Top incomes and racial inequality in South Africa
Evidence from tax statistics and household surveys
1993 – 2008

Elodie MORIVAL
Paris School of Economics - Master Analyse et politique économiques

Master's Thesis
August 2011

Under the supervision of Thomas PIKETTY - Examiner: Denis COGNEAU

Keywords: South Africa, income inequality, top income shares, income tax data, National accounts, household surveys

JEL Classifications: D31, C81

This master's thesis was done using STATA software

ACKNOWLEDGEMENTS

I would like to thank my supervisor Thomas Piketty for his insightful comments and suggestions and for his encouragement throughout the pursuit of this work. I am also thankful to Denis Cogneau who accepted to be the referee of my dissertation. I am very grateful to Facundo Alvaredo who agreed to discuss the subject of top incomes in South Africa with me, gave me access to tax tabulations and answered numerous questions about the functioning of tax data. Special thanks to Romain and my parents for their supports, to Jacques, Abel and Sarah for their help with STATA and to all the people I met in South Africa for their useful suggestions.

ABSTRACT

This paper shows that incomes in South Africa are highly unequally distributed. The share of the top 10 percent in gross income is over a half, that of the top 5 percent is over a third and that of the top 1 percent is above fifteen percent. Focusing on top incomes, we highlighted the fact that racial inequalities are still very important in post-apartheid South Africa. Despite an increase in the share of Blacks in the population - from 76,7% in 1995 to 79,2% in 2008 – top incomes are still mainly composed of Whites. Between 1995 and 2008 the share of Blacks in the top 10 percent rose from 25% to 36,5% and the share in the top 5 percent increased from 21% to 28%. On the contrary, the share of Blacks in the top 1 percent decreased from 19,5% to 16%. The increase in the top 10 percent and in the top 5 percent has been favoured by the positive discrimination implemented in the public sector. At the same time, the developments in financial sector, where most individuals belonging to the top 1 percent work, has complicated the access to high paying jobs for Blacks. These percentages are still very low compared to the racial composition of the population and the trend is still hesitant. With these results, we can conclude that the apartheid regime has left a very long-run footprint in South Africa and that the redistribution between interracial population groups is still an issue. To obtain these figures we used an unusual source of data. Instead of using tax statistics, we used five household surveys to cover the entire post-apartheid era. Using this type of data allows for a broader range of research than is possible with tax data. We can provide a full description of top income earners over the post-apartheid period: age, ethnicity, gender, level of education, business sectors and job occupations. Yet, household surveys are an imperfect source for the study of top incomes and two methodological sections assessing the level of quality of each survey precede our analysis of top incomes. In order to determine if our surveys can be used to examine top incomes, we successively compared them with National accounts and with tax statistics. Except for two years, 1993 and 2000, the surveys provide estimates of top income shares that are very close to the ones computed with tax data. The 1993 survey cannot be used to describe top incomes: measurements errors are too significant because the survey was undertaken during the democratic transition, a specific context of violence and political instability. The 2000 survey must be used with caution but true improvements have been done to enhance the quality of the data collected between the 2000 and the 2005 survey. The others surveys are used to explain why there are so few black people in the top income groups.

Contents

1. Introduction

In 1994 after more than two centuries of a racial regime, South Africa captured the attention of the world when the political power was transferred from the white minority to the black majority without massive bloodshed or civil war[1]. Afterward, South Africa became a global symbol of the struggle for human rights and racial equity and an economic model for many other African countries. Indeed, South Africa has made strong economic progress with a growth up to 5,5% in 2007 due to an increase in domestic consumption and a growing disposable income[2]. Nevertheless, South Africa has continued to face difficult challenges in overcoming its apartheid legacy and addressing high levels of unemployment and inequality. South Africa's inequality levels are still among the highest in the world and the issue of racial inequality continues to dominate the post-Apartheid era.

There are many studies on the evolution of income inequalities in South Africa such as Terreblanche (2004), Seeking and Natrass (2005), or Van der Berg and Louw (2003). Most of these studies establish a comparison between the evolution of inequalities and the reduction or the increase in poverty. They paint a highly unequal South Africa. Evidence from household surveys indicates that inequality increased both within the whole population and within each racial group[3]. At the same time, according to Leibbrandt et al. (2010) poverty has remained virtually constant – or fallen slightly – over the same period. If there is a consensus around the direction of post-Apartheid inequality and poverty trends, there is no agreement about the precise levels. The increase in inequality, after the end of the Apartheid era, may surprise some. Seventeen years after the advent of democracy, one should expect that the end of discrimination and segregation policies has brought a sharp decrease in inequality. The explanation given by South African literature is the following: the increasing inequality within the black population has prevented a significant decline in aggregate inequality. The declining interracial inequality was accompanied by a rise in intraracial inequality and the previous "racial society" seems to have been replaced by a "class society" (Terreblanche,

[1]The use of the words Black, Coloured, Indian and White refer to the apartheid-based racial classification of South Africans. The word "African" is used to regroup all historically disadvantaged racial groups: Blacks, Coloureds, Indians and Asians. This classification variable is also kept by household surveys in order to monitor changes in the life circumstances of those who were disadvantaged in the apartheid era.

[2] October 2008, IMF Country Report No. 08/348.

[3] The Gini coefficients are provided in Table A in the appendix.

2004) or at least by a new society where the racial footprint starts to grey to be replaced by a new social strata and more subtle socio-economic dynamics. "Race and class are no longer coterminous" as Seekings and Nattrass (2005) claimed in their book[4]. This conclusion is strengthened by the use of a decomposition of the Theil coefficient into "within-group" component i.e. the intraracial contribution to overall inequality and "between-group" component i.e. the interracial contribution to overall inequality. Analyzing these Theil coefficients – provided in the appendix – we observe a decline in the importance of between-race inequality and an increase in within-race inequality. In our opinion, this theory has to be analyzed with caution. In 2008, the Theil coefficient shows that 43% of total inequality is still a consequence of between-race inequality across the White group and the African group. Moreover, Leibbrandt et al. (2010) have shown that the bottom deciles of the income distribution and the poverty profile are still composed mainly by Africans. Despite the end of the apartheid regime, some inertia can remain so that inequality may still be particularly important among the richest South Africans. As a consequence, this question of intraracial inequality led us to the study of top incomes and their racial composition. Two questions must be tackled. Is the intraracial inequality driven by an increase in the number of the very rich Blacks? Did South Africa become a multiracial society even among the top income earners? The answer to this question is crucial and can influence policy initiatives: should economic and social policies address the increase in intraracial inequality and stop focusing solely their attentions on redistribution between racial groups?

There are 65,000 United States Dollar millionaires in South Africa and 120,000 South Africans are members of the global top 1 percent of wealth holders[5]. Yet, top incomes have never been studied in South African literature except through the consideration of the last decile. Alvaredo and Atkinson (2010) wrote a paper on top incomes in South Africa entitled "Colonial Rule, Apartheid and Natural resources: top incomes in South Africa, 1903 – 2007". This paper studies the evolution and the concentration of the top 5 percent, the top 1 percent, the top 0,5 percent and the top 0,1 percent from 1903 to 2007. They use income tax data, a method that allows them to cover a very long period. The paper shows that despite short-run movements, the share of the top 1 percent was halved between 1914 and 1993. On the contrary, since 2002, the study of top income shares shows no downward trend. The use of tax statistics presents two significant weaknesses that we would like to overcome. Income tax data are not available between 1994 and 2002, which prevents the authors from analyzing the

[4] Seekings and Nattrass, 2005, p343.
[5] James Davies and Anthony Shorrocks, 2010, Credit Suisse Global Wealth Databook.

dynamics of top incomes after the Apartheid regime. Moreover, tax tabulations are published with a classification by race only between 1956 and 1987.

