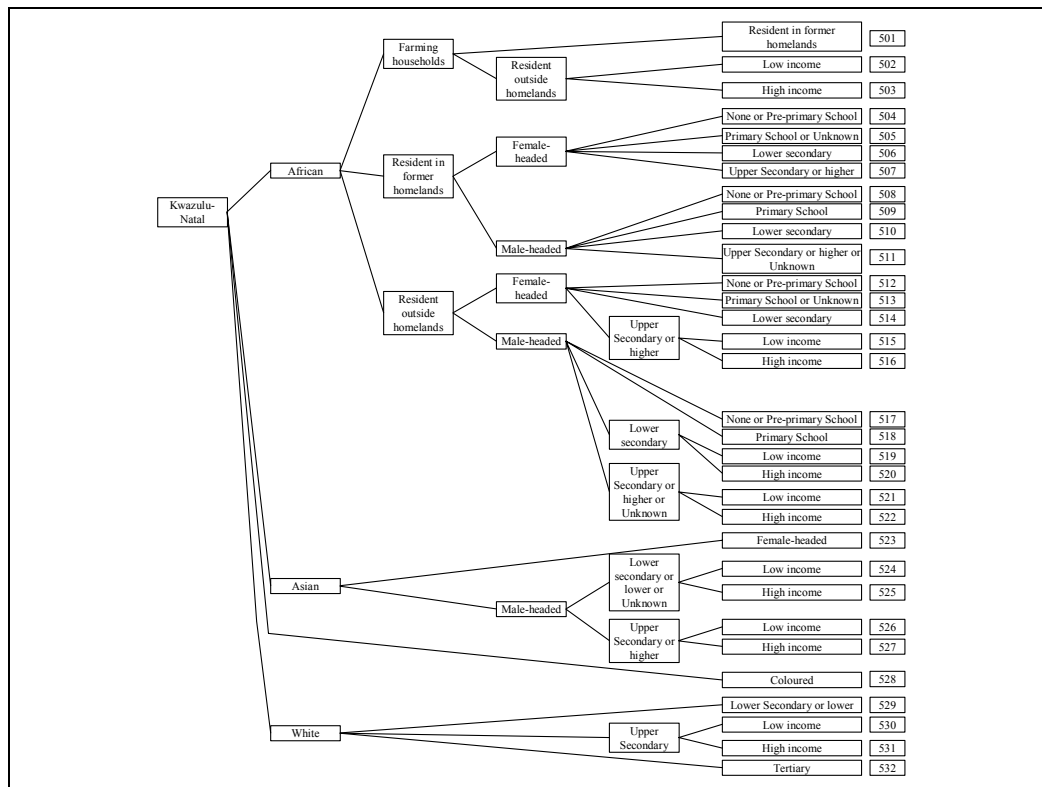


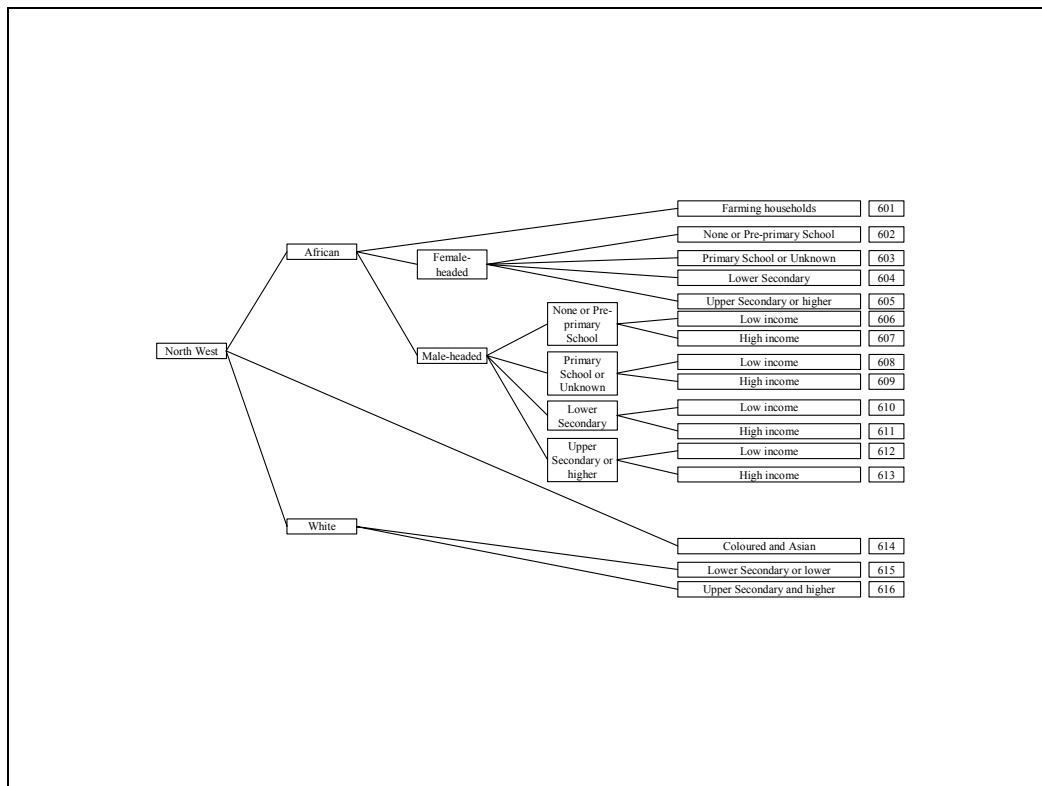
Table 10 contains the summary statistics for the 32 household groups in KwaZulu-Natal. Almost one in five South African households live in this province, making it the second largest of all the provinces. Large areas of KwaZulu-Natal previously fell within homelands areas, with a quarter of households still residing within these areas. Since agriculture forms an important livelihood strategy for African households, both within and outside the former homelands areas, three agricultural household groups are formed. Furthermore, about half the African households, both within and outside the former homelands areas, are female-headed. KwaZulu-Natal is home to over 73% of the South African Asian households. This justifies having four separate racial household groups in this province, with a fairly detailed disaggregation of African, and to a lesser extent, Asian households. There are four White household groups and only a single Coloured group.

Figure 5: KwaZulu-Natal household groups



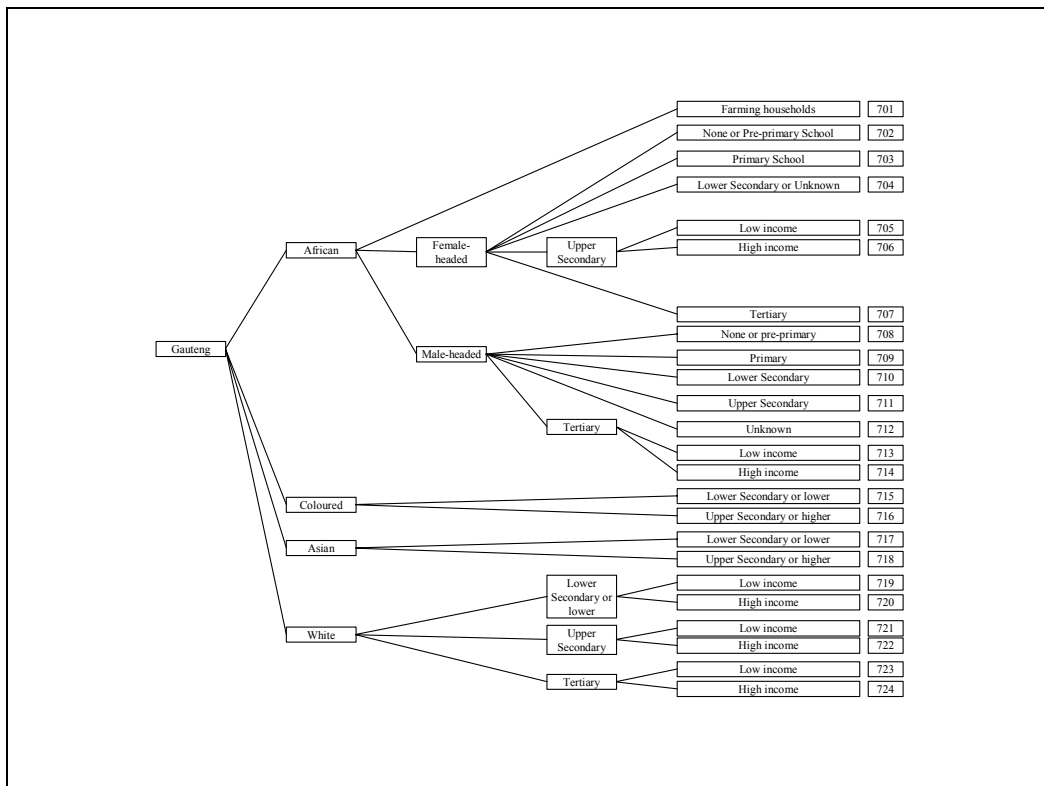
Over 92% of the almost 800,000 households in the North West province are African (see Table 11). African households are disaggregated into farming households and male-/female-headed households, giving a total of 13 household groups. Roughly 37% of the African households are female-headed. Coloured and Asian household combined make up just over 1% of the population and are represented by a single household group, while there are two White household groups. This gives a total of 16 household groups.

Figure 6: North West household groups



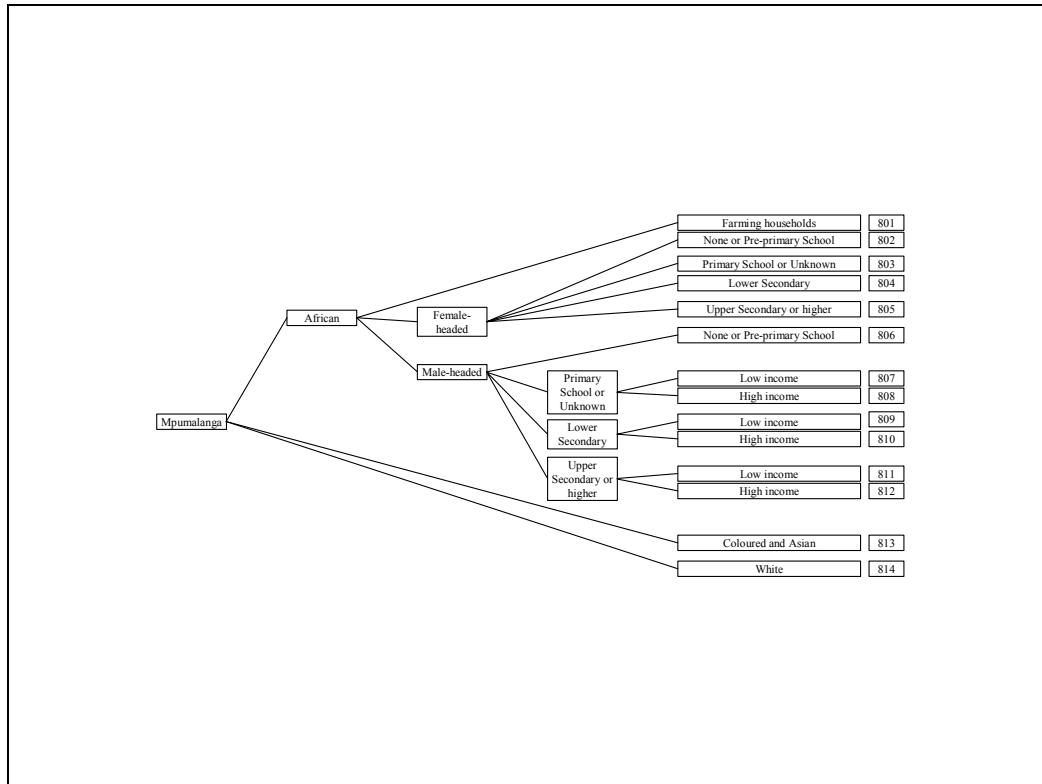
Gauteng is the largest of the South Africa provinces, with almost 28% of households living in this province (see Table 12). All racial groups are well represented and are included separately. Over 78% of the households are African, and hence 14 of 24 the household groups are African. These are disaggregated into farming, female- and male-headed households. Asian and Coloured households each have two household groups, while White households have six, reflecting the fact that over 40% of White households in South Africa live in Gauteng.

Figure 7: Gauteng household groups



Mpumalanga and Limpopo have very similar profiles and hence their household groupings are also quite similar. Both have 12 African household groups, and a single household group each for Coloured/Asian and White households, giving a total of 14 households. Mpumalanga is the smaller of the two and is home to just under 650,000 households, 92% of which are African. These households are disaggregated further into farming, female- and male-headed households (see Table 13).

Figure 8: Mpumalanga household groups



With just over one million households, the Limpopo province is roughly similar in size to the Western Cape. Almost 98% of the households are African. Table 14 lists the summary statistics for this province.

Figure 9: Limpopo household groups

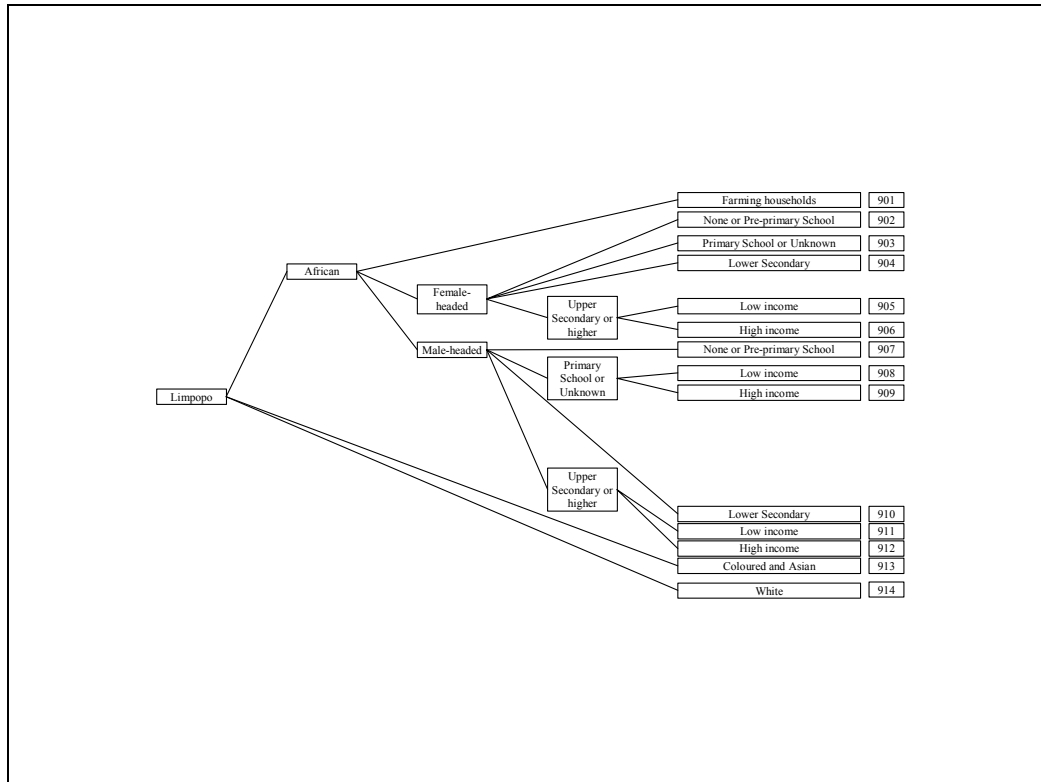


Table 6 to Table 14 can all be interpreted in the following way. The first column contains the code for each RHG. Columns three and four contain the number of observations, first at sample level (‘unweighted’) and then at the population level (weighted). The IES 2000 ‘household weights’ were used. Variable *totinc* is the total household income. The rest of the columns in the table report the mean and standard deviation (weighted) of *totinc*, as well as the interquartile range, the minimum, the 10<sup>th</sup> percentile (P10), median, 90<sup>th</sup> percentile (P90) and maximum of *totinc*.