Our idea is to use another type of data to study the trend of top income shares and its racial composition since the end of the Apartheid era. There are several statistical data on household incomes since 1994: the National accounts, the Income and Expenditure survey, the population Censuses, the new National and Income Dynamics Surveys and the tax data from 2002 to nowadays. All these sources have limitations. The National accounts provide means or totals and cannot be used to study the complete distribution of households' incomes. The household surveys are imperfect and suffer from the classical biases of surveys. The population censuses cannot be used to study top incomes since they indicate personal incomes in income classes. Moreover, almost two millions households reported no income at all in the 2000 population census. Tax data are a great source of information to study top incomes but tabulations are only available from 2002 and tell us nothing about the ethnicity, the profession or the level of education of taxpayers. Yet, knowing that tax data are probably the best source of data to accurately estimate top income shares we decided to precisely compare the shares found with household surveys and with tax data.

The first purpose of this study is to bring the available information within a common framework, comparing household surveys, national accounts and tax data, proposing corrections to incomplete or erroneous surveys, in order to select the household surveys which can give reliable estimates for the top income shares. The top percentile shares measure the share of total income accruing to the top percentile of adult individuals. Therefore, we first compared the denominator - the total income of the entire population - found in National accounts and in household surveys. Afterwards, we compared the entire distribution of income provided in household surveys and in tax tabulations. The second purpose of this master's thesis is to study the characteristics of top income earners: age, gender, ethnicity, business sector and level of education in order to better understand the trend of top income shares since the advent of democracy in South Africa. At the same time, we would like to examine the assumption according to which South Africa has become a "class society". This second part is only done if the distribution of income in household surveys is comparable to the one found in income tax data.

The rest of the paper is organized as follows. Section 2 describes the data used for the research reported in this paper. Section 3 establishes a comparison between households' income according to the surveys and according to National accounts. Section 4 compares the

distribution of income in tax statistics and surveys and suggests an explanation for the evolution of top income shares over the period. Section 5 is an empirical description of top income earners and is used as an attempt to explain the low percentage of Blacks among the top 10 percent, the top 5 percent and the top 1 percent through the analysis of their age, gender, business sector and level of education. Section 6 concludes.

2. Data

2.1 Income tax data

The data used by Alvaredo and Atkinson (2010) in their paper are income tax tabulations. Data employed are not in the form of individual tax records but in the form of published tabulations. These tabulations report for several income brackets the corresponding number of taxpayers and their total income. For 1993 - and for many years before 1993 - official statistics only give the number of taxpayers by income bracket, but not the exact amount of income assessed in that bracket. To overcome this issue, one can assume either than the income of every person in each bracket is equal to the lower bracket limit or to the upper bracket limit. This leads to less precise top income shares estimates. Income brackets go up to R 150,001 in 1993 and to R 5,000,001 after 2002[6]. No income tax data is available during the transition between 1993 and 2002. Income shares estimates for 1993 and from 2002 onwards are not directly comparable since the tax administration has improved and tax code has changed.

Income tax is levied on residents' worldwide income. Non-residents are taxed on their income from a South African source. Since 1990, the income tax is based on the individual. The total income declared is composed of: salary and wages; commission; pension; member's fees (for member of close corporation); director's fees / remuneration; business income; farming income; service benefits; fringe benefits; allowances; income from investments; rental income; annuity income from a retirement fund; lump sum benefits; gratuities; special remuneration: proto teams; income from sources outside South Africa; other. Taxable capital gains also form part of taxable income. The main part of individual income is composed of salary and wages, pension and retirement payments and investment income (interest and dividends)[7]. Some individuals may also have business income that is taxable as personal income.

Two issues have to be tackled to estimate top income shares with income tax data. The first is the need to relate the persons to a control total to define how many tax filers represent a given percentile. In South Africa, less than 10 percent of the population is registered in income tax data. Therefore, the computation of top income shares cannot be done below the

[6] 1 rand ≈ 0,1euros
[7] 2008 Tax Statistics, National Treasury and the South African Revenue Service

top 5 percent[8]. A large number of formal sector workers, earning taxable income below the tax threshold (the level of annual income below which no income tax is payable), are not required to register for tax data. Workers who earned an amount below the tax threshold only pay the "Standard income tax on employees" (SITE), a tax deducted directly from their remuneration by their employers. Consequently, an individual whose annual net remuneration does not exceed the tax threshold, who does not receive a travel allowance or any other income is not registered in the tax returns. Two types of thresholds exist: one for individuals aged younger than 65 years and one for individuals aged 65 years old and above. Between 1993 and 2007, the income tax threshold has increased significantly. Table D in the appendix reports the tax thresholds between 1993 and 2007. Even if employees who pay SITE-only do not need to register for income tax or submit tax, the tax tabulations contain some registered individual taxpayers with taxable income below the tax threshold. It is mainly individuals who need to declare other non-SITE income. The second issue is the need for a control total income used as the denominator in the top income shares estimation. This question is examined in section 3.

2.2 Household surveys

In order to provide more information on top income earners than those available in income tax data we decided to use household surveys to compute top income shares. In the following sections we used three kinds of surveys: the Project for Statistics on Living Standards and Development survey (PSLSD), the Income and Expenditure surveys (IES) and the National Income Dynamics survey (NIDS).

The 1993 PSLSD survey was undertaken during the democratic transition of South Africa. The survey took place in a specific context of South Africa's history. The main idea was to collect statistical information on living standards in order to help policy makers with the data required for planning strategies. The topics covered by the detailed questionnaire included demography, household services, household expenditures, educational and health status, remittances, employment and income. The survey contains about 9,000 households. The main instrument used in the survey was a comprehensive household questionnaire. Training was provided to fieldworkers in order to ensure consistency. All the questionnaires were checked when received. Where information was incomplete or appeared contradictory,

[8] In the first version of their paper, Alvaredo and Atkinson (2010) computed also the income share of the top 10 percent. In the forthcoming new version, they decided to start with the top 5 percent since the income share of the top 10 percent is affected by missing people.

the questionnaire was sent back to the relevant survey organization. The sample design adopted for the study was a two-stage self-weighting design in which the first stage units were census enumerator sub-districts and the second stage were households. Such a design automatically provides a representative sample and weights do not have to be added. Nevertheless, weights were provided to undertake issues such as violence, noncompliance and under-representation of Whites in the sample. To compute the top income shares we would have liked to use a definition of income as close as possible to the one used in income tax data. Unfortunately for the year 1993, we only had the monthly gross pay of people having a regular employment. This includes wages, self-employment income and eventually profit shares, bonuses and allowances.

The Income and Expenditure surveys were administered to a representative sample of households with national coverage. It is originally designed for the compilation of the Consumer Price Index (CPI). The surveys collect information on items and services acquired by South African households, various sources of income acquired by participating households (monetary or in-kind) and details on how they spent this income. The 2005 IES is the third of its kind, and follows similar surveys undertaken in 1995 and 2000. The Income and Expenditure survey for 2010 is not available yet but could be used to improve this work in the future. The 2005 IES has been improved significantly compared to the 1995 and 2000 IES. The survey was conducted over a period of one year, from September 2005 to August 2006, with sampled households participating for one month and new subsamples of households starting every month. On the contrary, for the 1995 and 2000 IES the survey was conducted during a period of four weeks. Information was also collected regarding income acquired by different members of the household during the survey month and during the 11 months prior to the survey. The data collection was done through an extensive interview with trained fieldworkers. In 2005, this questionnaire was split and conducted on five separate visits during the time of the survey. Only one interview was done for the previous IES. In 1995 and 2000, the IES contains about 30,000 households whereas the 2005 IES contains 25,000 households. A two-stage weighting procedure was applied for each sample of the IES. The definition we use for regular income is composed by the following incomes: salaries and wages; commission and director's fees; receipt from pension, social welfare grants and other annuity funds; net profit from business or professional practice/activities or farming conducted on a full-time basis; part-time work and cash allowances in respect of transport, housing and clothing; regular allowances; interest received and/or accrued on deposits, loans,

savings certificates, and dividends on building society shares; income from letting of fixed property; royalties; dividend received on shares; bonuses and income from overtime

The 2008 NIDS is the first national panel study used to measure changes in the well-being of South Africans. The survey follows about 7,305 households. The NIDS should document the dynamic of incomes, expenditures, assets, access to services, education, health and other dimensions of well-being. The design of NIDS envisaged data collection every two years. The National Income Dynamics Study used a combination of household and individual level questionnaires. The computation of the weights was the same in NIDS and in IES. This is essentially a two-stage procedure. In the first stage, the design weights were calculated as the inverse of the probability of inclusion. In the second, the weights were calibrated to the 2008 midyear estimates. Two sets of weights are thus provided, the design weights and the post-stratification weights. In 2008, the definition we used for "regular income" is composed by the following incomes: main and secondary jobs; casual wages; self employment income; 13th cheque; other bonus; profit share; extra piece-rate income; old age pension; grants; interest/dividend income; rental income; private pensions and annuities; inter-household remittance.

In the table below, we sum up the composition of the regular income we use for each type of surveys and for tax statistics.

Table 1: Categories composing total income in household surveys and in tax data

NIDS	IES	Income Tax data
Income from work		
Main and secondary job Casual wages Self employment income 13th cheque Other bonus and xtra piece-rate income	Salaries and wages Commission and director's fees Net profit from business or professional practice/activities or farming conducted on a full-time basis Bonuses and income from overtime	Salary Wages Commission Director's fees Business income Farming income Gratuities Special remuneration: Proto teams Member's fees (for Member of Close Corporation)

Income from capital and property		
Interest and dividend Profit share Rental income	Interest received and/or accrued on deposits, loans, savings certificates, and dividends on building society shares Net Income from letting of fixed property Royalties Dividend received on shares	Income from investments Lump sum benefits Rental income
All other income		
Old age pension and Grants Private pensions and annuities Inter-household remittance	Receipt from pension, social welfare grants and other annuity funds Part-time work and cash allowances in respect of transport, housing and clothing Regular allowances	Pension Annuity income from a retirement fund Service benefits Fringe benefits Allowances Income from sources outside south Africa

Most surveys impose top coding to limit the effect of measurement errors on aggregates, which limits the analysis of top incomes. The household surveys we used here are not top-coded. In each survey, sample weights are constructed to allow aggregation of estimates to the South African household population level. All the results presented in the following sections are weighted up to population totals to be representative of the all population and not only of the sample.

3. Household surveys and National accounts: comparison of total households' income reported in each data

Computing top income shares estimates with income tax data and with household surveys, we have noticed that the shares are higher with surveys than with tax statistics. This is not surprising and similar to the results provided by Burkhauser et al. (2009) for the United States comparing top incomes estimates between the internal CPS data and the IRS tax return data-based reported by Piketty and Saez (2003). Burkhauser et al. (2009) find close estimates with the two data except for the share of the top 1 percent of the distribution during 1993 – 2000. They conclude that the differences in inequality trends observed by other researchers using the two data sources are not primarily due to deficiencies in either data source but rather to the traditions of income inequality measurement used in the two literatures (definition of income, choice of income inequality index…)[9]. We would like to explain more precisely the differences between the estimates and particularly if these differences come from a divergence in the definition of income used for the numerator or the denominator of the top shares or if they come from deficiencies in either data. This section establishes a comparison between the two denominators used by Alvaredo and Atkinson (2010) and by us: households' total income declared in surveys and in National accounts. Indeed, a first assumption to explain why we find higher estimates can be that the denominator used by Alvaredo and Atkinson (2010) to compute their income shares is bigger than the one we use. Nevertheless, if the total income reported in surveys is underestimated compared to the one found in National accounts, one can also think that income will be under-declared in surveys compared to income tax data. In that case the underestimation of the denominator would be more or less compensated by the lower numerator and top income shares estimates would be close. As a consequence, section 4 establishes a comparison between household survey and tax statistics and is therefore complementary to the section 3 to explain these differences.

[9] Burkhauser and al (2009), p 22.

3.1 Presentation and ratio of coverage between household surveys and National accounts estimates

As we saw in the presentation of income tax data, one issue with this type of data is to find a control for total income. Indeed, the authors need to relate the amounts recorded in the tax data – the numerator of the top share – to a comparable control total for the entire population – the denominator of the top share. As explained by Atkinson, Piketty and Saez (2011), this is a matter of attention since different methods can be employed and may affect comparability between our income shares. As the tax records in South Africa only cover a small part of total households' income, Alvaredo and Atkinson (2010) – inspired by Kuznets (1953) – combined income tax data with National accounts estimates to obtain total income. Their denominator is the households' gross income (households' disposable income plus taxes on income and wealth paid by households) found in National accounts and adjusted to reflect the year of income tax assessment. The households' gross income after adjustment is close to the one reported in National accounts. The total income found in National accounts is higher than the total income found in tax data for two reasons: National accounts include all households and not only those with an income above the tax threshold and National accounts also include income that is not taxable (and not reported) such as transfers or post office savings bank interests (below a specified amount). As the authors acknowledge they "are understating the top income shares since this non-taxable income is omitted from the numerator"[10].

The denominator we used is the total households' regular income defined in Section 2 for each survey. We do not have the possibility to compare exactly the top income shares for each year, only 2 years are covered by the two data, 1993 and 2005. To overcome this issue the 2000 Income and expenditure survey will be compared with the fiscal year starting in March 2002 and the 2008 National Income Dynamics Survey will be compared with the fiscal year starting in March 2007. Tax data were not available between 1994 and 2002, but National accounts were still published annually, which allows us to know the denominator for these missing years.