Table 6: Western Cape household groups – summary statistics

Western Cape	Description	Observations (unweighted)	Observations (weighted)	Mean (totinc)	St.dev. (totinc)	Interquartile range	Min (totinc)	P10 (totinc)	Median (totinc)	P90 (totinc)	Max (totinc)
101	African, Female, Lower Secondary and lower	156	72,591	20,938	23,850	17,400	1,903	6,300	15,644	40,727	240,419
102	African, Male, Primary and lower	208	83,463	20,908	14,191	14,766	977	6,480	17,386	38,640	93,056
103	African, Male, Lower Secondary	110	47,301	24,747	18,081	16,560	2,464	6,480	20,315	44,720	99,527
104	African, Upper Secondary and higher	123	50,602	71,112	193,631	44,954	2,994	12,000	38,829	115,094	2,105,466
105	Asian & Coloured, Female, Primary and lower	256	90,036	28,233	20,371	24,240	2,700	8,064	24,154	53,127	113,818
106	Asian & Coloured, Female, Lower Secondary	132	48,972	38,385	39,249	27,516	3,785	11,060	29,068	70,905	324,000
107	Asian & Coloured, Female, Upper Secondary and higher	66	28,559	68,287	47,882	57,711	3,671	20,368	50,684	139,086	240,951
108	Asian & Coloured, Male, Primary and lower	483	137,238	34,763	39,379	24,262	3,579	10,995	24,510	62,199	330,620
109	Asian & Coloured, Male, Lower Secondary	360	140,361	54,998	45,684	52,041	1,407	13,650	44,071	103,272	301,700
110	Asian & Coloured, Male, Upper Secondary and higher, Low-income (split)	122	51,367	47,734	22,036	37,810	5,819	16,096	49,897	76,408	86,064
111	Asian & Coloured, Male, Upper Secondary and higher, High-income (split)	123	61,645	165,366	71,969	94,195	86,400	100,000	146,020	248,060	413,490
112	White, Lower Secondary and lower	83	46,612	68,427	68,138	60,080	1,800	14,407	40,640	159,794	302,069
113	White, Upper Secondary, Low-income (split)	106	56,236	62,409	26,994	41,298	11,428	25,358	62,111	98,160	118,978
114	White, Upper Secondary, High-income (split)	107	62,823	219,316	115,676	106,975	119,880	125,096	185,708	349,800	793,428
115	White, Tertiary, Low-income (split)	53	38,302	113,025	36,328	36,958	32,239	52,177	119,936	163,247	173,597
116	White, Tertiary, High-income (split)	53	38,961	389,154	239,821	236,032	176,000	199,196	286,000	739,090	1,200,000

Table 7: Eastern Cape household groups – summary statistics

Eastern Cape	Description	Observations (unweighted)	Observations (weighted)	Mean (totinc)	St.dev. (totinc)	Interquartile range	Min (totinc)	P10 (totinc)	Median (totinc)	P90 (totinc)	Max (totinc)
201	African, Agricultural	125	53,134	17,997	20,584	17,470	398	2,076	11,991	52,292	96,961
202	African, Homeland, Female, None	301	136,190	9,440	10,765	3,959	489	3,397	7,117	14,480	129,997
203	African, Homeland, Female, Primary	450	211,411	10,638	10,694	5,450	286	3,435	7,951	18,386	86,581
204	African, Homeland, Female, Lower Secondary	248	113,211	12,665	17,791	7,149	469	3,616	8,042	21,988	218,664
205	African, Homeland, Female, Upper Secondary and higher, Low-incom	94	39,736	10,385	5,897	8,560	2,066	3,635	8,500	19,748	25,003
206	African, Homeland, Female, Upper Secondary and higher, High-inco	96	34,774	71,537	52,808	46,345	25,308	30,800	55,172	119,529	399,446
207	African, Homeland, Male, None	189	87,889	11,938	10,262	8,283	360	2,052	9,466	20,154	67,864
208	African, Homeland, Male, Primary	344	158,845	13,559	17,471	10,555	754	2,352	9,128	28,846	190,230
209	African, Homeland, Male, Lower Secondary	191	82,542	18,793	30,203	14,089	541	3,435	9,240	38,762	261,768
210	African, Homeland, Male, Upper Secondary and higher, Low-income	72	29,317	22,750	16,541	29,389	572	4,580	15,094	48,610	51,500
211	African, Homeland, Male, Upper Secondary and higher, High-income	74	28,128	106,812	73,924	53,731	51,851	57,305	80,897	178,931	467,587
212	African, Non-Homeland, Female, None	56	19,438	10,664	10,178	9,360	1,026	1,536	8,181	19,966	51,384
213	African, Non-Homeland, Female, Primary	113	37,084	14,957	16,247	10,429	1,041	4,195	9,799	31,568	141,560
214	African, Non-Homeland, Female, Lower Secondary	97	31,292	15,178	12,131	14,136	1,680	3,970	11,404	31,200	71,000
215	African, Non-Homeland, Female, Upper Secondary and higher	64	21,532	56,058	46,370	73,840	2,381	8,208	48,300	105,993	268,410
216	African, Non-Homeland, Male, None	53	19,128	14,773	11,807	12,091	670	4,386	11,673	27,275	56,215
217	African, Non-Homeland, Male, Primary	176	64,730	17,437	16,798	14,931	703	4,320	12,208	36,800	143,656
218	African, Non-Homeland, Male, Lower Secondary	135	44,694	25,674	22,922	24,697	1,015	3,840	20,212	51,756	126,832
219	African, Non-Homeland, Male, Upper Secondary and higher	96	34,428	68,693	81,761	83,674	1,609	9,529	45,501	137,619	473,041
220	Asian & Coloured, Primary and lower	147	52,823	22,775	33,240	17,743	922	6,480	12,699	46,473	302,526
221	Asian & Coloured, Lower Secondary	104	34,675	36,079	41,990	28,654	3,640	6,818	24,000	96,045	256,894
222	Asian & Coloured, Upper Secondary and higher	49	17,821	124,036	100,051	136,321	5,320	21,450	99,686	250,838	492,545
223	White, Lower Secondary and lower	53	21,232	88,448	76,754	71,220	6,240	18,026	64,255	229,458	342,000
224	White, Upper Secondary	89	39,060	149,007	112,366	115,371	28,876	54,648	112,855	282,600	569,675
225	White, Tertiary	50	27,300	212,387	155,433	223,980	36,940	55,452	166,100	421,872	822,741

Table 8: Northern Cape household groups – summary statistics

Northern Cape	Description	Observations (unweighted)	Observations (weighted)	Mean (totinc)	St.dev. (totinc)	Interquartile range	Min (totinc)	P10 (totinc)	Median (totinc)	P90 (totinc)	Max (totinc)
301	African, Primary and lower	310	41,992	17,403	18,570	12,240	586	4,850	10,916	37,800	133,883
302	African, Lower Secondary and higher	187	26,377	37,812	67,012	33,098	1,927	5,380	16,344	77,572	626,419
303	Coloured & Asian, Lower Secondary and lower	382	52,208	19,535	30,313	12,427	915	4,974	11,245	36,541	310,440
304	Coloured & Asian, Upper Secondary and higher	240	34,155	53,887	72,385	56,900	1,234	7,089	25,246	121,080	451,048
305	White	191	32,515	195,770	362,297	130,736	6,480	33,202	106,000	390,874	3,480,000

Table 9: Free State household groups – summary statistics

Free State	Description	Observations (unweighted)	Observations (weighted)	Mean (totinc)	St.dev. (totinc)	Interquartile range	Min (totinc)	P10 (totinc)	Median (totinc)	P90 (totinc)	Max (totinc)
401	African, Agricultural	86	24,713	11,096	10,483	7,974	1,128	2,171	8,124	23,610	66,515
402	African, Female, None	144	42,950	10,564	15,644	4,758	496	4,270	7,621	16,251	192,000
403	African, Female, Primary	333	98,359	12,028	15,261	9,206	1,022	3,000	8,400	22,772	205,419
404	African, Female, Lower Secondary	165	49,980	17,792	22,889	16,913	1,590	3,600	10,400	34,472	149,697
405	African, Female, Upper Secondary and higher	98	31,175	40,810	40,807	54,275	1,212	4,971	24,573	92,000	257,045
406	African, Male, None	193	52,274	18,573	22,028	17,290	555	3,600	12,960	39,389	203,848
407	African, Male, Primary, Low-income (split)	267	69,869	7,719	3,826	6,263	825	2,960	7,353	13,224	15,280
408	African, Male, Primary, High-income (split)	267	71,179	37,798	23,406	19,553	15,354	17,846	32,576	61,257	264,400
409	African, Male, Lower Secondary, Low-income (split)	151	45,438	8,214	4,858	9,208	600	2,590	7,200	15,600	18,194
410	African, Male, Lower Secondary, High-income (split)	151	37,932	45,901	32,307	15,892	18,498	22,800	38,121	71,858	294,195
411	African, Male, Upper Secondary and higher, Low-income (split)	107	31,406	11,734	8,373	11,694	771	2,598	9,559	24,069	31,475
412	African, Male, Upper Secondary and higher, High-income (split)	108	33,799	94,361	72,796	66,231	31,798	37,172	72,544	180,000	535,285
413	Asian & Coloured	47	15,103	42,137	49,464	48,876	3,435	4,186	24,900	90,943	260,768
414	White, Lower Secondary and lower	57	27,905	98,212	141,163	65,386	6,240	23,267	59,898	157,820	1,096,539
415	White, Upper Secondary	92	42,034	178,817	420,288	102,399	12,480	41,075	90,895	254,000	4,997,943
416	White, Tertiary	50	24,131	223,039	162,968	166,970	39,665	57,714	195,442	398,793	847,333

Table 10: KwaZulu-Natal household groups – summary statistics

KwaZulu-Natal	Description	Observations (unweighted)	Observations (weighted)	Mean (totinc)	St.dev. (totinc)	Interquartile range	Min (totinc)	P10 (totinc)	Median (totinc)	P90 (totinc)	Max (totinc)
501	African, Agricultural, Homeland	71	29,396	27,004	68,688	13,911	893	4,657	11,152	40,086	728,003
502	African, Agricultural, Non-Homeland, Low-income (split)	55	22,467	7,633	2,905	3,875	1,628	3,240	7,735	11,624	12,081
503	African, Agricultural, Non-Homeland, High-income (split)	55	27,203	27,490	20,182	15,715	12,090	13,176	21,697	49,565	137,340
504	African, Homeland, Female, None	273	100,183	10,097	6,729	6,384	914	4,277	7,740	19,690	62,700
505	African, Homeland, Female, Primary	210	86,627	12,178	10,543	8,698	1,223	3,829	9,526	21,861	70,864
506	African, Homeland, Female, Lower Secondary	77	36,240	13,691	9,807	10,318	2,132	5,590	10,498	24,305	48,687
507	African, Homeland, Female, Upper Secondary and higher	59	29,445	44,194	39,697	48,158	3,435	6,287	38,024	93,174	258,099
508	African, Homeland, Male, None	171	68,107	13,340	10,378	10,080	1,356	4,233	9,840	26,200	56,640
509	African, Homeland, Male, Primary	191	78,067	19,081	24,123	13,704	879	5,359	13,800	40,438	257,626
510	African, Homeland, Male, Lower Secondary	85	36,120	21,620	23,820	16,314	2,580	5,021	13,573	56,754	124,776
511	African, Homeland, Male, Upper Secondary and higher	92	42,562	68,263	84,851	65,888	2,625	10,372	45,480	133,843	482,067
512	African, Non-Homeland, Female, None	259	113,400	15,078	11,726	10,873	2,280	6,000	11,254	30,032	79,869
513	African, Non-Homeland, Female, Primary	435	207,518	16,305	25,180	10,718	884	4,641	10,200	33,600	455,215
514	African, Non-Homeland, Female, Lower Secondary	217	111,464	20,135	20,431	14,310	1,145	5,468	13,662	42,858	142,800
515	African, Non-Homeland, Female, Upper Secondary and higher, Low-income (split)	100	51,942	11,349	5,533	10,501	2,521	4,315	11,292	19,417	21,260
516	African, Non-Homeland, Female, Upper Secondary and higher, High-income (split)	99	52,052	59,184	52,420	45,679	21,400	23,859	43,490	97,225	368,385

... Table 10 continued...

KwaZulu-Natal	Description	Observations (unweighted)	Observations (weighted)	Mean (totine)	St.dev. (totine)	Interquartile range	Min (totine)	P10 (totine)	Median (totine)	P90 (totine)	Max (totine)
517	African, Non-Homeland, Male, None	253	103,153	17,569	18,516	13,337	977	5,434	12,716	36,325	181,669
518	African, Non-Homeland, Male, Primary	447	198,493	21,783	20,870	18,146	664	6,120	16,536	41,469	283,751
519	African, Non-Homeland, Male, Lower Secondary, Low-income (split)	155	75,107	11,161	5,459	9,238	1,600	3,817	10,150	18,900	21,351
520	African, Non-Homeland, Male, Lower Secondary, High-income (split)	155	78,696	61,094	83,596	29,977	21,596	24,132	39,231	95,200	854,652
521	African, Non-Homeland, Male, Upper Secondary and higher, Low-income (split)	138	66,546	14,778	7,415	13,123	12	5,030	14,582	24,989	28,119
522	African, Non-Homeland, Male, Upper Secondary and higher, High-income (split)	138	68,277	101,865	146,590	67,974	28,153	32,500	63,200	214,248	1,552,595
523	Asian, Female, Lower Secondary and lower	75	41,224	45,332	59,618	28,909	4,169	12,390	33,833	87,232	468,275
524	Asian, Male, Lower Secondary and lower, Low-income (split)	69	33,463	25,357	10,274	18,766	6,519	12,400	26,144	39,931	45,154
525	Asian, Male, Lower Secondary and lower, High-income (split)	69	35,496	104,311	75,721	72,507	45,227	48,220	67,200	210,003	390,233
526	Asian, Male, Upper Secondary and higher, Low-income (split)	84	43,767	51,270	18,580	27,932	11,450	23,158	50,400	74,321	83,280
527	Asian, Male, Upper Secondary and higher, High-income (split)	85	41,818	158,162	68,472	83,339	83,616	87,329	146,042	260,000	427,262
528	Coloured	47	22,716	55,530	56,351	45,587	4,794	9,607	39,791	144,960	250,774
529	White, Lower Secondary and lower	38	21,773	79,863	68,810	72,388	10,965	18,392	60,293	165,276	460,068
530	White, Upper Secondary, Low-income (split)	78	47,883	63,382	27,091	37,418	8,880	24,005	61,430	100,802	112,020
531	White, Upper Secondary, High-income (split)	78	44,899	241,975	186,036	125,988	112,800	121,171	184,000	428,617	1,295,786
532	White, Tertiary	64	39,373	287,772	880,540	188,794	12,233	38,400	153,600	363,600	7,569,990

Table 11: North West household groups – summary statistics

North West	Description	Observations (unweighted)	Observations (weighted)	Mean (totinc)	St.dev. (totinc)	Interquartile range	Min (totinc)	P10 (totinc)	Median (totinc)	P90 (totinc)	Max (totinc)
601	African, Agricultural	78	22,038	36,986	72,146	20,172	1,022	3,291	12,821	80,160	348,586
602	African, Female, None	203	59,194	14,021	15,003	9,057	572	5,159	10,320	29,016	153,308
603	African, Female, Primary	371	105,202	15,369	17,402	10,157	782	4,334	10,576	30,402	151,073
604	African, Female, Lower Secondary	207	60,151	20,354	19,913	16,847	449	5,247	13,946	43,972	179,716
605	African, Female, Upper Secondary and higher	175	46,590	41,781	45,554	47,958	1,356	5,941	26,925	84,232	320,000
606	African, Male, None, Low-income (split)	133	39,412	7,947	3,249	4,108	834	3,435	7,986	12,609	13,098
607	African, Male, None, High-income (split)	134	37,985	32,156	25,565	19,892	13,200	14,445	24,000	58,506	199,996
608	African, Male, Primary, Low-income (split)	270	74,856	9,957	4,714	7,273	401	3,579	9,550	16,744	18,970
609	African, Male, Primary, High-income (split)	272	80,224	39,561	32,428	17,625	19,148	20,860	30,230	60,840	314,720
610	African, Male, Lower Secondary, Low-income (split)	199	58,752	13,696	7,430	14,032	1,041	4,580	14,007	24,300	25,949
611	African, Male, Lower Secondary, High-income (split)	197	58,553	45,425	22,694	19,671	26,000	27,625	37,689	74,900	172,767
612	African, Male, Upper Secondary and higher, Low-income (split)	171	46,489	17,166	9,059	14,195	1,867	5,384	16,720	30,354	35,401
613	African, Male, Upper Secondary and higher, High-income (split)	174	46,276	100,906	179,815	56,883	36,000	39,680	69,675	153,617	2,587,039
614	Asian & Coloured	47	11,443	53,709	61,626	55,936	821	7,052	27,993	126,156	360,000
615	White, Lower Secondary and lower	80	22,489	88,244	82,634	70,980	10,642	17,932	70,800	180,823	498,213
616	White, Upper Secondary and higher	92	24,696	251,112	554,006	96,191	24,031	71,576	140,226	336,000	4,665,813