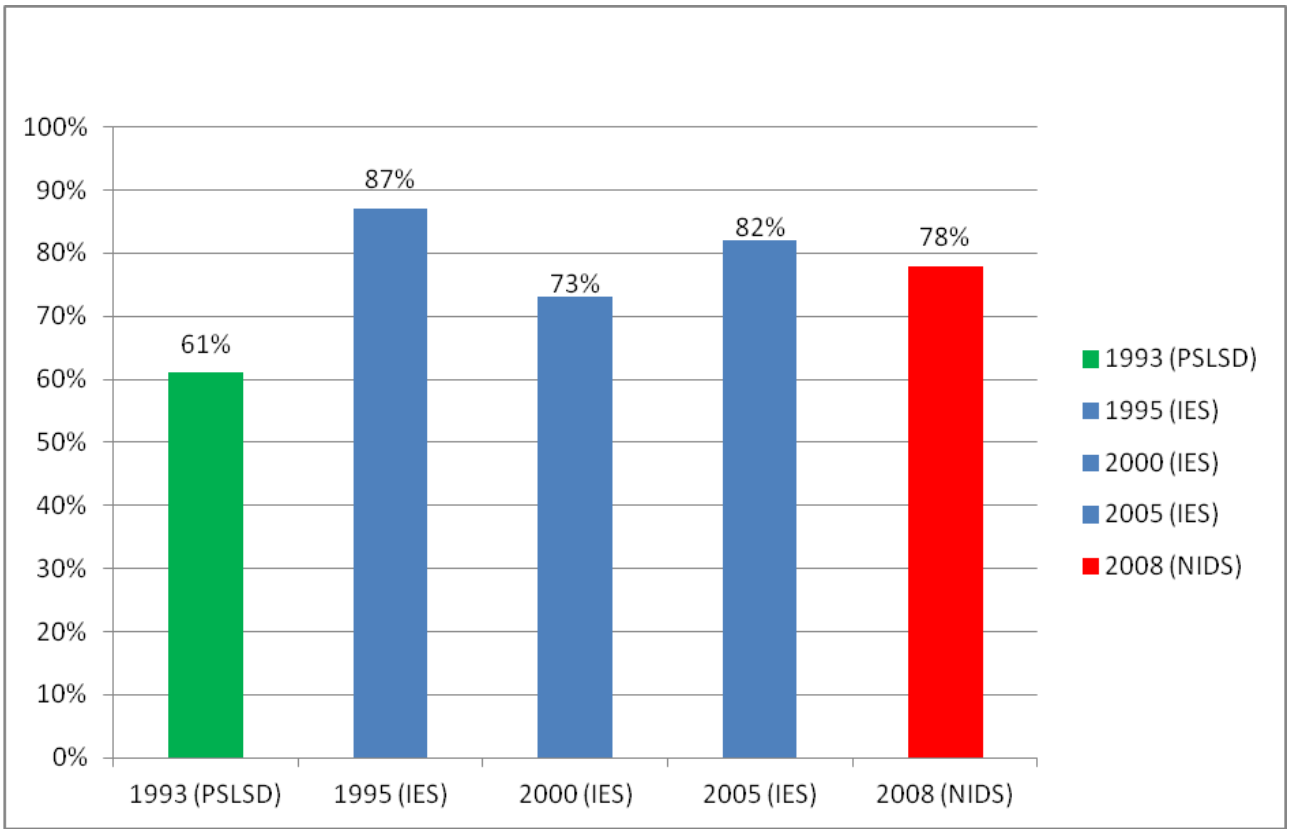
In the Figure 1 below we compare:

- The denominator used by Alvaredo and Atkinson: total households' gross income found in National accounts adjusted to reflect the year of income tax assessment.

[10] Alvaredo and Atkinson (2010), p9.

- The total households' income reported in household surveys.

Figure 1 – Gross households' income: household surveys estimates as a percentage of adjusted National accounts estimates



Source: own calculations using data from 1993 PSLSD, IES 1995, 2000, 2005 and NIDS 2008 data sets and National accounts adjusted by Alvaredo and Atkinson (2010)

At the aggregate level, household surveys usually report lower income than the National accounts estimates. Indeed, the aggregates resulted from the surveys and their counterparts in National accounts differ in definition and measurement methodology. It is not surprising to find a total households' income smaller in surveys than in National accounts. As Alvaredo (2007) noticed in his paper on the rich in Argentina, means of consumption and income from household surveys and National accounts differ not only because the rich might not be present in the surveys, but also because the two sources of information are different and measure different concepts. According to Alvaredo (2007), National accounts track money and are more likely to capture large transactions, while surveys follow people and are less likely to include large transactors.

The coverage between the denominator found with household surveys and the denominator used by Alvaredo and Atkinson (2010) vary significantly across the years. For the IES the ratio of coverage between the adjusted national accounts and the household surveys declines from 87% in 1995 to 73% in 2000.

Reconciling household survey data and National accounts data is a current problem. For instance, economists such as Ravaillon (2001) and Deaton (2005) tried to analyze and explain the difference in consumption growth found with household surveys and with National accounts. We did the same thing with income and tried to explain the difference between the two estimates of households' total income. Both surveys and National accounts can be at the root of the discrepancy. Noncompliance and income underestimation in household surveys can lead to measurement errors. The inclusion of specific type of incomes in National accounts which are absents from surveys can also overestimate the total households' income.

3.2 Analyzing the difference between household surveys and National accounts estimates: the issue of noncompliance and income underestimation in surveys

The issue of noncompliance

Not everyone who is asked to participate in a survey agrees to do so. If compliance is random, there is no bias in the survey's estimates. Indeed, non-response is only a problem if the non-respondents are a non-random sample of the total sample. According to Groves and Cooper (1998) this is often the case. They have shown in their paper that the probability of response is negatively related to almost all measures of socioeconomic status. Mathematically this will means that the probability to answer to an income survey is monotonically declining with the income. Failure to response is assumed to be higher with better-off households. In our opinion, this conclusion is not exactly true: if high-income households might be less likely to participate because of a high opportunity cost of their time, the poor face also the same issue. Townships can be dangerous for fieldworkers and inhabitants are hard to contact. Noncompliance explains probably some part of the shortfall between the surveys and the National accounts, but how much cannot be measured exactly. Noncompliance can be minimized through two kinds of procedures: imputation and weighting. Imputation can be used for item non-response i.e. blanks within the questionnaire. Imputation can be used to

impute values for the missing items. In the 2000 IES a category "undeclared income" was created. Where the total income was not reported and expenditure was shown, undeclared income was estimated equal to the value of the reported expenditure. Where the total income was significantly less than the total expenditure, undeclared income was estimated equal to the total expenditure minus the total income. Adding this "undeclared income" to the total income in the 2000 IES, we find a coverage of 73% with the National accounts instead of a coverage of 65% if we do not take into account this "undeclared income". Imputation can thus be used to improve the quality of the data and provide a better match with National accounts. A similar category is not provided in the others surveys. In addition, weights can be computed so that the achieved samples for surveys reflect truly the entire population. Rather than accept a poor match between the sample and the population (more women or Blacks in the sample for instance), we can use weights to bring the two more closely into line. This is known as "non-response weighting". As explained in section 2, a two-stage weighting procedure was applied for the IES and the NIDS. In the first stage, the design weights were calculated as the inverse of the probability of inclusion. In the second, the weights were calibrated to the previous census available for the IES or to the 2008 midyear estimates for the 2008 NIDS. Two sets of weights are provided, the design weights and the post-stratification weights. The post-stratification weights adjust the design weights such that the age-sex-race marginal totals in the data match the all population estimates. We use these weights to compute all the results presented in this paper. The detailed procedure to compute the weights is provided in the appendix. It should be noted that the non-response weighting procedure can reduce biases but do not eliminate them. Indeed, the reasons according to which an individual decides to take part in a survey are complex, and depend upon a lot of factors specific to each individual. Post-stratification only aligns the survey to the population along a small number of dimensions (such as age and sex).