Table 12: Gauteng household groups – summary statistics

Gauteng	Description	Observations (unweighted)	Observations (weighted)	Mean (totinc)	St.dev. (totinc)	Interquartile range	Min (totinc)	P10 (totinc)	Median (totinc)	P90 (totinc)	Max (totinc)
701	African, Agricultural	60	50,513	20,179	17,650	17,816	1,543	4,410	15,295	42,434	98,476
702	African, Non-Homeland, Female, None	115	91,271	17,365	12,309	12,696	3,000	6,480	12,489	35,080	54,279
703	African, Non-Homeland, Female, Primary	330	242,907	20,738	26,410	16,004	1,430	6,000	14,462	43,162	412,130
704	African, Female, Lower Secondary	339	254,590	26,534	27,996	19,516	1,920	7,343	16,739	57,928	230,153
705	African, Non-Homeland, Female, Upper Secondary, Low-income (split)	111	79,481	13,163	5,773	9,836	3,200	5,123	12,000	21,655	24,000
706	African, Non-Homeland, Female, Upper Secondary, High-income (split)	111	78,736	63,155	48,514	38,279	24,379	28,697	47,894	118,037	278,348
707	African, Non-Homeland, Female, Tertiary	26	25,399	99,336	56,475	71,160	7,382	21,744	96,000	182,910	227,093
708	African, Non-Homeland, Male, None	171	123,662	26,450	26,286	20,809	1,368	6,516	18,640	48,000	192,779
709	African, Non-Homeland, Male, Primary	615	448,047	29,433	35,797	24,052	489	6,740	23,808	52,180	619,152
710	African, Non-Homeland, Male, Lower Secondary	638	483,627	35,848	43,004	26,290	1,600	9,000	25,707	65,433	614,152
711	African, Non-Homeland, Male, Upper Secondary	603	442,762	53,976	141,520	41,749	1,800	10,400	32,160	111,600	3,815,376
712	African, Non-Homeland, Male, unknown	43	34,590	33,444	27,341	20,751	4,800	12,185	27,718	68,400	151,200
713	African, Non-Homeland, Male, Tertiary, Low-income (split)	33	24,920	49,900	25,698	50,054	7,123	17,048	57,028	84,000	84,480
714	African, Non-Homeland, Male, Tertiary, High-income (split)	33	28,077	203,930	124,801	124,442	84,720	96,567	162,000	438,000	630,033
715	Coloured, Lower Secondary and lower	96	59,719	44,767	49,031	30,421	4,117	8,840	32,935	103,680	288,569
716	Coloured, Upper Secondary and higher	62	45,315	107,452	142,546	53,251	11,153	20,837	52,608	258,000	651,290
717	Asian, Lower Secondary and lower	22	15,980	62,699	72,030	63,175	1,446	7,180	31,534	190,944	296,283
718	Asian, Upper Secondary and higher	44	29,390	147,189	99,134	114,864	12,325	50,000	129,000	311,774	398,000
719	White, Lower Secondary and lower, Low-income (split)	54	52,422	31,800	11,330	17,064	5,460	16,499	33,049	47,266	48,000
720	White, Lower Secondary and lower, High-income (split)	55	60,158	128,475	70,139	89,580	48,548	60,330	110,400	240,000	484,070
721	White, Upper Secondary, Low-income (split)	117	116,516	69,798	31,748	51,626	6,480	24,480	67,580	114,000	124,560
722	White, Upper Secondary, High-income (split)	118	113,642	231,262	117,642	111,558	124,611	138,100	209,363	316,000	882,868
723	White, Tertiary, Low-income (split)	75	87,511	123,189	48,125	66,306	33,849	65,670	116,209	195,600	203,207
724	White, Tertiary, High-income (split)	75	77,026	439,976	321,448	273,419	205,162	236,500	340,207	693,421	1,939,390

Table 13: Mpumalanga household groups – summary statistics

Mpumalanga	Description	Observations (unweighted)	Observations (weighted)	Mean (totinc)	St.dev. (totinc)	Interquartile range	Min (totinc)	P10 (totinc)	Median (totinc)	P90 (totinc)	Max (totinc)
801	African, Agricultural	92	23,309	26,714	43,886	13,772	1,388	5,274	14,872	42,297	273,497
802	African, Female, None	333	90,139	16,389	19,423	9,145	2,005	6,480	11,707	28,834	270,877
803	African, Female, Primary	283	77,549	18,463	19,806	12,376	1,759	5,040	12,070	37,681	175,492
804	African, Female, Lower Secondary	146	38,778	20,534	29,841	12,542	3,475	5,990	12,081	45,239	378,142
805	African, Female, Upper Secondary and higher	121	34,329	39,543	43,517	41,729	679	5,667	20,566	108,802	179,110
806	African, Male, None	281	79,891	22,024	25,869	16,884	1,957	6,000	14,669	41,633	297,000
807	African, Male, Primary, Low-income (split)	218	58,294	9,412	3,763	6,181	633	4,281	9,522	14,471	16,622
808	African, Male, Primary, High-income (split)	217	69,463	41,895	42,896	22,195	16,632	19,584	32,411	64,878	492,169
809	African, Male, Lower Secondary, Low-income (split)	114	28,706	12,963	5,901	10,176	2,366	5,613	12,509	21,136	24,593
810	African, Male, Lower Secondary, High-income (split)	114	31,736	59,495	46,147	32,253	24,726	28,380	43,582	111,908	368,704
811	African, Male, Upper Secondary and higher, Low-income (split)	113	34,036	16,748	8,238	14,264	1,189	6,332	16,471	27,718	33,081
812	African, Male, Upper Secondary and higher, High-income (split)	113	31,182	88,848	69,153	58,884	33,433	37,231	67,263	167,327	469,971
813	Asian & Coloured	37	9,751	82,754	69,049	94,438	2,755	11,671	74,049	181,129	290,888
814	White	95	41,248	147,739	100,564	140,245	7,118	24,152	130,274	272,460	486,377

Table 14: Limpopo household groups – summary statistics

Limpopo	Description	Observations (unweighted)	Observations (weighted)	Mean (totinc)	St.dev. (totinc)	Interquartile range	Min (totinc)	P10 (totinc)	Median (totinc)	P90 (totinc)	Max (totinc)
901	African, Agricultural	95	32,766	18,085	18,936	14,290	1,725	5,913	10,937	34,640	140,685
902	African, Female, Non & pre-Primary	592	224,716	14,640	22,866	7,546	339	5,180	9,120	26,518	327,293
903	African, Female, Primary	420	139,955	13,020	18,158	8,107	1,172	3,778	8,957	23,619	317,328
904	African, Female, Lower Secondary	244	81,393	13,254	19,120	8,830	1,340	3,619	8,509	24,560	250,884
905	African, Female, Upper Secondary and higher, Low-income (split)	142	44,842	8,693	3,848	6,098	1,978	3,965	8,428	14,992	16,333
906	African, Female, Upper Secondary and higher, High-income (split)	142	39,701	64,300	72,305	57,231	16,520	18,656	48,370	107,488	746,514
907	African, Male, None	252	96,166	17,678	17,150	14,187	818	5,400	13,304	35,083	171,801
908	African, Male, Primary, Low-income (split)	238	84,302	7,365	3,417	5,291	750	2,708	7,118	12,250	14,153
909	African, Male, Primary, High-income (split)	240	74,842	44,319	59,142	24,235	14,160	15,600	26,528	88,716	570,200
910	African, Male, Lower Secondary	268	91,424	27,089	68,270	24,056	1,145	3,794	14,125	52,834	1,602,586
911	African, Male, Upper Secondary and higher, Low-income (split)	184	50,301	13,072	8,581	12,516	1,132	4,116	10,186	27,200	32,462
912	African, Male, Upper Secondary and higher, High-income (split)	184	45,827	120,363	84,315	92,452	33,514	43,200	88,000	236,589	545,065
913	Asian & Coloured	15	3,655	91,935	116,657	91,440	4,800	6,000	65,201	305,400	373,280
914	White	86	21,728	194,615	223,669	142,861	9,228	48,000	128,464	342,899	1,323,288

#### 4. Factor groups for a South Africa SAM

Since households earn the largest share of their income from labour, the link between household and factor groups is important. Factors are first disaggregated by province and race, and hence the primary direct link between factor and households is via these two dimensions.<sup>6</sup> Factor groups are also formed around skill groups, which links in with the education attribute in the household groups.

Labour income and occupation data can be sourced from either the Income and Expenditure Survey of 2000 (IES 2000) or the Labour Force Survey of September 2000 (LFS 2000:2). The LFS 2000:2 is designed specifically to gather factor data, and as such contains more probing questions about activities of workers. One can therefore assume that the occupation code data in the LFS is more accurate.<sup>7</sup> However, as discussed in PROVIDE (2005b), there are some concerns about the quality of the LFS 2000:2 factor income data. The IES 2000 factor income data is also not of a very high quality. Consequently a combined IES-LFS factor income variable was created (see 2005b, for a detailed discussion). The LFS occupation codes were, however, used throughout (see Table 15).

Table 15: Occupation codes

Factor code	Description
0	Not applicable/not working
1	Legislators, senior officials and managers
2	Professionals
3	Technical and associate professionals
4	Clerks
5	Service workers and shop and market sales workers
6	Skilled agricultural and fishery workersta
7	Craft and related trades workers
8	Plant and machine operators and assemblers
9	Elementary Occupation
10	Domestic workers
11	Not adequately or elsewhere defined, unspecified

Source: LFS 2000:2

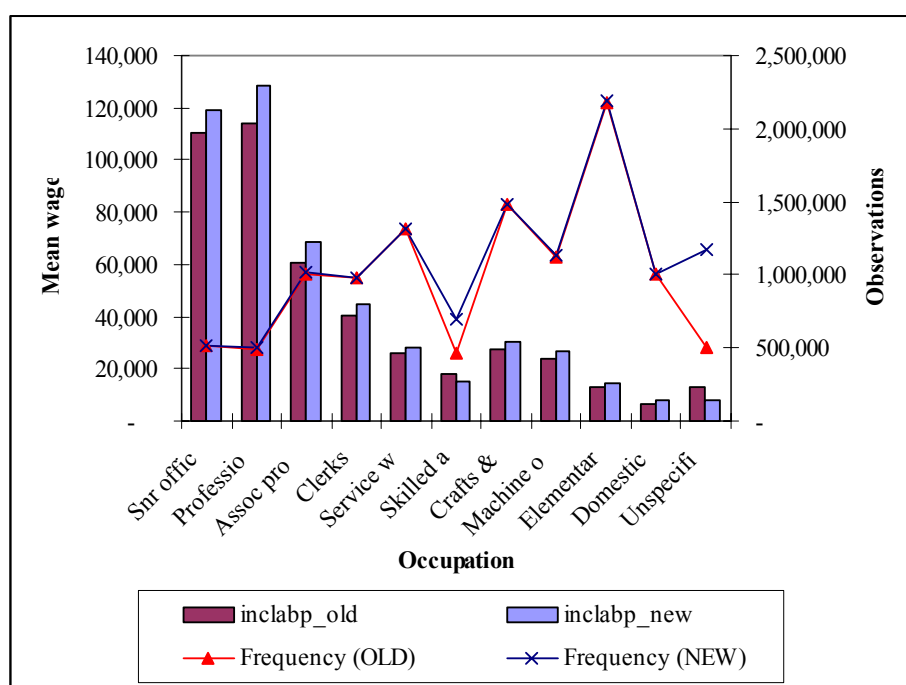
All respondents reporting zero factor income fall under factor code 0. Thus, all employment figures are based on people who actually earned an income in the year 2000. The combined IES-LFS factor income variable created has two ‘versions’ –

<sup>6</sup> A limited number of people live in mixed-race households where the race of one or more of the household members is not the same as the head of the household. About 0.3% of African, 1.2% of Coloured, 1.5% of Asian and 1.0% of White household members report a different race group than the heads of their respective households.

<sup>7</sup> The IES 2000 only asks respondents a single question to determine their occupation code, while the LFS 2000:2 has a series of questions on the topic.

*inclabp\_old* and *inclabp\_new* (see Figure 10). The ‘old’ version is the original combined IES-LFS variable, while the ‘new’ variable was scaled so that the sum of the individual household members’ wages equal the total wage income reported by the household. Unfortunately there are many more unspecified workers for the new factor income variable since household members that previously reported no labour income and no occupation code now ‘receives’ income from labour. This approach to scaling up the data still needs some further consideration, and hence the old variable is still used for the formation of factor groups at this stage (see PROVIDE, 2005b for a discussion).

Figure 10: Average wage and number of observations per factor group



The formation of factor groups for the SAM depended mainly on the number of observations available within a given factor group. All workers were first disaggregated by province and race, and thereafter into occupation groups as per Table 15. In cases where there were too few observations to justify a single factor group to be included in the SAM, the group was merged with another factor group with a similar skills profile. Typically, factor codes 1 to 3 were grouped together as highly skilled workers, codes 4 – 5 as skilled, codes 6 – 8 as semi-skilled and codes 9 – 11 as unskilled.<sup>8</sup>

<sup>8</sup> Unfortunately code 11 (not adequately defined) may possibly include semi-skilled, skilled or high-skilled workers as well, but this is virtually impossible to determine. One option would be to allocate the workers to certain skill classes depending on their reported wage level. This option is being explored for possible inclusion in a later version of the PROVIDE SAM.

The basic principles followed when forming the factor groups were the following. Asian and Coloured workers were grouped together whenever these two racial groups were also grouped for the household groups in the relevant province. Once the province-race subgroups were formed, each initially with 11 occupation codes, occupations were grouped together when the number of observations were very low. Groups were aggregated until the number of observations in the subgroup was at least 5% of the provincial total, provided that there remained at least two factor groups per province-race subgroup.

Table 16 to Table 24 provides descriptions and summary statistics of the factor groups in each province. The first column contains the code for each factor group. Columns three and four contain the number of observations, first at sample level ('unweighted') and then at the population level (weighted). The LFS 2000:2 'person weights' were used. Variable *inclabp\_old* is the wage or salary income. The rest of the columns in the table report the mean and standard deviation (weighted) of *inclabp\_old*, as well as the interquartile range, the minimum, the 10<sup>th</sup> percentile (P10), median, 90<sup>th</sup> percentile (P90) and maximum of *inclabp\_old*.

Table 16: Western Cape factor groups – summary statistics (*inclabp\_old*)

Western Cape	Description	Observations (unweighted)	Observations (weighted)	Mean ( <i>inclabp_old</i> )	St.dev. ( <i>inclabp_old</i> )	Interquartile range	Min ( <i>inclabp_old</i> )	P10 ( <i>inclabp_old</i> )	Median ( <i>inclabp_old</i> )	P90 ( <i>inclabp_old</i> )	Max ( <i>inclabp_old</i> )
101	African, High-skilled & Skilled	179	71,647	32,591	36,470	23,000	1,300	5,880	19,000	72,000	201,500
102	African, Semi-skilled	172	74,851	17,057	11,047	10,000	600	7,800	14,400	28,800	78,000
103	African, Unskilled	373	143,399	11,368	6,575	8,600	90	4,800	9,672	20,400	36,400
104	Coloured & Asian, High-skilled	243	112,653	66,839	46,300	54,712	1,560	18,000	60,000	120,000	294,688
105	Coloured & Asian, Clerks	206	90,791	41,469	31,822	26,740	2,400	11,960	32,400	82,200	207,564
106	Coloured & Asian, Services and sales	195	77,878	28,773	24,806	26,050	300	6,000	20,488	67,840	149,794
107	Coloured & Asian, Craft and trade	278	115,567	27,905	19,708	18,324	1,000	9,360	24,000	54,000	133,532
108	Coloured & Asian, Machine and plant operators	283	98,687	28,598	19,033	15,600	2,400	10,800	24,000	51,000	132,600
109	Coloured & Asian, Elementary	991	264,059	14,924	11,417	11,920	1,200	5,200	10,800	29,355	72,600
110	Coloured & Asian, Agriculture and fisheries & Domestic workers & Unspecified	430	147,509	12,282	29,680	9,400	34	1,200	7,280	20,580	288,000
111	White, High-skilled	224	172,464	137,185	121,379	120,000	1,200	36,000	102,000	282,000	800,000
112	White, Skilled	119	89,799	50,650	35,066	35,000	180	18,000	43,488	89,000	320,000
113	White, Semi- & Unskilled	126	73,105	58,412	68,948	74,040	230	6,440	36,000	142,440	668,296

Table 17: Eastern Cape factor groups – summary statistics (*inclabp\_old*)

Eastern Cape	Description	Observations (unweighted)	Observations (weighted)	Mean (inclabp_old)	St.dev. (inclabp_old)	Interquartile range	Min (inclabp_old)	P10 (inclabp_old)	Median (inclabp_old)	P90 (inclabp_old)	Max (inclabp_old)
201	African, High-skilled	302	118,521	50,895	35,702	36,000	600	12,000	47,800	88,884	247,405
202	African, Skilled	336	139,220	20,796	22,669	21,810	240	2,808	11,400	50,136	162,000
203	African, Agriculture and fisheries	155	65,757	6,848	8,129	6,840	40	780	3,600	16,800	41,660
204	African, Craft and trade	221	93,891	10,804	11,798	8,400	300	1,800	6,600	24,000	75,000
205	African, Machine and plant operators	151	60,158	18,783	12,965	15,000	2,400	5,200	15,600	37,684	72,000
206	African, Elementary	511	220,575	9,665	15,630	9,120	60	1,200	6,000	21,600	276,000
207	African, Domestic workers & Unspecified	343	134,769	4,560	5,726	3,066	30	1,440	3,600	6,600	72,700
208	Coloured & Asian, High-skilled & Skilled	133	54,491	47,689	59,469	43,346	1,200	7,800	30,000	114,000	416,600
209	Coloured & Asian, Semi- & Unskilled	241	101,448	13,058	13,906	13,680	240	2,400	7,920	27,600	96,000
210	White, High-skilled	109	59,691	96,962	81,193	66,000	4,200	26,880	72,000	202,200	540,000
211	White, Skilled	87	45,603	49,032	28,726	33,000	5,760	14,400	48,000	84,000	162,000
212	White, Semi- & Unskilled	64	29,882	44,421	30,996	46,480	1,200	10,500	36,000	84,000	132,000