Income underestimation

Even for respondents who agree to participate and answer all the survey questions, measurement error is still a concern in survey data. The reliability of income reported in the household surveys is related to the accuracy and completeness with which respondents consent to share information with the survey takers. Respondents might under-report their earnings either through forgetfulness or through fear of the taxation authority. As Johnson and

Moore (2008) explain, respondents may "guestimate" their answers to questions, "even if respondents' guesses overall are not biased, such approximation reduces the estimation efficiency of the data"[11]. In the developing world, surveys are known to detect almost exclusively wages and pensions, self-employment income and public transfers, while capital income is often largely neglected. Misreporting and underreporting of some income sources in the household surveys can explain some part of the discrepancy between the total income found in surveys and in National accounts. Statistics South Africa – the national statistical service of South Africa – proposes in its *Analysis of results* (2008) of the 2005 IES a comparison between the household surveys and the National accounts by source of income. In the household survey, almost 75% of the annual gross income is derived from work activities: salaries and wages, self-employment and other business income. This corresponds approximately to 96% of the "compensation of employees" find in National accounts. The total income from capital reported in the 2005 IES is only 4% of the total reported in National accounts. The definition used for income from capital is different in the two data. National accounts report "income from capital and property" which correspond to interest (income) plus dividends plus rent (income) plus property income attributed to insurance policy holders. In the household survey the equivalent is composed by interest, income from dividends and rent received. However, this difference in definition is not sufficient to explain the significant discrepancy between the total income from capital in surveys and in National accounts. Another possible explanation is that households who receive only small amounts of taxable interests or dividend income may forget to report these amounts in the IES questionnaire. As noticed by Johnson and Moore (2008) households may not think they have "received" this income, particularly in the case of interests earned on bank accounts and money market funds. Moreover, these types of earning have a great variability that can lead households to underestimate them, especially if these incomes are not an important source of disposable income. In any cases, this conclusion prevents us to decompose the total income of top income earners by source of income in the following sections since income from capital will probably be under-reported.

[11] Johnson and Moore (2008), p2.

3.3 Analyzing the difference between household surveys and National accounts estimates: do National accounts provide an overestimation of total households' income?

There are significant differences between income concepts used in National accounts and those used for tax purposes or household surveys. Discrepancies between National accounts and household surveys are expected and can be tolerated if they are not too large. Indeed, both series are not directly comparable and a number of factors account for the differences between the two, but in unknown proportions. First, numbers may differ because of definitional differences for the sample and the income. National accounts do not rely on a sample in the same way as household surveys which relate to a relatively small number of households. The household sector in National accounts is combined with non-profit institutions serving households (charities, universities, trade unions…). The definition of income is also different. The National accounts concept of personal income includes items that are not asked for in the surveys. Two items are added in National accounts: employer contributions to funds (notably pension and medical aid) and imputed rent. Imputed rent is recorded (but not included in our definition of regular income) in the 2005 IES and in the 2008 NIDS but not asked for in the other surveys. Employer contributions can explain a significant part of the difference. According to Simkins (2004) employer contributions to funds can run as high as 25% for salaried professionals and managers, but are, on average, lower for less skilled formal sector workers and are inexistent in the case of the self-employed or workers in the informal sector. On average, 10% of compensation of employees is assumed to consist of employer contributions to funds. The total in household surveys is not corrected for the non-household elements and for the difference between the income definitions.

The total gross households' income used as a denominator by Alvaredo and Atkinson (2010) is different from the total found in household surveys and in income tax data. Aaberge and Atkinson (2010) faced the same problem. In order to compute the top income shares in Norway they used - as Alvaredo and Atkinson (2010) - the total households' income series of the National accounts as a point of departure. For every single year, Aaberge and Atkinson (2010) found that the household income total exceeded the total reported in the income tax tabulations. In 1950, for example, the household income total was higher than the total recorded in the tax statistics by about 55%. According to Aaberge and Atkinson (2010) "in

part this difference reflects the incomes of those not covered by the tax statistics; in part the difference reflects differences in definition or in the valuation of income"[12]. As we said earlier, National accounts include incomes (such as transfers or tax-free income), which are missing from the income tax statistics. According to Aaberge and Atkinson (2010), the second of these differences is too high to use the National accounts household income totals ("These [differences] means that we cannot simply take the total household income series")[13]. An alternative approach to the National accounts is to adjust the series of tax data using other information about the income of those not covered. During the twentieth century, they use the estimates of total assessed income, including those not covered by the tax statistics. The highest ratio found between the estimates of total assessed income and National accounts is 72 percent. Aaberge and Atkinson (2010) decided to use as a control total for income a fixed percentage (72 per cent) of the household income total recorded in the National accounts. According to the authors "a reasonable first approximation to an income concept that allows for those not covered, but is otherwise defined in the same way is a fixed percentage of the household income total"[14]. Since 72 percent was the highest ratio found by the authors it is also an upper bound, which should lead, as the authors acknowledged, to an under-estimation of the top income shares. This approach is close to the one adopted for Sweden by Roine and Waldenström (2005), where they took a constant percentage of total personal income. If estimates of total assessed income, including those not covered by the tax statistics, are available in South Africa, a similar approach could be interesting in the paper of Alvaredo and Atkinson (2010) in order to see if National accounts overestimate households' income. In Section 4, we will see that the total income declared in tax data is very close to the one reported in the 2005 IES. Yet, in 2005, the total households' income in National accounts is higher than the total reported in survey by about 18%. Households' income in National accounts might be too high compared to the real value of households' income, which can lead to an underestimation of top income shares.

[12] Aaberge and Atkinson (2010), p5
[13] Aaberge and Atkinson (2010), p28
[14] Aaberge and Atkinson (2010), p5

3.4 Interpreting the evolution of the ratio of coverage between household surveys and National accounts estimates

We managed so far to broadly explain why surveys and National accounts yield different estimates. The deviation between the two estimates is crucial but the trend in the ratio over time is also important. The 1993 PSLSD survey and the 2008 NIDS are not directly comparable with the others surveys. The 1995, 2000 and 2005 Income and Expenditure surveys are easier to compare. Indeed these surveys try to ensure continuity in the items included and in the data definition so that all IES can be compared over time. In 1995, the coverage is 87%, in 2000 it decreases to 73% and in 2005 the coverage increases to 82%. Despite the willingness to keep all the Income and Expenditure surveys comparable, coverage and content in household surveys are subject to discontinuities resulting from changes in sample size and improvement in methods to collect information. The methodology to compute aggregates in National accounts did not change between 1995 and 2005. We can explain the rise in the discrepancy between 1995 and 2000 by an impoverishment in the quality of the data. We can think that biases presented above – noncompliance and measurement error – are more important in the 2000 IES. Households' income as reported by the IES increased by 20% from 1995 to 2000 and from 40% based on the denominators used by Alvaredo and Atkinson (2010). The increase of 20% was below the 38% increase in the Consumer Price Index (CPI) over the same period. This implies a decline in households' income in real terms between 1995 and 2000. This trend does not appear plausible. The changes measured from 2000 to 2005 and from 1995 to 2005 appear more plausible and are always well above the increase of the CPI which indicates that household income increased in real terms. Moreover, Seekings and Nattrass (2005) noticed some inconsistencies with the weights used in the 2000 IES. It seems that the weights estimated by Statistics South Africa for racial group in 2000 underestimate the size of the white population. This is potentially consequential for the total households' income since the white population is the wealthier population in South Africa. It also appears that the 2000 IES undersamples higher-income African households and that Statistics South Africa did not adjust the weights to take it into account. These criticisms highlight the fact that the quality of the survey was better in 1995 and in 2005 than in 2000.