Table 18: Northern Cape factor groups – summary statistics (*inclabp\_old*)

Northern Cape	Description	Observations (unweighted)	Observations (weighted)	Mean (inclabp_old)	St.dev. (inclabp_old)	Interquartile range	Min (inclabp_old)	P10 (inclabp_old)	Median (inclabp_old)	P90 (inclabp_old)	Max (inclabp_old)
301	African, High-skilled & Skilled	110	18,919	34,807	58,801	37,000	225	3,600	19,800	65,000	474,000
302	African, Semi- & Unskilled	414	66,465	11,354	15,222	8,840	220	2,400	5,880	25,412	161,200
303	Coloured & Asian, High-skilled & Skilled	167	31,386	46,422	62,173	54,580	900	4,800	32,500	84,109	402,000
304	Coloured & Asian, Semi- & Unskilled	548	95,453	12,278	21,041	9,400	150	1,920	6,000	26,000	360,000
305	White, High-skilled & Skilled	150	29,785	81,127	80,726	49,583	3,600	21,600	60,000	144,000	540,000
306	White, Semi- & Unskilled	100	19,902	107,349	154,886	91,608	15	10,800	60,000	200,000	900,000

Table 19: Free State factor groups – summary statistics (*inclabp\_old*)

Free State	Description	Observations (unweighted)	Observations (weighted)	Mean (inclabp_old)	St.dev. (inclabp_old)	Interquartile range	Min (inclabp_old)	P10 (inclabp_old)	Median (inclabp_old)	P90 (inclabp_old)	Max (inclabp_old)
401	African, High-skilled & Skilled	411	143,070	33,476	32,684	41,840	100	3,408	23,400	73,862	240,000
402	African, Semi-skilled	821	250,872	15,395	15,180	18,336	180	2,400	10,800	32,346	237,087
403	African, Unskilled	778	231,241	6,466	6,951	4,800	80	1,200	4,160	15,660	56,312
404	Coloured & Asian, High-skilled & Skilled	31	14,320	35,566	39,025	36,600	1,200	2,600	30,000	84,000	174,483
405	Coloured & Asian, Semi- & Unskilled	23	10,875	10,409	9,146	9,960	300	2,400	8,400	19,500	42,000
406	White, High-skilled & Skilled	194	92,975	70,802	68,597	59,100	6,000	14,400	48,000	156,000	370,400
407	White, Semi- & Unskilled	79	32,832	55,829	105,757	57,000	1,214	3,600	36,000	114,000	1,532,000

Table 20: KwaZulu-Natal factor groups – summary statistics (*inclabp\_old*)

KwaZulu-Natal	Description	Observations (unweighted)	Observations (weighted)	Mean (inclabp_old)	St.dev. (inclabp_old)	Interquartile range	Min (inclabp_old)	P10 (inclabp_old)	Median (inclabp_old)	P90 (inclabp_old)	Max (inclabp_old)
501	African, High-skilled	358	171,702	48,369	42,408	36,000	1,200	7,800	39,600	84,000	276,000
502	African, Skilled	513	245,615	20,155	27,873	16,200	240	4,800	14,400	42,000	508,896
503	African, Agriculture and fisheries	262	105,652	8,626	21,094	5,400	30	1,200	5,040	14,400	354,000
504	African, Craft and trade	395	177,617	16,069	13,787	13,000	240	4,200	12,480	30,000	112,880
505	African, Machine and plant operators	399	171,230	20,944	15,429	15,952	1,800	6,240	16,800	39,600	103,044
506	African, Elementary	926	374,624	10,570	10,395	8,900	240	2,800	7,200	22,100	119,600
507	African, Domestic workers & Unspecified	573	254,762	9,257	65,692	4,000	1	1,800	4,800	10,800	1,414,976
508	Coloured, High-skilled & Skilled	31	36,139	36,438	28,924	33,000	1,800	8,400	24,000	72,000	120,000
509	Coloured, Semi- & Unskilled	29	27,035	24,394	21,397	21,000	2,400	4,848	17,280	49,920	96,000
510	Asian, High-skilled & Skilled	306	165,725	48,020	40,281	41,700	2,000	12,000	36,000	95,500	300,000
511	Asian, Semi- & Unskilled	231	122,522	31,352	27,082	22,291	100	7,200	24,000	63,400	180,000
512	White, High-skilled & Skilled	238	190,846	92,015	124,599	79,600	3,000	21,600	66,000	180,000	1,598,000
513	White, Semi- & Unskilled	72	57,397	54,122	45,324	60,000	1	6,000	48,000	120,000	192,200

Table 21: North West factor groups – summary statistics (*inclabp\_old*)

North West	Description	Observations (unweighted)	Observations (weighted)	Mean (inclabp_old)	St.dev. (inclabp_old)	Interquartile range	Min (inclabp_old)	P10 (inclabp_old)	Median (inclabp_old)	P90 (inclabp_old)	Max (inclabp_old)
601	African, High-skilled & Skilled	706	217,200	32,944	35,152	34,700	600	5,500	24,000	67,400	540,000
602	African, Semi-skilled	805	266,790	19,488	15,337	16,200	110	4,200	18,000	36,000	140,000
603	African, Unskilled	813	254,389	9,600	9,682	8,940	60	1,800	6,260	21,600	100,000
604	Coloured & Asian, High-skilled & Skilled	41	16,277	35,412	41,296	36,000	3,000	7,280	21,800	63,432	250,000
605	Coloured & Asian, Semi- & Unskilled	19	7,517	17,279	15,752	24,000	1,680	3,000	11,100	46,800	48,000
606	White, High-skilled & Skilled	113	45,662	73,760	78,249	66,000	8,000	22,800	48,000	132,000	540,000
607	White, Semi- & Unskilled	77	30,634	94,093	92,965	78,144	300	24,000	74,400	162,000	540,000

Table 22: Gauteng factor groups – summary statistics (*inclabp\_old*)

Gauteng	Description	Observations (unweighted)	Observations (weighted)	Mean (inclabp_old)	St.dev. (inclabp_old)	Interquartile range	Min (inclabp_old)	P10 (inclabp_old)	Median (inclabp_old)	P90 (inclabp_old)	Max (inclabp_old)
701	African, High-skilled	432	231,122	56,982	55,436	48,000	1,500	12,000	42,000	114,000	375,000
702	African, Clerks	301	156,914	40,212	183,862	22,400	3,600	10,400	26,400	60,000	3,785,376
703	African, Services and sales	549	274,671	20,807	17,821	13,600	520	5,280	17,500	36,000	139,179
704	African, Craft and trade	562	296,404	20,522	15,099	14,600	500	6,000	17,400	36,000	108,000
705	African, Machine and plant operators	514	264,348	23,919	15,169	15,600	1,248	10,400	20,800	40,000	186,000
706	African, Elementary	684	361,448	15,558	12,939	11,700	180	3,600	13,000	30,000	180,000
707	African, Domestic workers & Agriculture and fisheries & Unspecified	705	394,189	9,413	8,612	7,200	111	2,400	7,280	18,000	96,500
708	Coloured, High-skilled & Skilled	114	56,398	60,267	68,972	39,000	1,200	18,980	36,842	192,000	424,960
709	Coloured, Semi- & Unskilled	89	40,530	22,763	18,924	20,400	1,200	4,800	18,826	42,000	96,000
710	Asian, High-skilled & Skilled	76	56,045	69,857	47,504	44,160	6,000	24,000	63,000	144,000	229,404
711	Asian, Semi- & Unskilled	17	13,046	51,739	52,058	51,600	6,000	6,480	42,000	108,000	180,000
712	White, High-skilled	338	346,496	140,286	152,315	99,000	6,000	43,080	96,000	276,000	1,500,000
713	White, Skilled	206	204,449	58,156	53,221	37,200	1,440	19,200	48,000	96,545	390,000
714	White, Semi- & Unskilled	135	137,230	54,689	56,183	51,600	300	4,680	42,000	114,000	360,000

Table 23: Mpumalanga factor groups – summary statistics (*inclabp\_old*)

Mpumalanga	Description	Observations (unweighted)	Observations (weighted)	Mean (inclabp_old)	St.dev. (inclabp_old)	Interquartile range	Min (inclabp_old)	P10 (inclabp_old)	Median (inclabp_old)	P90 (inclabp_old)	Max (inclabp_old)
801	African, High-skilled	222	67,705	47,178	39,804	51,940	200	5,400	39,000	91,800	276,000
802	African, Skilled	344	99,412	17,243	17,877	18,600	200	2,400	11,400	42,000	107,905
803	African, Semi-skilled	732	219,788	17,518	16,511	18,000	90	2,880	12,000	36,600	130,000
804	African, Unskilled	889	249,087	9,467	13,620	7,200	1	1,800	6,000	20,000	312,000
805	Coloured & Asian, High-skilled & Skilled	29	12,261	50,926	30,087	42,000	6,480	18,000	49,400	79,200	140,400
806	Coloured & Asian, Semi- & Unskilled	18	6,428	49,073	54,482	74,400	4,200	6,000	12,900	142,300	174,000
807	White, High-skilled & Skilled	66	41,711	86,748	59,249	84,900	8,400	18,000	89,803	156,000	276,000
808	White, Semi- & Unskilled	50	31,646	72,714	70,411	88,072	250	6,000	48,000	184,100	336,000

Table 24: Limpopo factor groups – summary statistics (*inclabp\_old*)

Limpopo	Description	Observations (unweighted)	Observations (weighted)	Mean (inclabp_old)	St.dev. (inclabp_old)	Interquartile range	Min (inclabp_old)	P10 (inclabp_old)	Median (inclabp_old)	P90 (inclabp_old)	Max (inclabp_old)
901	African, High-skilled	390	124,474	54,332	44,368	48,395	260	8,400	48,000	102,000	540,000
902	African, Skilled	357	125,159	18,486	21,755	19,200	360	3,000	10,200	43,385	180,000
903	African, Semi-skilled	596	230,814	15,162	26,639	13,440	450	2,400	8,400	30,540	312,000
904	African, Unskilled	825	310,758	8,794	12,202	6,000	30	1,800	4,800	20,400	120,000
905	Coloured & Asian, High-skilled & Skilled	19	14,357	47,436	66,982	66,060	5,100	5,400	6,600	102,000	242,000
906	Coloured & Asian, Semi- & Unskilled	5	3,921	14,761	15,697	11,500	4,800	4,800	8,400	45,600	45,600
907	White, High-skilled & Skilled	80	36,141	78,459	60,077	80,200	3,000	18,000	66,000	162,000	276,000
908	White, Semi- & Unskilled	37	15,755	91,141	134,840	75,000	300	12,000	48,960	144,000	629,000

The household-factor link is not the only important factor link in a SAM. Also important is the value-added sub-matrix, which shows the link between factors and activities, i.e. it shows the flow of resources (value added) from activities (industries) to factors. The total flow of resources is equal to the wage multiplied by quantity or the number of workers in the case of labour. In some cases a modeller using the SAM may be interested in actual employment levels. In such instances it is necessary to specify a factor use matrix, which shows the employment levels by factors and activities. If, for example, the factor use data is used in a CGE model, it is possible to generate actual changes in employment levels (at industry and/or occupation level, depending on the closure rules selected) that relates to real employment data, rather than data on hypothetical relative employment changes.

In order to estimate the factor use matrix the following steps are followed. In addition to the value-added sub-matrix that is extracted from the IES/LFS 2000 database (see PROVIDE, 2005b for details), an average wage sub-matrix is also extracted. This average wage sub-matrix is defined over factors ( $f$ ) and activities ( $a$ ), say  $avwage(f, a)$ . The average wage data is read into the SAM estimation process as a parameter and stays unchanged as the cell entries of the value-added sub-matrix are changed during the SAM estimation process. Once the final SAM has been estimated, the new entries in the value-added sub-matrix are divided by the average wage estimates, which gives a factor use matrix defined over  $f$  and  $a$ , say  $factuse(f, a)$ . All this takes place within the SAM estimation process (see PROVIDE, 2005a for more on the estimation process).

The extraction of a detailed average wage matrix (parameter) is not straightforward. In the original LFS 2000:2 data there is only a single agricultural activity account. However, in the fully disaggregated SAM this account is split into numerous accounts for each of the nine provinces, thus giving 79 agricultural accounts.<sup>9</sup> It is therefore necessary to split the value-added data as well as the average wage data from the single agricultural activity account into 79 agricultural activities. The Agricultural Survey of 1996 (SSA, 1999) was conducted at this higher disaggregated level. This survey is used to obtain agricultural employment ratios by race and province, as well as the ratios of total value-added payments from activities to factors.

The following process is therefore followed to create the average wage sub-matrix or parameter: Firstly, employment data is extracted from the LFS 2000:2. This involves extracting the weighted number of workers reporting positive wage income. Next, the

---

<sup>9</sup> Not all of these accounts contain data, and consequently nine are dropped from the final SAM, giving 70 agricultural activities spread over nine provinces.

employment data for agriculture is split using the aforementioned employment ratios so that total employment for each province-race sub-group remains consistent with the employment level reported in the LFS 2000:2, while the distribution of workers between different province-race groups is consistent with the ratios from the Agricultural Census of 1996. The total value added ratios, also calculated from the Agricultural Census of 1996, are used to split the total value added of the single LFS 2000:2 agricultural activity into the 79 agricultural activities. Finally, an estimate of average wages for each factor-activity subgroup is obtained by dividing the value-added data by the factor use matrix.

There were also some other non-agricultural industries for which the data in the LFS 2000:2 was not disaggregated at the same level as the PROVIDE SAM activity disaggregation. For these industries it is simply assumed that the average wage is the same in each industry. This assumption is necessary because no external source of information is available on the distribution of wages between those industries as was the case of the agricultural industries.

## **5. Alternative household and factor groupings**

Various other possible household and factor groupings were explored and used in previous versions of the PROVIDE SAMs. Although the household and factor groups described in sections 3 and 4 are currently incorporated in the PROVIDE SAM, this section describes some alternative household and factor groupings that can be used. The Stata code used to form household and factor groups is set up so that these groups can be recreated fairly easily.

### **5.1. Household groups for a previous version of the National SAM**

A former version of the PROVIDE National SAM used the same household classification as the Western Cape SAM compiled by McDonald and Punt (2001).<sup>10</sup> In this classification scheme households are first divided into race groups, thereafter into rural and urban households, and finally by income group. While McDonald and Punt (2001) used total household income to form income groups, the current household income groups are based on an adult equivalent per capita income of each household.

Although there is some opposition against the continued racial focus of South African economic analyses, a racial classification of households remains important in a social accounting context given the large differences in behavioural characteristics

---

<sup>10</sup> The classification was initially developed to be suitable at a national level and then applied to the Western Cape province in order to maintain consistent accounts.

between race groups, driven largely by differences in income levels and the history of segregation. The majority of the estimated 42.5 million (IES 2000) South Africans are classified as African (80.9%). Coloured and White people make up 8.6% and 8.1% respectively, while 2.5% of the population is classified as Asian.<sup>11</sup> Table 25 shows the household racial composition in South Africa. Note this differs slightly from the population composition due to differences in average household sizes between racial groups.

Table 25: Urban-rural and racial household composition in South Africa

Number of households				Percentages*		
	Urban	Rural	Total	Urban	Rural	Total
African	5,064,111	3,536,648	8,600,758	46.2%	32.2%	78.4%
Coloured	761,537	118,510	880,047	6.9%	1.1%	8.0%
Asian	258,966	6,913	265,880	2.4%	0.1%	2.4%
White	1,156,570	70,228	1,226,799	10.5%	0.6%	11.2%
Total	7,241,184	3,732,299	10,973,483	66.0%	34.0%	100.0%

Source: IES 2000

Note (\*): Expressed as cell percentage of total number of households

Consumer preferences are often dictated or influenced by various cultural factors, income levels of households and demographic characteristics such as household size and structure. Table 26 shows the average adult equivalent per capita income of households by racial group. Large differences are apparent, with White households earning on average more than twice as much as Asians, four times as much as Coloureds and about seven times as much as Africans. Economic theory predicts (and evidence shows) that income is an important determinant of expenditure patterns of households, e.g., Engel’s Law states that low-income households will spend a larger proportion of their income on necessities such as food. This affects the overall expenditure pattern of the household as well. A simple statistical test (Hotelling’s T<sup>2</sup>-test) is used to compare expenditure patterns between poor and non-poor households.<sup>12</sup> Poor households spend an average of 54.2c on food per R1.00 spent compared to 30.0c of non-poor households.<sup>13</sup> This is probably the most important factor causing the null

<sup>11</sup> These figures are calculated by multiplying the number of households (see Table 25) by the average household size (see Table 26).

<sup>12</sup> For simplicity poor households are defined as those households with an adult equivalent per capita income which is less than or equal to the 40<sup>th</sup> percentile of adult equivalent per capita income. This is equal to R5,146 per annum per adult equivalent (2000 prices).

<sup>13</sup> The expenditure categories compared are labelled *gfood* (food, beverages and tobacco), *gcloth* (clothing and footwear), *ghouse* (housing, water, electricity and fuels), *gfurn* (furnishings, equipment and maintenance), *gheed* (health and education), *gtrans* (transport), *genter* (entertainment, hotels, cafes and restaurants), *gmisc* (miscellaneous), *gtax* (household income and indirect taxes) and *gsav* (households savings).

hypothesis that poor and non-poor households have similar expenditure patterns to be rejected. The test results (Stata output) appears below: <sup>14</sup>

Table 26: Average income (adult equivalent) and household size measures by race

	African	Coloured	Asian	White
Adult equivalent p.c. income	11,296	18,158	31,608	76,669
A – no. of adults	3.03	3.24	3.30	2.41
K – no of children under 10	0.97	0.93	0.62	0.38
H – household size	4.00	4.17	3.93	2.79
E – adjusted household size*	3.04	3.21	3.15	2.34

Source: IES 2000

Note (\*): See discussion below for an explanation of the adult equivalent household size.

```
. hotel g* [aweight = wgtselect], by(poor);
```

```
-> poor = Poor
```

Variable	Obs	Weight	Mean	Std. Dev.	Min	Max
gfood	11492	4388722.37	54.19415	16.71918	0	100
gcloth	11492	4388722.37	6.719574	7.31477	0	70.98266
ghouse	11492	4388722.37	7.019147	10.17506	0	100
gfurn	11492	4388722.37	3.616409	5.552243	0	100
gheed	11492	4388722.37	14.49425	8.462165	0	100
gtrans	11492	4388722.37	3.553086	5.312033	0	69.69719
gender	11492	4388722.37	3.440273	6.684102	0	95.72431
gmisc	11492	4388722.37	5.459808	7.907538	0	100
gtax	11492	4388722.37	.8216886	3.069754	0	74.64694
gsav	11492	4388722.37	.6816212	3.873976	0	64.58732

```
-> poor = Non-poor
```

Variable	Obs	Weight	Mean	Std. Dev.	Min	Max
gfood	14685	6584431.96	29.97955	18.62405	0	100
gcloth	14685	6584431.96	6.132645	6.355863	0	100
ghouse	14685	6584431.96	9.626351	11.31506	0	100
gfurn	14685	6584431.96	4.465699	6.801213	0	74.40686
gheed	14685	6584431.96	13.06647	9.548521	0	92.15281
gtrans	14685	6584431.96	4.28087	8.311693	0	80.649
gender	14685	6584431.96	5.055506	6.750529	0	75.25454
gmisc	14685	6584431.96	14.53832	13.24652	0	96.50835
gtax	14685	6584431.96	7.555742	12.1028	0	96.96163
gsav	14685	6584431.96	5.298847	9.895168	0	95.1045

2-group Hotelling's T-squared = 13060.805

F test statistic: ((26177-10-1)/(26177-2) (10)) x 13060.805 = 1305.6314

H0: Vectors of means are equal for the two groups

F(10,26166) = 1305.6314

Prob > F(10,26166) = 0.0000

Also important in determining expenditure patterns is the household's size (see Table 26). White households, for example, are on average much smaller (2.79) than African (4.00), Coloured (4.17) and Asian (3.97) households. The structure of households also differs between racial groups. About 24% and 22% of African and Coloured household members are children under the age of 10, while this figure is only

<sup>14</sup> The Stata<sup>®</sup> software (referred to throughout as Stata) is a registered trademark of the Stata Corporation (StataCorp, 2001).

16% and 14% for Asian and White households respectively. Hotelling's T<sup>2</sup>-test can be used to compare expenditure patterns of households with above average and below average household sizes respectively.<sup>15</sup> The Stata output below shows that the null hypothesis that small and large households have similar expenditure patterns can be rejected at a 1% significance level.<sup>16</sup>

```
. hotel g* [aweight = wgtselect], by(sizegr);
```

```
-> sizegr = Below average H
```

Variable	Obs	Weight	Mean	Std. Dev.	Min	Max
gfood	13047	5664933.34	36.64072	21.85218	0	100
gcloth	13047	5664933.34	6.006014	6.72588	0	100
ghouse	13047	5664933.34	9.755571	12.23523	0	100
gfurn	13047	5664933.34	3.920403	6.164076	0	100
gheed	13047	5664933.34	12.96497	9.458073	0	100
gtrans	13047	5664933.34	4.016164	7.323931	0	80.649
gender	13047	5664933.34	4.949443	7.250132	0	79.8722
gmisc	13047	5664933.34	12.89884	14.10039	0	100
gtax	13047	5664933.34	5.223456	10.68921	0	90.8438
gsav	13047	5664933.34	3.624424	8.704856	0	92.78826

```
-> sizegr = Above average H
```

Variable	Obs	Weight	Mean	Std. Dev.	Min	Max
gfood	13130	5308220.99	42.89085	20.5547	0	100
gcloth	13130	5308220.99	6.753046	6.7788	0	70.98266
ghouse	13130	5308220.99	7.332868	9.219082	0	96.01681
gfurn	13130	5308220.99	4.345464	6.525383	0	71.08553
gheed	13130	5308220.99	14.35524	8.766181	0	100
gtrans	13130	5308220.99	3.961649	7.213912	0	76.55553
gender	13130	5308220.99	3.833256	6.165895	0	95.72431
gmisc	13130	5308220.99	8.782066	9.450737	0	96.50835
gtax	13130	5308220.99	4.477187	9.474214	0	96.96163
gsav	13130	5308220.99	3.268368	7.969265	0	95.1045

2-group Hotelling's T-squared = 1503.1762

F test statistic: ((26177-10-1)/(26177-2)(10)) x 1503.1762 = 150.26594

H0: Vectors of means are equal for the two groups

F(10,26166) = 150.2659

Prob > F(10,26166) = 0.0000

A racial classification is necessary from a policy analysis point of view. Due to obvious inequalities between racial groups many current social policies aim to improve conditions of previously disadvantaged groups or individuals. A racial disaggregation of households will allow policy analysts to evaluate the impact and efficiency of such policies.

Table 25 shows that about two thirds of South African households live in urban areas. There are various reasons why an urban-rural split is justified. Firstly, as in most

<sup>15</sup> The average weighted household size is 3.9 members.

<sup>16</sup> It has to be mentioned that about two thirds of poor households have above average household sizes (average household size of poor households is 5.1). In contrast to this about two thirds of non-poor households have below average household sizes (average household size of non-poor households is 3.1). Given the correlation between poverty status and household size group it is difficult to say the expenditure pattern differences are driven largely by income or household size differences.

developing countries, households in rural areas are typically more impoverished than their urban counterparts. In South Africa the average adult equivalent per capita income in rural areas is only R7,438 compared to R25,986 in urban areas. Secondly, rural households are also typically larger in size. The average rural household has 4.6 members, compared to the 3.5 members in urban households.<sup>17</sup> Finally, price differences between urban and rural areas further cause expenditure patterns to differ.

Income is often a very important determinant of expenditure patterns of households. Given the differences in household size and structure between households of different racial groups as well as between urban and rural households, caution is needed when disaggregating households on the basis of total household income. Often household-level income (or expenditure) is used as an indicator of households' well-being, but this introduces a degree of bias. The size and structure of households affects expenditure levels and patterns and hence the income level required to fund expenditure.

Consider the size of the household. The size of a household will to a large extent determine the household-level food and clothing expenditure since larger households require more of these necessities to survive. If two households earn the same total income, the larger household will typically spend a larger proportion of its income on food and clothes – and most likely on other goods and services perceived to be necessities. An effective way to deal with this problem is using per capita consumption or income figures. However, large households may also benefit from economics of scale on shared goods – be they necessities or luxuries – such as housing.

The structure of the household is also important. It is typically assumed that a young child does not require the same level of expenditure on food and clothes than an adult. The World Health Organization estimates a young child's nutritional needs at 64% of that of an adult (see Leibbrandt and Woolard, 1999). If one compares two households of equal size and equal household income, the household with relatively more children can be regarded as better off.

The adult equivalence scale adjusts the actual household size to take into account differences in the size and structure of households. The adjusted household size variable  $E$  is constructed using the formula  $E = (A + \alpha K)^\theta$ , where  $A$  refers to the number of adults in a household and  $K$  the number of children. The parameters  $\alpha$  and  $\theta$  control for the size and structure. The lower the value of  $\alpha$  ( $\alpha < 1$ ), the lower the weight of children in the adjusted household size variable. Similarly, the lower the value of  $\theta$  ( $\theta < 1$ ), the more households are perceived to benefit from scale economies. May

---

<sup>17</sup> Table 25 shows that over 3.5 million of the 3.7 million rural households (94.8%) are African. The statistics presented here are therefore dominated by the characteristics of African rural households.

(1995, cited in Leibbrandt and Woolard, 1999) suggested setting  $\alpha = 0.5$  and  $\theta = 0.9$  for South Africa. Although values for  $\alpha$  and  $\theta$  estimated by Leibbrandt and Woolard are different from that used by May, they find that the poverty profile is fairly insensitive to values of  $\alpha$  and  $\theta$ . We follow May in the calculation of adult equivalence scales.<sup>18</sup>

Table 27: Adjusted household size ( $E$ ) by race and location

	Urban	Rural
African	2.75	3.45
Coloured	3.22	3.18
Asian	3.16	2.95
White	2.34	2.37

The adult equivalent income is used to disaggregate households further into various income groups (see Table 1), thus forming 30 representative households. These RHGs are saved as variable *hhgradinc*.

## 5.2. Household groups for a previous version of the series of provincial SAMs

In previous SAM versions household groups were disaggregated by province within each of the regions, and hence provincial-level household characteristics need to be considered. Here we explore some of the possibilities for forming provincial-level household and factor groups.

### 5.2.1. *Provincial-level household groups*

The previous National SAM disaggregated households first by race (African, Coloured, Asian and White) and thereafter by location (rural and urban). Thereafter each race-location sub-group was further disaggregated into a number of income groups. The number of groups formed depended loosely on the number of observations and the dispersion of income within each particular sub-group. A similar approach was followed for the previous provincial-level household groups. In cases where a particular province-race-location sub-group contained very few observations, rural and urban households were merged. In some cases two racial groups also had to be merged due to under-representation of household groups. The end result was a total of 184 RHGs in nine provinces, ranging from 14 groups in Limpopo to 26 groups in the Western and Northern Cape provinces respectively.

<sup>18</sup> During a conversation Murray Leibbrandt mentioned that there seems to be a move away from adult equivalence scales in favour of per capita welfare measures, mainly due to the arbitrariness of the selected age level below which people are regarded as 'children', as well problems surrounding the estimation of the parameters. There is also no clear evidence that the adult equivalent approach is better (or worse) than a per capita welfare measure. The current household grouping in the SAM, which is described in detail in section 3, does not use adult equivalent scales.

The number of observations in each province-race-location sub-group determined to a large extent the number of final groups that were formed. Table 28 gives a summary of the number of observations (sample level) of each of the sub-groups. As expected some race groups are poorly represented in certain locations in South Africa. Asian households in particular are not generally well represented in rural areas, with KwaZulu-Natal the only province in which more than 10 rural Asian households (sample level) were interviewed. Coloured households are also generally poorly represented in rural areas in many of the provinces. If too few households make up a household group there may be some concerns about the representativity of an RHG made up of households of that specific population-race-location sub-group. Consequently it was decided to ignore the urban-rural split when a certain province-race-location sub-group contains less than 10 observations. As a result the urban-rural split was removed for Coloured households in all provinces except the Western Cape, Eastern Cape and Northern Cape. The same was done for Asian households in all provinces except KwaZulu-Natal and for White households in Limpopo.

Also evident from Table 28 is that some racial groups as a whole are not adequately represented in certain provinces. Again the Asian race group stands out as an example, with only the Eastern Cape, KwaZulu-Natal, Gauteng and Mpumalanga containing more than 15 Asian households. Coloured households are also poorly represented in Limpopo. As a rule of thumb it was decided that a race group had to be represented by at least 15 observations (combined urban and rural) in order for it to 'qualify' as a separate province-race sub-group. The shaded cells in Table 29 show those province-race sub-groups that were merged with another province-race sub-group. In each instance Hotelling's  $T^2$ -test was used to determine which other race group's expenditure patterns match the under-represented race group's expenditure pattern closest.

Hotelling's  $T^2$ -test revealed that expenditure patterns of Asian households in the Western Cape, Northern Cape, Freestate and North West are a closer match with White household expenditure patterns than Coloured or African households. These households were consequently merged with urban White households, since the majority of the Asian households live in urban areas. The fact that Asian household incomes match White household incomes more closely is likely to be part of the reason for the similarities in expenditure patterns (see Table 30). Interestingly though, Asian households in Limpopo display expenditure patterns that match Coloured households more closely. As a result Asian and Coloured households were merged in this province. Although this merged group still only contains 14 observations it was decided to leave it in place as a separate household groups rather than merging the group with a third racial group. The statistical test results appear in the appendix (section 8.1).

Next, a decision had to be taken about the number of income groups that were to be formed within each of the sub-groups. Table 31 uses shading to show the final sub-groups (province-race-location) before the sub-groups were split into income groups. The number of observations within each sub-group was used as a guideline in this process. Any sub-group containing less than 45 households were not sub-divided into income groups. Sub-groups with 46 to 150 households were divided into three income groups around the 50<sup>th</sup> and 75<sup>th</sup> percentiles.<sup>19</sup> Sub-groups with more than 150 households were split into five income groups around the 25<sup>th</sup>, 50<sup>th</sup>, 75<sup>th</sup> and 87.5<sup>th</sup> percentiles.<sup>20</sup> Sampling weights were used throughout (IES 2000).

Table 32 shows the composition of sub-groups (race-location) and the number of sample observations found in each group. It also summarises some weighted income statistics (range, median, mean and standard deviation), while the last column shows the suggested number of income groups based on the number of observations (as explained in the previous paragraph).

---

<sup>19</sup> Duociles with the top duocile split into an upper and lower group.

<sup>20</sup> Quartiles with the top quartile split into two groups.

Table 28: Number of survey-level households per province, by race and location

	African		Coloured		Asian		White		Total
	Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural	
Western Cape	513	84	1,069	457	14		326	76	<b>2,539</b>
Eastern Cape	1,125	1,853	241	36	18		176	17	<b>3,466</b>
Northern Cape	412	80	496	126	5		145	46	<b>1,310</b>
Freestate	1,466	604	34	7	6	1	167	31	<b>2,316</b>
KwaZulu-Natal	1,692	2,052	39		371	11	242	15	<b>4,422</b>
North-West	1,286	1,299	33	6	7		144	28	<b>2,803</b>
Gauteng	3,159	79	149		64	1	479	12	<b>3,943</b>
Mpumalanga	1,017	1,127	22	2	15		81	13	<b>2,277</b>
Limpopo	795	2,206	3	1	9	1	77	9	<b>3,101</b>
<b>Total</b>	<b>11,465</b>	<b>9,384</b>	<b>2,086</b>	<b>635</b>	<b>509</b>	<b>14</b>	<b>1,837</b>	<b>247</b>	<b>26,177</b>

Table 29: Merging urban-rural households within certain race groups

	African		Coloured		Asian		White		Total
	Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural	
Western Cape	513	84	1,069	457	14		326	76	<b>2,539</b>
Eastern Cape	1,125	1,853	241	36	18		176	17	<b>3,466</b>
Northern Cape	412	80	496	126	5		145	46	<b>1,310</b>
Freestate	1,466	604	41		7		167	31	<b>2,316</b>
KwaZulu-Natal	1,692	2,052	39		371	11	242	15	<b>4,422</b>
North-West	1,286	1,299	39		7		144	28	<b>2,803</b>
Gauteng	3,159	79	149		65		479	12	<b>3,943</b>
Mpumalanga	1,017	1,127	24		15		81	13	<b>2,277</b>
Limpopo	795	2,206	4		10		86		<b>3,101</b>

Table 30: Mean income of households per province, by race and location (weighted)

	African		Coloured		Asian		White	
	Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural
Western Cape	29,934	19,200	59,781	23,540	112,507		169,593	126,518
Eastern Cape	32,105	12,822	39,428	12,911	145,208		147,212	136,840
Northern Cape	23,842	23,673	33,564	17,779	201,713		174,422	221,823
Freestate	24,972	12,588	41,615	14,084	66,767	72,000	131,083	410,358
KwaZulu-Natal	29,523	14,608	60,814		76,656	23,595	182,217	124,147
North-West	32,735	19,720	47,146	23,401	116,867		132,685	365,024
Gauteng	35,262	22,303	68,442		121,880	50,000	167,748	133,234
Mpumalanga	34,539	17,814	51,733	10,677	113,010		156,521	98,018
Limpopo	44,087	17,074	92,511	6,000	121,998	10,800	162,248	224,688