The main issue when one wants to use household surveys to study the trend of top income shares is that there are not all comparable. Some surveys provide better quality data than others and this can affect our estimates. It is better to use them for static rather than for dynamic studies. Except for 1993, one can be satisfied by the discrepancy we found between the households' income recorded in surveys and in National accounts. The discrepancy is always below 30 percent. In Ravaillon's paper (2001), which studies how well National accounts and surveys agree, Ravaillon found that aggregate household expenditure from India's National Sample Survey accounts only for about 60% of private consumption from the National accounts. The discrepancies we found in the South African surveys are always lower compared to others developing countries. Taking into account the over-estimation of gross households' income in National accounts and the biases in household surveys, the differences can basically be explained. Nevertheless, the differences between the denominator used in the paper of Alvaredo and Atkinson (2010) and the one used in this paper explain only a part of the divergence between our top income shares. To complete this conclusion we need to provide a second analysis, comparing household surveys and income tax data.

4. Household surveys and income tax data: analysis of the income distribution in each data and computation of top income shares

There are a number of differences between income tax data and surveys such as there are differences between National accounts and household surveys. The population covered, the unit of observation, the sample size and the motivation people face in providing data are different. The 2005 IES cover approximately 85,000 individuals and the 2005 tax data file contains approximately 4 millions tax records out of a population of about 45 millions. As explained in the previous section, household surveys have some drawbacks but using this type of data allow for a broader range of research than is possible with tax data. Income tax data includes only South African residents whose gross income is above the "tax threshold" which changes every year and is legislatively prescribed. Income tax filers represent therefore less than 10% of the population. The main advantage of household surveys is to provide the entire distribution of income and to cover the entire South African population with the sample weights. Yet, they suffer from other issues such as unit and item non-response or measurement errors. For tax data this issue is less significant but taxpayers can also intentionally underreport certain types of income in order to reduce tax liabilities. Tax evasion and tax avoidance are the survey counterpart of nonresponse and incomplete response. Tax data and household surveys are not free of problems regarding the under-reporting of income. They must be read with caution but there are still informative to study the dynamics of income concentration.

The comparisons between income tax data and household surveys are more restricted than the comparisons between household surveys and National accounts. In 1993, we can compare the total income declared in the 1993 PSLSD survey and in the income tax data but we cannot compare the entire income distribution. For 1993, SARS only provides the number of taxpayers by income brackets and not the income declared in each bracket. The income assessed is computed using either the upper or the lower bound of the income bracket. Moreover, the last income bracket is only R150,001 which is too low to study the distribution of top incomes. Tax data are not available in 2000 and we will compare the 2000 Income and expenditure survey with the fiscal year starting in March 2002. For the 2008 National Income Dynamics Survey, we can use the 2007 data file but the results are based on incomplete assessments.

4.1 Total income reported in tax data and in household surveys

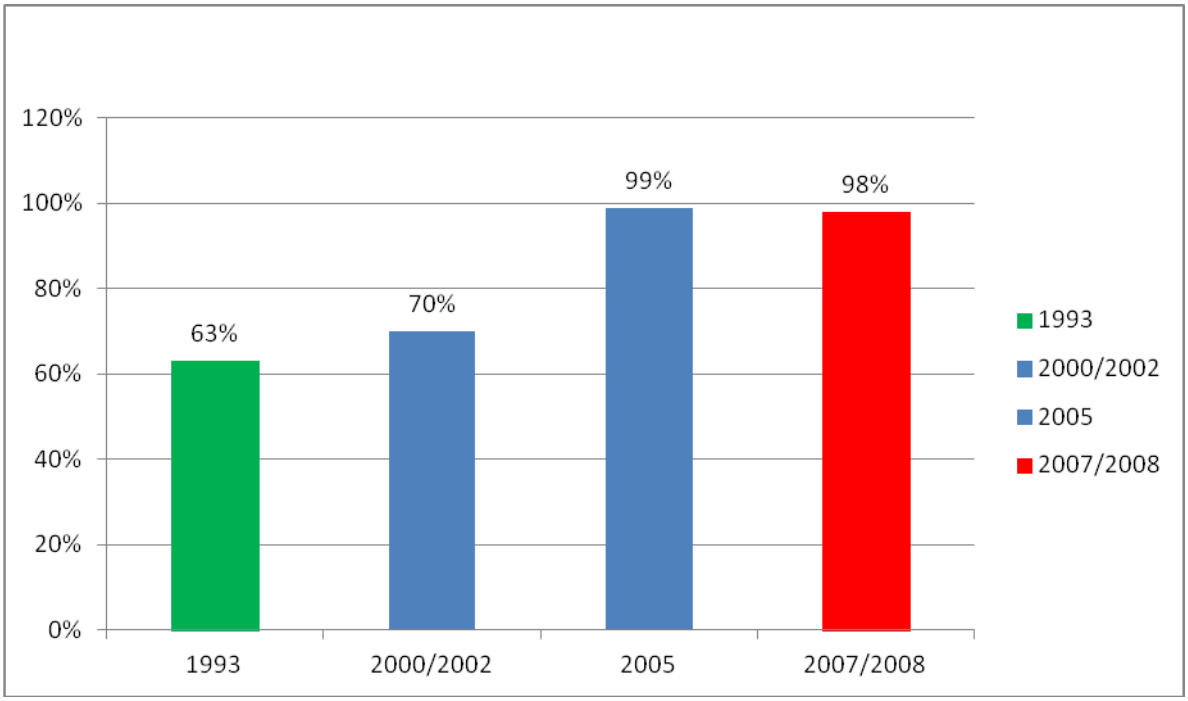
Household surveys are rarely used to study top incomes. There are often criticized and seen as not reliable. For instance, according to Alvaredo (2007) in his paper on Argentina: "microdata (…) do not offer valuable information when targeting the top, as the rich are missing either for sampling reasons, low response rates or ex-post elimination of 'extreme' values". Extreme values have not been eliminated from our surveys since the incomes are not top-coded. Yet, as shown in previous sections, surveys suffer from others biases. In order to measure how far household surveys are affected by these biases we compared them with income tax data. First, we did a comparison between the total income reported by taxpayers and by individuals in tax data and surveys. Since a similar total does not mean a comparable distribution we next studied the entire distribution of income in 2005 and 2007.

In the Figure 2 below we compare:

- The total income reported by taxpayers above the tax threshold.

- The total individual income reported above the same tax threshold in household surveys. The 2000 IES is used as a proxy for the 2002 tax file.

Figure 2 – Total individual income: surveys as a percentage of tax data estimates



Source: own calculations using data from 1993 PSLSD, IES 1995, 2000, 2005 and NIDS 2008 data sets and income tax tabulations provided by Alvaredo

In 1993, the total income declared above the tax threshold (R11,285) in tax data was less than R145,695 million (using the lower bound of the income brackets). In the 1993 PSLSD survey the total income declared above R11,285 was about R91,300 million. In the 1993 PSLSD surveys, we use an imprecise and low definition of income: the survey only reports the monthly gross pay for individual having a regular employment. As a consequence, the total income is underestimated by 37% in the survey. This discrepancy reflects to the one we found with National accounts. In 2002, the total income above R40,000 reported in tax data is about R331,851 million. Using the 2000 IES we found a total income above the tax threshold of R233,000 million. According to National accounts, gross households' income increased from R677,743 million in 2000 to R825,682 million in 2002. If this growth was equally shared among individuals we can partly explain the discrepancy between the two aggregates: 18 percent of the difference can be explained by the growth in income between 2000 and 2002. Yet, the underestimation in the 2000 IES compared to the 2002 tax file is about 30 percent. The last 12 percent results probably from the biases we exposed earlier. In 2005, income tax data reported a total income above the tax threshold of R483,008 million against R483,000 million in 2005 IES. In 2007, income tax data reported a total income above the tax threshold of R477,055 million against R470,000 million in 2008 NIDS. Despite this close figure, the 2008 NIDS might underestimate the total income since the 2007 data file have been computed from incomplete assessments. Since a similar total does not mean a comparable distribution, we would like to study the entire distribution of top incomes provided in income tax data and in surveys. We do this comparison only in 2005 and 2007 since this is not possible in 1993 and 2000.