Table 31: Final sub-groups before splitting by income

	African		Coloured		Asian		White		Total
	Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural	
Western Cape	513	84	1,069	457	340			76	<b>2,539</b>
Eastern Cape	1,125	1,853	241	36	18		176	17	<b>3,466</b>
Northern Cape	412	80	496	126	150			46	<b>1,310</b>
Freestate	1,466	604	41		174			31	<b>2,316</b>
KwaZulu-Natal	1,692	2,052	39		371	11	242	15	<b>4,422</b>
North-West	1,286	1,299	39		151			28	<b>2,803</b>
Gauteng	3,159	79	149		65		479	12	<b>3,943</b>
Mpumalanga	1,017	1,127	24		15		81	13	<b>2,277</b>
Limpopo	795	2,206	14				86		<b>3,101</b>

Table 32: Composition of sub-groups, number of observations and suggested number of income groups within each sub-group

<b>Western Cape</b>		<b>African</b>		<b>Coloured</b>		<b>Asian</b>		<b>White</b>				
	<b>Subgroup No.</b>	<b>Urban</b>	<b>Rural</b>	<b>Urban</b>	<b>Rural</b>	<b>Urban</b>	<b>Rural</b>	<b>Urban</b>	<b>Rural</b>	<b>Total</b>	<b>Weighted no. of obs</b>	<b># income groups</b>
	11	513								513	234,019	5
	12		84							84	19,441	3
	13			1,069						1,069	467,627	5
	14				457					457	81,813	5
	15					14		326		340	232,061	5
	16								76	76	18,522	3
										<b>2,539</b>	<b>1,053,484</b>	<b>26</b>

<b>Eastern Cape</b>		<b>African</b>		<b>Coloured</b>		<b>Asian</b>		<b>White</b>				
	<b>Subgroup No.</b>	<b>Urban</b>	<b>Rural</b>	<b>Urban</b>	<b>Rural</b>	<b>Urban</b>	<b>Rural</b>	<b>Urban</b>	<b>Rural</b>	<b>Total</b>	<b>Weighted no. of obs</b>	<b># income groups</b>
	21	1,125								1,125	389,460	5
	22		1,853							1,853	859,403	5
	23			241						241	80,629	5
	24				36					36	15,810	1
	25					18				18	7,230	1
	26							176		176	79,529	5
	27								17	17	8,354	1
										<b>3,466</b>	<b>1,440,414</b>	<b>23</b>

<b>Northern Cape</b>		<b>African</b>		<b>Coloured</b>		<b>Asian</b>		<b>White</b>				
	<b>Subgroup No.</b>	<b>Urban</b>	<b>Rural</b>	<b>Urban</b>	<b>Rural</b>	<b>Urban</b>	<b>Rural</b>	<b>Urban</b>	<b>Rural</b>	<b>Total</b>	<b>Weighted no. of obs</b>	<b># income groups</b>
	31	412								412	56,637	5
	32		80							80	11,144	3
	33			496						496	70,414	5
	34				126					126	15,851	5
	35					5		145		150	26,594	5
	36								46	46	6,607	3
										<b>1,310</b>	187,247	<b>26</b>

<b>Freestate</b>		<b>African</b>		<b>Coloured</b>		<b>Asian</b>		<b>White</b>				
	<b>Subgroup No.</b>	<b>Urban</b>	<b>Rural</b>	<b>Urban</b>	<b>Rural</b>	<b>Urban</b>	<b>Rural</b>	<b>Urban</b>	<b>Rural</b>	<b>Total</b>	<b>Weighted no. of obs</b>	<b># income groups</b>
	41	1,466								1,466	442,237	5
	42		604							604	146,838	5
	43			34	7					41	12,859	1
	44					6	1	167		174	87,438	5
	45								31	31	8,877	1
										<b>2,316</b>	698,247	<b>17</b>

<b>KwaZulu-Natal</b>		<b>African</b>		<b>Coloured</b>		<b>Asian</b>		<b>White</b>				
	<b>Subgroup No.</b>	<b>Urban</b>	<b>Rural</b>	<b>Urban</b>	<b>Rural</b>	<b>Urban</b>	<b>Rural</b>	<b>Urban</b>	<b>Rural</b>	<b>Total</b>	<b>Weighted no. of obs</b>	<b># income groups</b>
	51	1,692								1,692	853,386	5
	52		2,052							2,052	832,813	5
	53			39						39	19,614	1
	54					371				371	189,970	5
	55						11			11	5,759	1
	56							242		242	148,349	5
	57								15	15	5,588	1
										<b>4,422</b>	<b>2,055,479</b>	<b>23</b>

<b>North-West</b>		<b>African</b>		<b>Coloured</b>		<b>Asian</b>		<b>White</b>				
	<b>Subgroup No.</b>	<b>Urban</b>	<b>Rural</b>	<b>Urban</b>	<b>Rural</b>	<b>Urban</b>	<b>Rural</b>	<b>Urban</b>	<b>Rural</b>	<b>Total</b>	<b>Weighted no. of obs</b>	<b># income groups</b>
	61	1,286								1,286	327,242	5
	62		1,299							1,299	408,700	5
	63			33	6					39	9,841	1
	64					7		144		151	42,599	5
	65								28	28	5,970	1
										<b>2,803</b>	<b>794,352</b>	<b>17</b>

<b>Gauteng</b>												
	Subgroup No.	African		Coloured		Asian		White		Total	Weighted no. of obs	# income groups
		Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural			
	71	3,159								3,159	2,342,999	5
	72		79							79	73,435	3
	73			149						149	97,250	5
	74					64	1			65	44,050	3
	75							479		479	497,223	5
	76								12	12	9,385	1
										<b>3,943</b>	3,064,341	<b>22</b>

<b>Mpumalanga</b>												
	Subgroup No.	African		Coloured		Asian		White		Total	Weighted no. of obs	# income groups
		Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural			
	81	1,017								1,017	262,731	5
	82		1,127							1,127	334,423	5
	83			22	2					24	6,410	1
	84					15				15	4,356	1
	85							81		81	37,187	3
	86								13	13	3,304	1
										<b>2,277</b>	648,410	<b>16</b>

Limpopo	Subgroup No.	African		Coloured		Asian		White		Total	Weighted no. of obs	# income groups
		Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural			
	91	795								795	155,381	5
	92		2,206							2,206	850,882	5
	93			3	1	9	1			14	3,189	1
	94							77	9	86	21,728	3
										<b>3,101</b>	1,031,180	<b>14</b>

<b>Grand total</b>	<b>26,177</b>	10,973,154	<b>184</b>
--------------------	---------------	------------	------------

### 5.2.2. Provincial-level factor groups

The previous National SAM disaggregated factors along racial lines and then further divided these groups into one of 11 factor categories as specified in Table 15. A similar approach was followed for the previous version of the provincial-level factor groups, except that factors were first disaggregated by province. Labour income data was used in the formation of two sub-matrices, the value-added sub-matrix (factors-activities) and the functional distribution sub-matrix (households-factors). In the IES 2000 dataset households are the only beneficiaries of wage income, i.e. total value added equals total household income from factors.

As was previously the case in the formation of household groups, there were some concerns about the representativity of certain province-race-factor sub-groups. Table 33 shows the number of observations (survey level) for each occupation type by race and province. Only those workers reporting positive wage income are included in the table.

The small number of Asian workers, particularly outside of KwaZulu-Natal, made the fairly detailed disaggregation into eleven factor groups slightly problematic. The same could be said of Coloured workers in some of the provinces. Looking at relative numbers of workers, one also sees a clear pattern emerging whereby Coloured and African workers tend to be classified more as skilled or semi- and unskilled, while relatively large numbers of White workers are skilled or highly skilled.

In order to maintain consistency it was decided to merge Asian workers with their White counterparts in the Western Cape, Northern Cape, Free State and the North West and with Coloured workers in Limpopo. This ensured that the racial classification in each province was similar for households and factors. Table 34 shows the number of workers by race and province. The shaded groups show those cells that are merged.

Table 33: Number of wage-earning workers per race, province and occupation type

Description	Western Cape	Eastern Cape	Northern Cape	FreeState	KwaZulu-Natal	North-West	Gauteng	Mpumalanga	Limpopo
African Legislators senior officials and managers	7	29	7	13	35	30	70	23	29
African Professionals	16	43		26	47	36	85	23	105
African Technicians and associate professionals	19	183	14	96	211	165	233	128	202
African Clerks	22	95	27	68	126	127	281	82	102
African Service workers and shop market sales workers	87	164	43	155	278	174	453	191	174
African Skilled agricultural and fishery workers	29	104	15	96	151	83	70	105	103
African Craft and related trades workers	79	153	79	292	307	297	482	271	225
African Plant and machine operators and assemblers	34	141	51	379	352	286	477	225	165
African Elementary occupations	212	372	143	375	734	329	526	460	373
African Domestic Workers	85	246	64	250	354	256	450	196	178
African Unspecified	43	116	31	121	238	132	240	156	164
Coloured Legislators senior officials and managers	45	3	7	1	1	1	5	1	1
Coloured Professionals	48	4	4	1			13		
Coloured Technicians and associate professionals	125	22	32	3	3	6	21	3	
Coloured Clerks	181	32	39	8	14	14	38	4	
Coloured Service workers and shop market sales workers	171	31	61	9	10	4	20	5	
Coloured Skilled agricultural and fishery workers	56	11	24	2		1			
Coloured Craft and related trades workers	242	42	74	4	11	3	29	7	1
Coloured Plant and machine operators and assemblers	253	45	48	2	2	2	11		
Coloured Elementary occupations	912	81	238	6	2	3	16	2	1
Coloured Domestic Workers	160	31	110	3	1	2	3	1	
Coloured Unspecified	228	16	46	2	3	2	10	4	

Asian Legislators senior officials and managers	4	7		4	34	2	17	4	4
Asian Professionals		6	1	1	32		11	2	1
Asian Technicians and associate professionals	3	3	1	1	56	2	14	3	
Asian Clerks	1	3	1	1	84	1	11	2	2
Asian Skilled agricultural and fishery workers	4	5	2		49	4	12	1	9
Asian Service workers and shop market sales workers					3				
Asian Craft and related trades workers			1	1	66		3	3	
Asian Plant and machine operators and assemblers	1	1			60		1		
Asian Elementary occupations		1			36		3	2	
Asian Domestic workers					3				
Asian Unspecified	1	1	1	1	33	1	9	1	1
White Legislators senior officials and managers	75	35	19	32	48	10	106	10	13
White Professionals	53	23	14	29	35	13	99	9	12
White Technicians and associate professionals	73	36	30	23	58	20	112	21	13
White Clerks	67	58	47	67	50	26	115	13	30
White Service workers and shop market sales workers	47	24	27	16	22	10	73	7	4
White Skilled agricultural and fishery workers	17	6	30	12	7	4	2	4	
White Craft and related trades workers	37	24	30	29	21	32	57	25	15
White Plant and machine operators and assemblers	8	11	5	9	9	7	9	1	9
White Elementary occupations	19	1	6	7	8	5	14	5	2
White Domestic workers	1			1			2		
White Unspecified	44	17	18	22	22	13	42	10	8

Table 34: Workers by race and province (summary table)

	Western Cape	Eastern Cape	Northern Cape	Freestate	KwaZulu-Natal	North-West	Gauteng	Mpumalanga	Limpopo
African	633	1,646	474	1,871	2,833	1,915	3,367	1,860	1,820
Coloured	2,421	318	683	41	47	38	166	27	3
Asian	14	27	7	9	456	10	81	18	17
White	441	235	226	247	280	140	631	105	106
<i>Percentages</i>									
African	18.04%	73.94%	34.10%	86.30%	78.35%	91.06%	79.32%	92.54%	93.53%
Coloured	68.99%	14.29%	49.14%	1.89%	1.30%	1.81%	3.91%	1.34%	0.15%
Asian	0.40%	1.21%	0.50%	0.42%	12.61%	0.48%	1.91%	0.90%	0.87%
White	12.57%	10.56%	16.26%	11.39%	7.74%	6.66%	14.86%	5.22%	5.45%

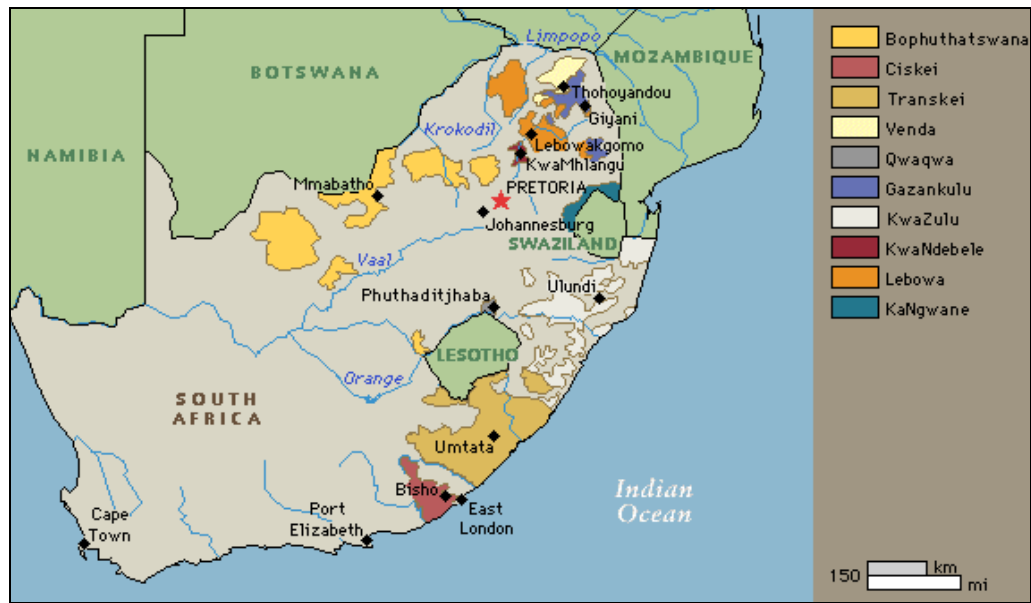
### 5.3. Household groups with a geographical- and agricultural focus

Various other permutations of households groups can be formed. An array of classification variables was created. Section 5.3.1 shows how a variable was created that indicates which households live in areas formerly classified as homelands areas (only African households). Two variables indicating whether households are classified as agricultural households or non-agricultural households (broad defined and strictly defined) are discussed in section 5.3.2. The normal *location* variable (urban/rural) is modified so that urban areas are split into metropolitan areas and secondary cities/small towns (see section 5.3.3). Finally, section 5.3.4 explains how magisterial districts in the IES 2000 were mapped to so-called ‘nodal areas’ for the implementation of a governmental rural development program.

#### 5.3.1. Former homelands areas

During the 1960s and 70s the South African government, as part of their Apartheid policy, set aside various areas known as homelands. The homelands would typically be made up of Africans of a specific ethnic group, depending on the geographic positioning and dominant ethnic group of the region. Figure 11 shows the ten homelands areas that existed in South Africa. Transkei, Boputhatswana, Venda and Ciskei (collectively referred to as the TBVC states) were the most prominent of the homelands. Homelands were either partially self-governed or in some cases independent from the Republic. The former homelands areas constitute less than 13% of the total land area of South Africa, but is still today home to 27.1% of the population and more than one third of all Africans (IES 2000). Given decades of under funding, poor management, and economic and geographical isolation, it can be expected that households in homelands areas will behave differently to economic shocks. The proposed household grouping therefore separates out households living in former homelands areas.