4.2 Are top incomes underestimated in household surveys?

Household surveys are not without shortcomings. As shown by Michal Brewer et al. (2008) in the United Kingdom, nonresponse, incomplete response and measurement errors, particularly affect the top income ranges. In order to evaluate how far household surveys underestimate or not top incomes we need to combine household survey data with information on upper income ranges from tax sources. In the following table we did a comparison between the 2005 IES and the tax data for each income brackets. The tax threshold for 2005 is R60,000 for people aged above 65 years old. To avoid missing people in tax data, we only started our comparison from the income bracket R60 001 – 70 000.

Table 2: Income tax tabulations and household survey, 2005

	Tax statistics		Household survey	
Income brackets In 2005 rand	**Number of taxpayers**	**Total income, in million 2005 rand**	**Number of individuals**	**Total Income, in million 2005 rand**
60 001 – 70 000	207 869	13 545	348 678	22 700
70 001 – 80 000	249 257	18 737	390 218	29 300
80 001 – 90 000	237 387	20 140	286 495	24 300
90 001 – 100 000	242 071	23 055	260 903	24 600
100 001 – 110 000	216 990	22 726	230 310	24 300
110 001 – 120 000	188 764	21 721	233 843	27 200
120 001 – 130 000	180 877	22 571	145 190	18 300
130 001 – 140 000	139 228	18 775	92 969	12 500
140 001 – 150 000	118 495	17 166	148 383	21 500
150 001 – 200 000	356 550	61 420	338 742	58 800
200 001 – 300 000	312 046	75 436	295 105	72 200
300 001 – 400 000	126 927	43 639	105 510	36 200
400 001 – 500 000	60 587	26 923	59 391	26 900
500 001 – 750 000	57 136	34 197	53 938	31 800
750001 - 1000000	19 215	16 454	22 197	19 000
1000001-2000000	18 278	24 412	3 417	4 610
2000001-5000000	5 155	14 533	12 875	28 600
>=5 000 001	853	7 548	None	None
Total	**3 856 999**	**483 008**	**3 028 162**	**483 000**

Source: own calculations using the 2005 IES and income tax tabulations provided by Alvaredo.

There were 3,856,999 taxpayers in 2005 for a total income of R483,008 million. These results are very similar to those found with household surveys. Individuals with an income above R60,000 are 3,028,162 and declare a total income of R483 000 million. There are fewer individuals in the survey than in tax data but the total income is not affected. Individuals

earning less than R120 000 are more numerous in the survey, which leads to a higher total income for these income brackets. The highest income in the 2005 IES goes up to R5,000,000 but not above. The distribution of taxpayers and income is more or less similar between the two data except for the income brackets R1,000,001 – 2,000,001 and R2,000,001-5,000,000. Indeed, there are only 3417 individuals who declare an income between R1,000,001 and R2,000,000 in the survey against 18 278 in tax data. On the contrary, there are 28 600 individuals in the 2005 IES who declare an income between R2,000,001 and R5,000,000 against 5 155 in tax data. This is unlikely to affect the top income shares since the richest 1 percent earned more than R284,108 in 2005. Summing taxpayers and incomes declared above 1 million we find a closer match for total income and for the number of individuals. If differences still exist between the two data we are far from the differences presented by Alvaredo (2007) on Argentina. He found 698 tax files with income above $1,000,000 and 26 tax files with income above $5,000,000 in his tax data. On the contrary, the top 160 individuals in his household surveys only have income between $500,000 and $1,000,000. Also, Székeley and Hilgert (1999) found - in sixteen different countries - that total income of the ten richest households in surveys were very similar to the average wage of a manager of a medium to large size firm. This is definitely not the case with the 2005 IES since the ten richest households earn between 2 and 5 million rand annually. The good quality of the 2005 IES relies mainly on two factors: a large sample of about 25,000 households and an improvement in collecting data compared to previous IES[15]. The 2005 IES can thus be seen as an exception among household surveys. Even if the 2007 tax file and the 2008 NIDS are not directly comparable we decided to do the same comparison to see if the 2005 IES is an exception. The results are presented in table 3. Analyzing this table, we have to keep in mind that the 2008 NIDS is only a panel of 7,302 households.

Table 3: Income tax tabulations and household survey, 2007/2008

	Tax statistics		Household survey	
Income brackets In 2005 rand	**Number of taxpayers**	**Total income, in million 2005 rand**	**Number of individuals**	**Total Income, in million 2005 rand**
70 001 – 80 000	100 755	7 564	283 101	21 200
80 001 – 90 000	116 100	9 885	379 969	32 100

[15] Details concerning the improvements in data quality are provided in section 2, p11.

90 001 – 100 000	123 338	11 717	200 507	19 000
100 001 – 110 000	133 415	14 033	214 635	22 500
110 001 – 120 000	151 633	17 434	187 734	21 600
120 001 – 130 000	125 989	15 733	61 155	7 650
130 001 – 140 000	124 636	16 832	153 118	20 700
140 001 – 150 000	118 913	17 227	80 436	11 600
150 001 – 200 000	376 130	64 647	369 062	44 200
200 001 – 300 000	339 263	82 276	307 918	76 800
300 001 – 400 000	150 712	51 839	49 950	16 500
400 001 – 500 000	77 391	34 419	137 874	64 200
500 001 – 750 000	76 086	45 623	26 411	15 700
750001 - 1000000	25 419	21 779	30 847	29 800
1000001-2000000	23 860	31 849	25 424	28 300
2000001-5000000	7 017	20 088	None	None
>=5 000 001	1 490	14 370	None	None
Total	**2,611,364**	**485 000**	**2,617,651**	**466 000**

Source : own calculations using the 2005 IES and income tax tabulations provided by Alvaredo.

Top incomes seem more underestimated in the 2008 NIDS than in the 2005 IES even if the 2007 figures are affected by incomplete assessment. The higher income in the survey is only above R1,5 million. Yet, there are 25,424 individuals earning between 1 million and 2 millions in the surveys against 23,860 in tax data. Finally there are about 7,000 individuals who earn more than one million rand annually according to tax data and who are not recorded in the survey. If the total income declared above a certain amount is the same in the two data, fluctuations in income brackets are unlikely to affect the top income shares. As in the 2005 IES, the number of individuals and the total incomes are overestimated in the surveys until R120 000 and start being underestimated afterward. These individuals probably underestimate their income, which leads to an overestimation of income and individuals in the first income brackets and to an underestimation in the highest income brackets.

We would like to better understand the implication of this underestimation on the computation of top income shares. In the following table, we compared the total income

declared by top income earners using the income thresholds corresponding to the top 5 percent and to the top 1 percent in tax data. For instance, in 1993 the top 5 percent correspond to individuals earning more than 43 266 rand annually and the top 1 percent to people earning more than 88 972 rand annually. In the table 4 below we compared the total income declared in the two data above the income thresholds computed with tax statistics.

Table 4: Total income declared by individuals belonging to the top 5 percent and the top 1 percent using the income threshold computed from tax statistics.