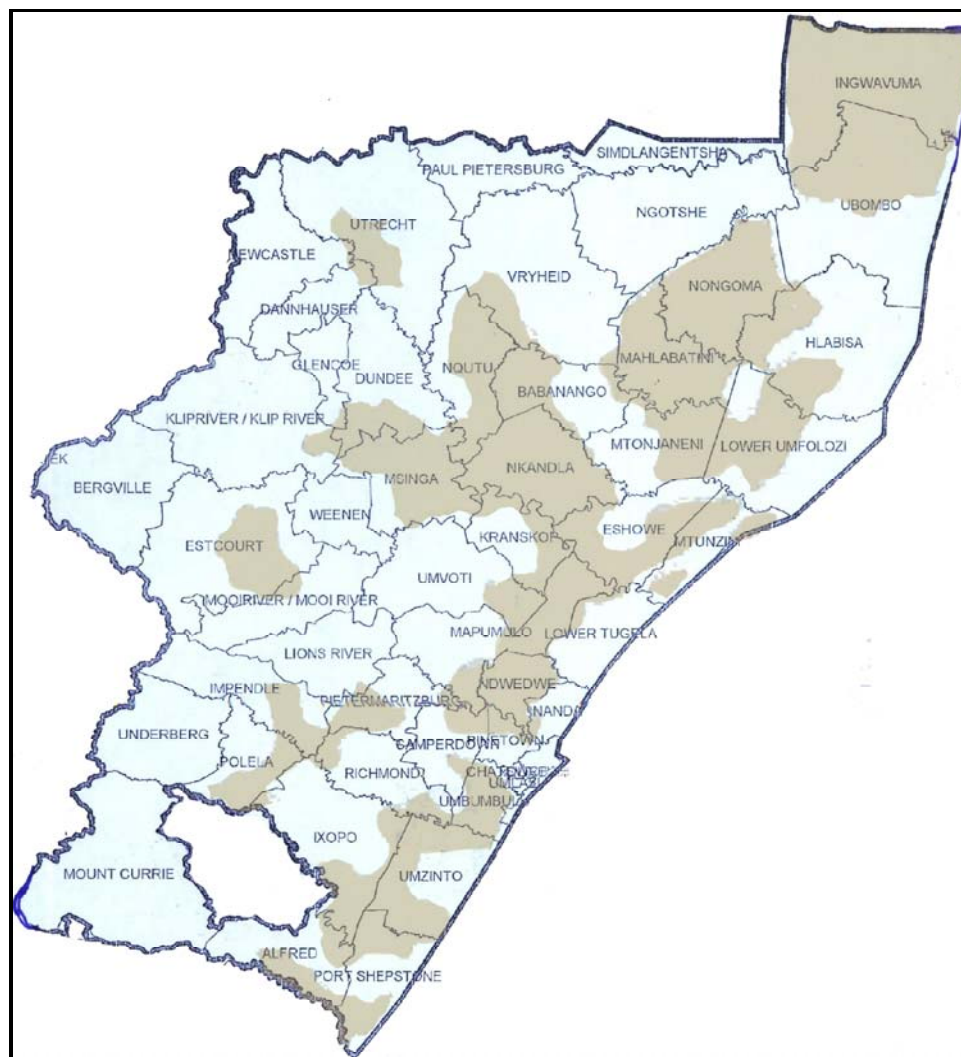
Figure 11: Former homelands in South Africa



Source: Unknown

Although homelands do not exist any longer today, all these areas, with the exception of KwaZulu, can easily be mapped to the 2000 magisterial boundaries that demarcate magisterial districts today. In KwaZulu, however, the former homelands boundaries are not the same as the 2000 magisterial district boundaries. By overlaying maps of the current magisterial districts in KwaZulu-Natal and an old map showing the KwaZulu homeland boundaries, one can see (approximately) which magisterial districts were formerly mostly or entirely part of KwaZulu (see Figure 12). The Stata do-file *households.do* shows how the magisterial districts were mapped to the homelands areas. A full listing of magisterial districts and the related codes is available from the author.

Figure 12: Magisterial districts and homeland areas in KwaZulu-Natal



Source: [http://www.lib.utexas.edu/maps/map\\_collection\\_guide.html](http://www.lib.utexas.edu/maps/map_collection_guide.html) and National Department of Land Affairs: Surveys and Mappings. Graphic work by Jacques Murdoch.

Note: The shaded areas represent areas formerly part of KwaZulu

### 5.3.2. *Agricultural households*

The suggested household grouping distinguishes between agricultural and non-agricultural households. Both the income and expenditure sides of the household accounts are used to try and determine which households can be declared agricultural households.

On the expenditure side information on home production for home consumption (HPHC) is used. Household that are involved in HPHC to such an extent that expenditure on inputs plus the value of home consumption (variable *hhhphc*) makes up 50% or more of total food expenditure, are declared agricultural households. Total food expenditure is defined here as the sum of normal food expenditure (variable *Cfood*) plus *hhhphc*. The average expenditure

on home production for home consumption as a share of food expenditure is 13.9% for those 5464 households that do produce agricultural goods for own consumption (variable *hphcshfood*). Only 321 households spend more than 50% of their food budget on home production for home consumption. These households are considered agricultural households.

```
. sum hphcshfood if hphcshfood > 0
```

Variable	Obs	Mean	Std. Dev.	Min	Max
hphcshfood	5464	.1387389	.1759424	.0001756	1

```
. sum hphcshfood if hphcshfood > 0.5
```

Variable	Obs	Mean	Std. Dev.	Min	Max
hphcshfood	321	.6714229	.1445378	.5001907	1

Income from agricultural related activities is defined as follows. All household members that earn income from labour (wages and salaries) either due to being employed as a skilled agricultural worker (*factnorace* = 6) or as an employee in the service of the agricultural sector (*activities* = 1) are considered. These individual incomes are added up to give a total household-level wage income from agricultural related activities (variable *sumaginclab*). This income measure is then expressed as a share of total household income (variable *sumaginclabpsh*). Income from the sale of home produce, expressed as a share of total household income (variable *inchphcsh*), is added to the wage income share to give an indication of the share of income from agricultural related activities (variable *agincsh*).

On average those 2702 households that have members that are employed in the agricultural sector or as skilled agricultural workers earn about 72.4% of their income from this source (variable *sumaginclabpsh*). About 1007 households earn income from the sale of home produce, contributing on average 4.43% to household income (variable *inchphcsh*). The sum of these two items (variable *agincsh*) indicates that those 3582 households that earn income from agricultural related activities earn an average of 55.9% of their income from this source. About 2066 of these households earn more than 50% of their income from this source and are declared agricultural households.

```
. sum sumaginclabpsh if sumaginclabpsh > 0
```

Variable	Obs	Mean	Std. Dev.	Min	Max
sumaginclabpsh	2702	.7240944	.2955603	.001408	1

```
. sum inchphcsh if inchphcsh > 0
```

Variable	Obs	Mean	Std. Dev.	Min	Max
inchphcsh	1007	.0443948	.1168926	.000031	1

```
. sum agincsh if agincsh > 0
```

Variable	Obs	Mean	Std. Dev.	Min	Max
agincsh	3582	.5586847	.3942237	.000031	1

```
. sum agincsh if agincsh > 0.5
```

Variable	Obs	Mean	Std. Dev.	Min	Max
agincsh	2066	.8654988	.1627271	.5009634	1

Agricultural households are identified by variable *agrigh*, which is created by the following Stata command:

```
gen agrigh = 1 if hphcshfood > 0.5 | agincsh > 0.5;
```

As shown below a total of 2338 households – 8.9% of households – are defined as agricultural households (sample level).<sup>21</sup> When sampling weights are used there are approximately 711771 agricultural households (6.5%). This suggests that agricultural households were over-sampled in IES 2000. Agricultural households earn an average income of R22819 compared to non-agricultural households' average income of R46587.

```
. tab agrigh
```

Agricultural households	Freq.	Percent	Cum.
non-agric	23839	91.07	91.07
agricultural	2338	8.93	100.00
Total	26177	100.00	

```
. tab agrigh [aweight = wgtselect], sum(totinc)
```

Agricultural households	Summary of Total household income			Obs.
	Mean	Std. Dev.	Freq.	
non-agric	46587.422	109364.55	10261384	23839
agricultu	22818.896	95426.601	711770.58	2338
Total	45045.683	108671.31	10973154	26177

### 5.3.3. Metropolitan Households

A third major proposed household sub-group is urban-metropolitan areas. The Local Government Municipal Structures Act (1998), as directed by the Constitution, makes provision for three types of municipalities. These are metropolitan municipalities (Category A), local municipalities (Category B), and district areas or municipalities (Category C). Category A municipalities can only be established in metropolitan areas. Using population statistics and production and employment data the Municipal Demarcation Board compiled a

<sup>21</sup> The number 2338 is slightly less than the sum of 2066 and 321 since some households fall into the agricultural household category under both the income-side and expenditure-side definitions.

list of ten places that should be considered potential candidates for metropolitan areas.<sup>22</sup> These ten areas were (ranked according to size): Johannesburg, Durban, Pretoria, Cape Town, East Rand, Port Elizabeth, Vaal, East London, Pietermaritzburg and Bloemfontein. If Botshabelo were included under Bloemfontein this area would appear higher on the list. After consideration the Board decided that Pretoria (Tshwane), Johannesburg, East Rand (Ekurhuleni), Durban (eThekweni), Cape Town and Port Elizabeth (Nelson Mandela) be declared metropolitan areas, and hence the six metropolitan municipalities in South Africa were formed in these areas.<sup>23</sup>

Next, the process of mapping metropolitan municipalities to magisterial districts in the IES 2000 was embarked on. Since the smallest geographical areas in the IES 2000 data are magisterial districts it is necessary to determine which magisterial districts fall in the larger metropolitan areas. Often metropolitan areas cut through the middle of magisterial districts, especially in the Gauteng province, so the process is not straightforward. Using maps from various municipal Internet websites and a map of the South African magisterial district boundaries the following mapping was used decided on:

- *City of Tshwane*: Pretoria, Wonderboom and Shoshanguve magisterial districts.<sup>24</sup>
- *City of Johannesburg*: Johannesburg, Roodepoort, Soweto and Randburg magisterial districts.<sup>25</sup>
- *Ekurhuleni Metropolitan Municipality*: Kemptonpark, Germiston, Alberton, Boksburg, Brakpan, Benoni, Springs and Nigel magisterial districts.<sup>26</sup>
- *eThekweni Municipality*: Comprises mainly of the Durban magisterial district.<sup>27</sup>
- *City of Cape Town*: Currently comprises of Bellville, Goodwood, Cape Town, Simon's Town, Wineberg, Mitchell's Plain, Kuilsriver, Somerset West and Strand magisterial districts.<sup>28</sup>
- *Nelson Mandela Metropolitan Municipality*: Uitenhage and Port Elizabeth magisterial districts.<sup>29</sup>

---

<sup>22</sup> See [http://www.local.gov.za/DCD/dcdlibrary/dma/dma\\_prelim.htm](http://www.local.gov.za/DCD/dcdlibrary/dma/dma_prelim.htm) and <http://www.demarcation.org.za/municprofiles2003/index.asp>

<sup>23</sup> See <http://www.info.gov.za/structure/local-gov.htm>.

<sup>24</sup> See <http://www.tshwane.gov.za>.

<sup>25</sup> See <http://www.joburg.org.za>.

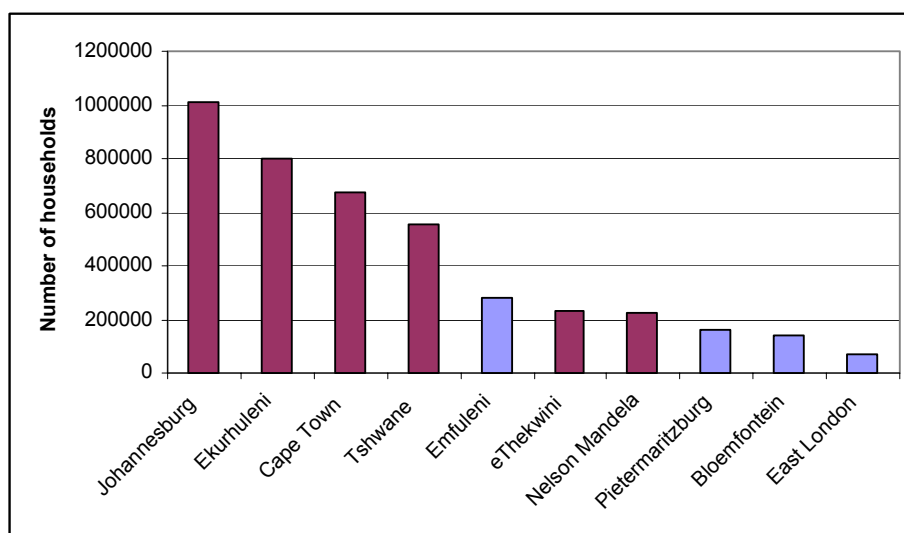
<sup>26</sup> From the graphics it appears as if only half of the Nigel magisterial district is included in this metropolitan municipality. See <http://www.ekurhuleni.com/ekurhuleni/index.jsp>.

<sup>27</sup> See <http://www.durban.gov.za/eThekweni/>.

<sup>28</sup> The Paarl and Wellington magisterial districts will probably be added to the metropolitan area within the near future. See <http://www.capetown.gov.za>.

The Demarcation Board also investigated the following areas for possible inclusion into the final list of metropolitan areas: Vaal, East London, Pietermaritzburg and Bloemfontein (including Botshabelo). These areas are also mapped to magisterial districts. The Emfuleni (or Lekoa-Vaal) municipality in the Vaal metropolitan region comprises of the Vanderbijlpark and Vereeniging magisterial districts. The remaining large urban areas of East London, Pietermaritzburg and Bloemfontein/Botshabelo can all be mapped directly to their similarly named magisterial districts.

Figure 13: Municipalities/Cities and population size (number of households)



Note: Lightly shaded bars represent municipal districts/cities not classified as Category A municipalities (metropolitan municipalities).

Source: IES 2000

#### 5.3.4. Nodal areas

During his State of the Nation address in 2001 President Thabo Mbeki identified 13 municipal areas that would be targeted for rural development areas. These 13 municipalities were called “nodal areas” for the implementation of these programmes (see Figure 14). These municipal areas are also mapped to the magisterial districts in order to identify households that fall with these areas, although these areas are not taken into account for the proposed household account disaggregation.<sup>30</sup>

- *Western Cape*: Central Karoo
- *Eastern Cape*: Chris Hani,<sup>31</sup> Amatole, Ukhahlamba and O.R. Tambo

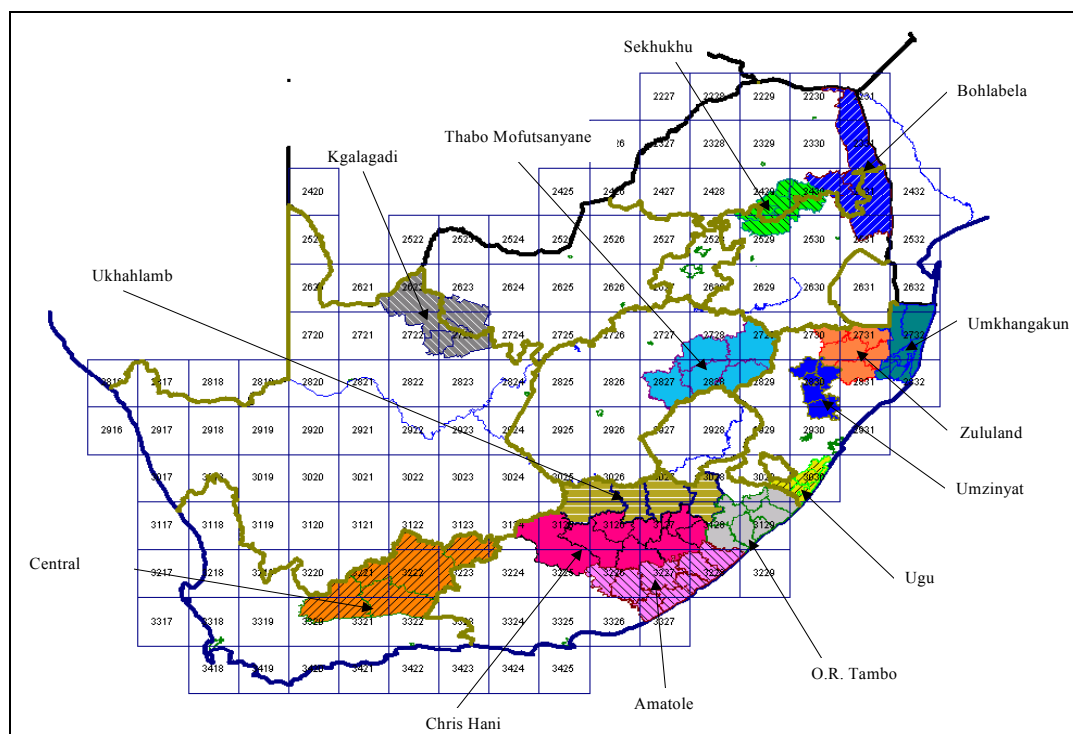
<sup>29</sup> See <http://www.routes.co.za/municipalities/ec/nelsonmandela.html>.

<sup>30</sup> The mapping file is available from the author.

<sup>31</sup> Formerly North East

- *Free State*: Thabo Mofutsanyane.
- *KwaZulu-Natal*: Ugu, Umzinyathi, Zululand District and Umkhangakunde.
- *Mpumalanga/Limpopo*: Bohlabela and Sekhukhune.<sup>32</sup>
- *Northern Cape/North West*: Kgalagadi.<sup>33</sup>

Figure 14: Nodal areas



Source: Chief Directorate Surveys and Mappings, Department of Land Affairs

## 6. Concluding remarks

This technical discussion outlined the steps taken for form representative household and factor groups for the South African PROVIDE SAM. When forming such household and factor groups the aim should be to group households and factors with similar preferences and characteristics. This ensures that the assumption that each household or factor group member is affected in the same way by a policy shock is not too unrealistic. In the pursuit of reducing intra-group heterogeneity there is a temptation to form large numbers of household or factor groups. However, there is a conflict between having large numbers of household and factor groups (which ensure a greater degree of homogeneity within the groups) and limiting the

<sup>32</sup> Both these municipal districts are ‘transfrontier’ municipalities, i.e. they stretch across the border between Limpopo and Mpumalanga.

<sup>33</sup> Kgalagadi is also a ‘transfrontier’ municipality.

number of groups (which keeps data analysis manageable). However, since it is always possible to aggregate household and factor groups at a later stage (and not *vice versa*) the approach here was to form as many groups as the data allowed.

## 7. References

- Arndt, C., Cruz, A., Jensen, H.T., Robinson, S. and Tarp, F. (1998). "Social Accounting Matrices for Mozambique 1994 and 1995," *International Food Policy Research Institute (IFPRI): Trade and Macroeconomics Division Discussion Paper*, 28.
- Bhorat, H. and Leibbrandt, M. (1996). "Understanding Unemployment. The Relationship between the Employed and the Jobless." In *Against the Current. Labour and Economic Policy in South Africa*, edited by Baskin, J. Randburg: NALEDI.
- Bourguignon, F., Robilliard, A. and Robinson, S. (2002). "Representative versus real households in the macro-economic modelling of inequality," *Mimeo*.
- Chulu, O. and Wobst, P. (2001). "A 1998 Social Accounting Matrix for Malawi," *International Food Policy Research Institute (IFPRI): Trade and Macroeconomics Division Discussion Paper*, 69.
- Cockburn, J. (2001). "Trade Liberalisation and Poverty in Nepal. A Computable General Equilibrium Micro Simulation Analysis," *Mimeo*.
- Decaluwé, B., Patry, A., Savard, L. and Thorbecke, E. (1999). "Poverty Analysis Within a General Equilibrium Framework," *CRÉFA Working Paper*, 9909.
- Fontana, M. and Wobst, P. (2001). "A Gendered 1993-94 Social Accounting Matrix for Bangladesh," *International Food Policy Research Institute (IFPRI): Trade and Macroeconomics Division Discussion Paper*, 74.
- Leibbrandt, M. and Woolard, I. (1999). "Household Incomes, Poverty and Inequality in a Multivariate Framework," *DPRU Working Papers*, 99(31).
- Lewis, J.D. (2001). "Policies to Promote Growth and Employment in South Africa." World Bank, Washington.
- May, J. (1998). "Poverty and Inequality in South Africa," *Report prepared for the Office of the Executive Deputy President and the Inter-Ministerial Committee for Poverty and Inequality*.
- McDonald, S. (2002). "The Commodity Protocols of the Lomé Convention: The Case of Beef Exports from Botswana," *Journal of Agricultural Economics*, 53(2): 407-427.
- McDonald, S. and Punt, C. (2001). "A Social Accounting Matrix for the Western Cape." Department of Agriculture, Western Cape Government.
- PROVIDE (2003). "Social Accounting Matrices and Economic Modelling," *PROVIDE Background Paper Series*, 2003:4. PROVIDE Project, Elsenburg. Available online at [www.elsenburg.com/provide](http://www.elsenburg.com/provide).
- PROVIDE (2005a). "Compiling a Social Accounting Matrix for South Africa: 2000," *PROVIDE Technical Paper Series*, Forthcoming. PROVIDE Project, Elsenburg.
- PROVIDE (2005b). "Creating an IES-LFS 2000 Database in Stata," *PROVIDE Technical Paper Series*, 2005:1. PROVIDE Project, Elsenburg. Available online at [www.elsenburg.com/provide](http://www.elsenburg.com/provide).
- SSA (1999). *Agricultural Survey, 1996*, Pretoria: Statistics South Africa.
- SSA (2002a). *Income and Expenditure Survey 2000*, Pretoria: Statistics South Africa.
- SSA (2002b). *Labour Force Survey September 2000*, Pretoria: Statistics South Africa.
- Thomas, M. and Bautista, R.M. (1999). "A 1991 Social Accounting Matrix (SAM) for Zimbabwe," *International Food Policy Research Institute (IFPRI): Trade and Macroeconomics Division Discussion Paper*, 36.
- Thurlow, J. and Van Seventer, D.E. (2002). "A Standard Computable General Equilibrium Model for South Africa," *International Food Policy Research Institute: TMD Discussion Paper*, 100.
- Thurlow, J. and Wobst, P. (2003). "Poverty-Focused Social Accounting Matrices for Tanzania," *International Food Policy Research Institute (IFPRI): Trade and Macroeconomics Division Discussion Paper*, 112.

## 8. Appendix:

### 8.1. Hypothesis testing (Hotelling's T<sup>2</sup>-test)<sup>34</sup>

```
. hotel g* if prov == 1 & (race == 1 | race == 3) [weight = wgtselect],
by(race) notab;
(analytic weights assumed)
```

```
2-group Hotelling's T-squared = 66.657988
F test statistic: ((611-10-1)/(611-2)(10)) x 66.657988 = 6.5672894
```

```
H0: Vectors of means are equal for the two groups
      F(10,600) = 6.5673
      Prob > F(10,600) = 0.0000
```

```
. hotel g* if prov == 1 & (race == 2 | race == 3) [weight = wgtselect],
by(race) notab;
(analytic weights assumed)
```

```
2-group Hotelling's T-squared = 28.071747
F test statistic: ((1540-10-1)/(1540-2)(10)) x 28.071747 = 2.7907478
```

```
H0: Vectors of means are equal for the two groups
      F(10,1529) = 2.7907
      Prob > F(10,1529) = 0.0020
```

```
. hotel g* if prov == 1 & (race == 4 | race == 3) [weight = wgtselect],
by(race) notab;
(analytic weights assumed)
```

```
2-group Hotelling's T-squared = 4.7966495
F test statistic: ((416-10-1)/(416-2)(10)) x 4.7966495 = .46923746
```

```
H0: Vectors of means are equal for the two groups
      F(10,405) = 0.4692
      Prob > F(10,405) = 0.9096
```

```
. hotel g* if prov == 2 & (race == 1 | race == 3) [weight = wgtselect],
by(race) notab;
(analytic weights assumed)
```

```
2-group Hotelling's T-squared = 111.87596
F test statistic: ((2996-10-1)/(2996-2)(10)) x 111.87596 = 11.153966
```

```
H0: Vectors of means are equal for the two groups
      F(10,2985) = 11.1540
      Prob > F(10,2985) = 0.0000
```

```
. hotel g* if prov == 2 & (race == 2 | race == 3) [weight = wgtselect],
by(race) notab;
(analytic weights assumed)
```

```
2-group Hotelling's T-squared = 68.554959
F test statistic: ((295-10-1)/(295-2)(10)) x 68.554959 = 6.6449175
```

```
H0: Vectors of means are equal for the two groups
      F(10,284) = 6.6449
      Prob > F(10,284) = 0.0000
```

---

<sup>34</sup> All racial combinations are tested. In cases where more than one of the null hypotheses were rejected the largest probability level was assumed to be the 'stronger' result. IES 2000 survey weights are used throughout.

```
. hotel g* if prov == 2 & (race == 4 | race == 3) [weight = wgtselect],
by(race) notab;
(analytic weights assumed)
```

```
2-group Hotelling's T-squared = 9.2901708
F test statistic: ((211-10-1)/(211-2)(10)) x 9.2901708 = .88901156
```

```
H0: Vectors of means are equal for the two groups
      F(10,200) = 0.8890
      Prob > F(10,200) = 0.5444
```

```
. hotel g* if prov == 3 & (race == 1 | race == 3) [weight = wgtselect],
by(race) notab;
(analytic weights assumed)
```

```
2-group Hotelling's T-squared = 40.810605
F test statistic: ((497-10-1)/(497-2)(10)) x 40.810605 = 4.0068594
```

```
H0: Vectors of means are equal for the two groups
      F(10,486) = 4.0069
      Prob > F(10,486) = 0.0000
```

```
. hotel g* if prov == 3 & (race == 2 | race == 3) [weight = wgtselect],
by(race) notab;
(analytic weights assumed)
```

```
2-group Hotelling's T-squared = 31.25241
F test statistic: ((627-10-1)/(627-2)(10)) x 31.25241 = 3.0802376
```

```
H0: Vectors of means are equal for the two groups
      F(10,616) = 3.0802
      Prob > F(10,616) = 0.0008
```

```
. hotel g* if prov == 3 & (race == 4 | race == 3) [weight = wgtselect],
by(race) notab;
(analytic weights assumed)
```

```
2-group Hotelling's T-squared = 7.4763457
F test statistic: ((196-10-1)/(196-2)(10)) x 7.4763457 = .71295049
```

```
H0: Vectors of means are equal for the two groups
      F(10,185) = 0.7130
      Prob > F(10,185) = 0.7116
```

```
. hotel g* if prov == 4 & (race == 1 | race == 3) [weight = wgtselect],
by(race) notab;
(analytic weights assumed)
```

```
2-group Hotelling's T-squared = 31.989866
F test statistic: ((2077-10-1)/(2077-2)(10)) x 31.989866 = 3.1851115
```

```
H0: Vectors of means are equal for the two groups
      F(10,2066) = 3.1851
      Prob > F(10,2066) = 0.0005
```

```
. hotel g* if prov == 4 & (race == 2 | race == 3) [weight = wgtselect],
by(race) notab;
(analytic weights assumed)
```

```
2-group Hotelling's T-squared = 26.263809
F test statistic: ((48-10-1)/(48-2)(10)) x 26.263809 = 2.1125238
```

```
H0: Vectors of means are equal for the two groups
      F(10,37) = 2.1125
```

Prob > F(10,37) = 0.0485

```
. hotel g* if prov == 4 & (race == 4 | race == 3) [weight = wgtselect],
by(race) notab;
(analytic weights assumed)
```

2-group Hotelling's T-squared = 7.080227  
 F test statistic: ((205-10-1)/(205-2)(10)) x 7.080227 = .67663253

H0: Vectors of means are equal for the two groups

F(10,194) = 0.6766  
 Prob > F(10,194) = 0.7454

```
. hotel g* if prov == 5 & (race == 1 | race == 3) [weight = wgtselect],
by(race) notab;
(analytic weights assumed)
```

2-group Hotelling's T-squared = 644.49871  
 F test statistic: ((4126-10-1)/(4126-2)(10)) x 644.49871 = 64.309219

H0: Vectors of means are equal for the two groups

F(10,4115) = 64.3092  
 Prob > F(10,4115) = 0.0000

```
. hotel g* if prov == 5 & (race == 2 | race == 3) [weight = wgtselect],
by(race) notab;
(analytic weights assumed)
```

2-group Hotelling's T-squared = 36.794061  
 F test statistic: ((421-10-1)/(421-2)(10)) x 36.794061 = 3.6003736

H0: Vectors of means are equal for the two groups

F(10,410) = 3.6004  
 Prob > F(10,410) = 0.0001

```
. hotel g* if prov == 5 & (race == 4 | race == 3) [weight = wgtselect],
by(race) notab;
(analytic weights assumed)
```

2-group Hotelling's T-squared = 308.26486  
 F test statistic: ((639-10-1)/(639-2)(10)) x 308.26486 = 30.390947

H0: Vectors of means are equal for the two groups

F(10,628) = 30.3909  
 Prob > F(10,628) = 0.0000

```
. hotel g* if prov == 6 & (race == 1 | race == 3) [weight = wgtselect],
by(race) notab;
(analytic weights assumed)
```

2-group Hotelling's T-squared = 13.359971  
 F test statistic: ((2592-10-1)/(2592-2)(10)) x 13.359971 = 1.3313546

H0: Vectors of means are equal for the two groups

F(10,2581) = 1.3314  
 Prob > F(10,2581) = 0.2074

```
. hotel g* if prov == 6 & (race == 2 | race == 3) [weight = wgtselect],
by(race) notab;
(analytic weights assumed)
```

2-group Hotelling's T-squared = 6.7637532  
 F test statistic: ((46-10-1)/(46-2)(10)) x 6.7637532 = .53802582

H0: Vectors of means are equal for the two groups

```

          F(10,35) =    0.5380
    Prob > F(10,35) =    0.8511

```

```

. hotel g* if prov == 6 & (race == 4 | race == 3) [weight = wgtselect],
by(race) notab;
(analytic weights assumed)

```

```

2-group Hotelling's T-squared = 3.9609319
F test statistic: ((179-10-1)/(179-2) (10)) x 3.9609319 = .37595286

```

```

H0: Vectors of means are equal for the two groups

```

```

          F(10,168) =    0.3760
    Prob > F(10,168) =    0.9557

```

```

. hotel g* if prov == 7 & (race == 1 | race == 3) [weight = wgtselect],
by(race) notab;
(analytic weights assumed)

```

```

2-group Hotelling's T-squared = 75.064467
F test statistic: ((3303-10-1)/(3303-2) (10)) x 75.064467 = 7.4859808

```

```

H0: Vectors of means are equal for the two groups

```

```

          F(10,3292) =    7.4860
    Prob > F(10,3292) =    0.0000

```

```

. hotel g* if prov == 7 & (race == 2 | race == 3) [weight = wgtselect],
by(race) notab;
(analytic weights assumed)

```

```

2-group Hotelling's T-squared = 33.924649
F test statistic: ((214-10-1)/(214-2) (10)) x 33.924649 = 3.2484452

```

```

H0: Vectors of means are equal for the two groups

```

```

          F(10,203) =    3.2484
    Prob > F(10,203) =    0.0007

```

```

. hotel g* if prov == 7 & (race == 4 | race == 3) [weight = wgtselect],
by(race) notab;
(analytic weights assumed)

```

```

2-group Hotelling's T-squared = 35.857694
F test statistic: ((556-10-1)/(556-2) (10)) x 35.857694 = 3.5275168

```

```

H0: Vectors of means are equal for the two groups

```

```

          F(10,545) =    3.5275
    Prob > F(10,545) =    0.0002

```

```

. hotel g* if prov == 8 & (race == 1 | race == 3) [weight = wgtselect],
by(race) notab;
(analytic weights assumed)

```

```

2-group Hotelling's T-squared = 71.454693
F test statistic: ((2159-10-1)/(2159-2) (10)) x 71.454693 = 7.1156551

```

```

H0: Vectors of means are equal for the two groups

```

```

          F(10,2148) =    7.1157
    Prob > F(10,2148) =    0.0000

```

```

. hotel g* if prov == 8 & (race == 2 | race == 3) [weight = wgtselect],
by(race) notab;
(analytic weights assumed)

```

```

2-group Hotelling's T-squared = 23.135043
F test statistic: ((39-10-1)/(39-2) (10)) x 23.135043 = 1.75076

```

```

H0: Vectors of means are equal for the two groups
      F(10,28) = 1.7508
      Prob > F(10,28) = 0.1181

. hotel g* if prov == 8 & (race == 4 | race == 3) [weight = wgtselect],
by(race) notab;
(analytic weights assumed)

2-group Hotelling's T-squared = 15.399544
F test statistic: ((109-10-1)/(109-2) (10)) x 15.399544 = 1.4104256

H0: Vectors of means are equal for the two groups
      F(10,98) = 1.4104
      Prob > F(10,98) = 0.1869

. hotel g* if prov == 9 & (race == 1 | race == 3) [weight = wgtselect],
by(race) notab;
(analytic weights assumed)

2-group Hotelling's T-squared = 142.84951
F test statistic: ((3011-10-1)/(3011-2) (10)) x 142.84951 = 14.242225

H0: Vectors of means are equal for the two groups
      F(10,3000) = 14.2422
      Prob > F(10,3000) = 0.0000

. hotel g* if prov == 9 & (race == 2 | race == 3) [weight = wgtselect],
by(race) notab;
(analytic weights assumed)

2-group Hotelling's T-squared = 43.694187
F test statistic: ((14-10-1)/(14-2) (10)) x 43.694187 = 1.0923547

H0: Vectors of means are equal for the two groups
      F(10,3) = 1.0924
      Prob > F(10,3) = 0.5320

. hotel g* if prov == 9 & (race == 4 | race == 3) [weight = wgtselect],
by(race) notab;
(analytic weights assumed)

2-group Hotelling's T-squared = 58.711597
F test statistic: ((96-10-1)/(96-2) (10)) x 58.711597 = 5.3090274

H0: Vectors of means are equal for the two groups
      F(10,85) = 5.3090
      Prob > F(10,85) = 0.0000

```

### **Technical Papers in this Series**

<b>Number</b>	<b>Title</b>	<b>Date</b>
TP2003:1	Creating a 1995 IES Database in Stata	September 2003
TP2003:2	Creating a 1995 OHS and a combined OHS-IES Database in Stata	September 2003
TP2003:3	A Standard Computable General Equilibrium Model Version 3: Technical Documentation	September 2003
TP2004:1	SeeResults: A spreadsheet Application for the Analysis of CGE Model Results	November 2004
TP2004:2	The Organisation of Trade Data for inclusion in Social Accounting Matrix	December 2004
TP2005:1	Creating a 2000 IES-LFS Database in Stata	February 2005
TP2005:2	Forming Representative Household and Factor Groups in a SAM	March 2005

### **Other PROVIDE Publications**

Background Paper Series

Working Papers

Research Reports