Year	Individuals earning more than: current rand	Total income in tax data: million current rand	Total income in surveys: million current rand	Discrepancy between tax data and surveys
1993	> 46 366	110 000	46 900	- 57%
	> 88 972	37 400	19 500	- 48%
2000/2002	> 81 513	273 904	140 000	- 49%
	> 207 240	126 889	46 000	- 64%
2005	> 109 704	384 803	360 000	- 6,5%
	> 284 108	167 710	158 000	- 6%
2007 / 2008	> 107 634	434 121	375 000	- 13,5%
	> 317 459	182 700	147 000	- 20%

Source: own calculations using data from 1993 PSLSD, IES 2000, 2005 and NIDS 2008 data sets and income tax tabulations provided by Alvaredo

Firstly, one needs to know that the income thresholds computed from tax statistics diverge from the income thresholds computed from surveys. Using household surveys we found slightly lower income thresholds for the top 5 percent and for the top 1 percent. As a consequence, in the Table 4 above, there are fewer individuals in the household surveys than in the tax tabulations. We also have to keep in mind the Figure 2 where we have shown that the total income declared in each survey is below the one assessed in tax files. Total income was underestimated by 37% in 1993, by 30% in 2000 and by 1% and 2% respectively in 2005 and 2008. In table 4, the discrepancy is higher than the one found in Figure 2 for the total income. This tends to confirm the assumption made by Groves and Cooper (1998): noncompliance and underestimation of income are higher among better-off households. The probability for an individual to underreport her income is not independent of her characteristics.

Finally, one should remember two things from this section. First, the total income declared above the tax threshold in household surveys and in tax data are very similar. If we do not take into account the entire distribution of income, this means that the total income reported in surveys could be used as a good proxy to measure the total households' income or at least to evaluate the overestimation of National accounts estimates. 87 per cent is the higher ratio found between the total households' income declared in survey and in National accounts. This might be used as an upper bound or at least as an indication to highlight the fact that National accounts provide a too large denominator, which might lead to an underestimation of the top income shares. Second, studying the distribution of income we confirm that top incomes tend to be underestimated in surveys and that South African surveys are not an exception. To evaluate the impact on the computation of top income shares we need to define our own income thresholds instead of using those computed with the tax tabulations. We do that in the next section.

The goal of these two previous sections was not to declare either the household surveys or the tax data superior. On the contrary, we have tried to document important similarities and differences between the two data sources. Both data have strengths and weaknesses that need to be understood and carefully considered before attempting to use them to answer any set of research questions.

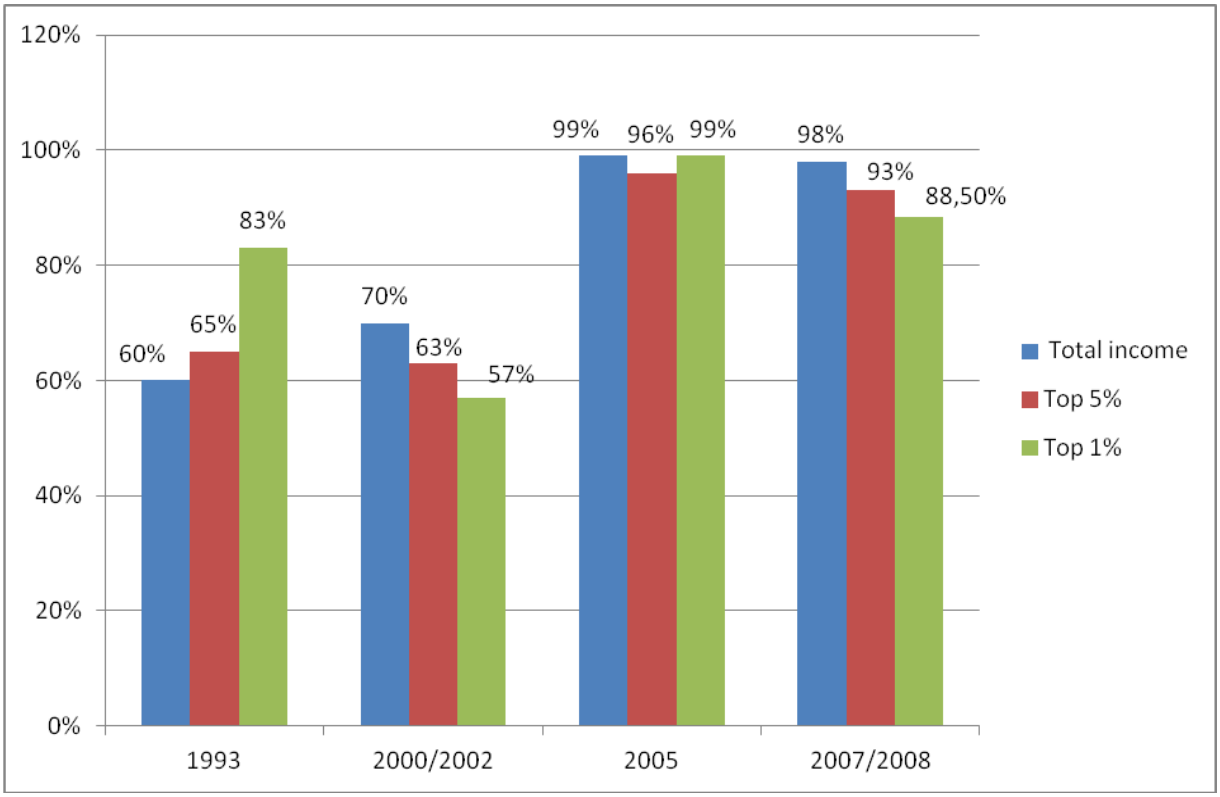
4.3 The computation of top income shares

In this section, we present the computation of top income shares using household surveys. This is a prelude to compare and explain why we find higher top income shares using surveys instead of tax statistics.

To compute top income shares, the first step consists in estimating the income thresholds corresponding to each of the percentiles P90, P95, P99 that define our top income groups. To estimate our income threshold we used the same definition as Alvaredo and Atkinson (2010). Since only a small fraction of the South African population files up a tax return, Alvaredo and Atkinson (2010) need a control total for the population to compute their percentile and income share. They use the adult population defined as all residents aged above 15. We use the same definition in household surveys to compute our income thresholds and

define our top income groups. Since the household surveys use official censuses to compute the post-stratification weights, we have approximately the same number of individuals aged above 15 in the two data. In 1993, Alvaredo and Atkinson (2010) used a control total for population of 25 023 million individuals against 26 681 million according to the 1993 PSLSD survey. In 2005, the population is equal to 31 789 million in the paper of Alvaredo and Atkinson (2010) and equal to 31 867 million in the household survey. Finally in 2007, it is respectively 32 561 million and 32 534 million in the household survey. As a consequence, we nearly have the same number of individuals in the top 5 percent and in the top 1 percent in the two data. Yet, we found slightly lower thresholds with household surveys than with tax data due to the underestimation of total income. These income thresholds are available in the appendix in Table E. In the figure below we compared the total income declared by the top 5 percent and the top 1 percent in the surveys and in the tax data. Compared to the Table 4 above, we have here the same number of individuals in the top income groups but the income thresholds are different for each data.

Figure 3: Total income declared by the top 5 percent and the top 1 percent: household surveys as a percentage of tax data estimates



Source: own calculations using data from 1993 PSLSD, IES 1995, 2000, 2005 and NIDS 2008 data sets and income tax tabulations provided by Alvaredo.

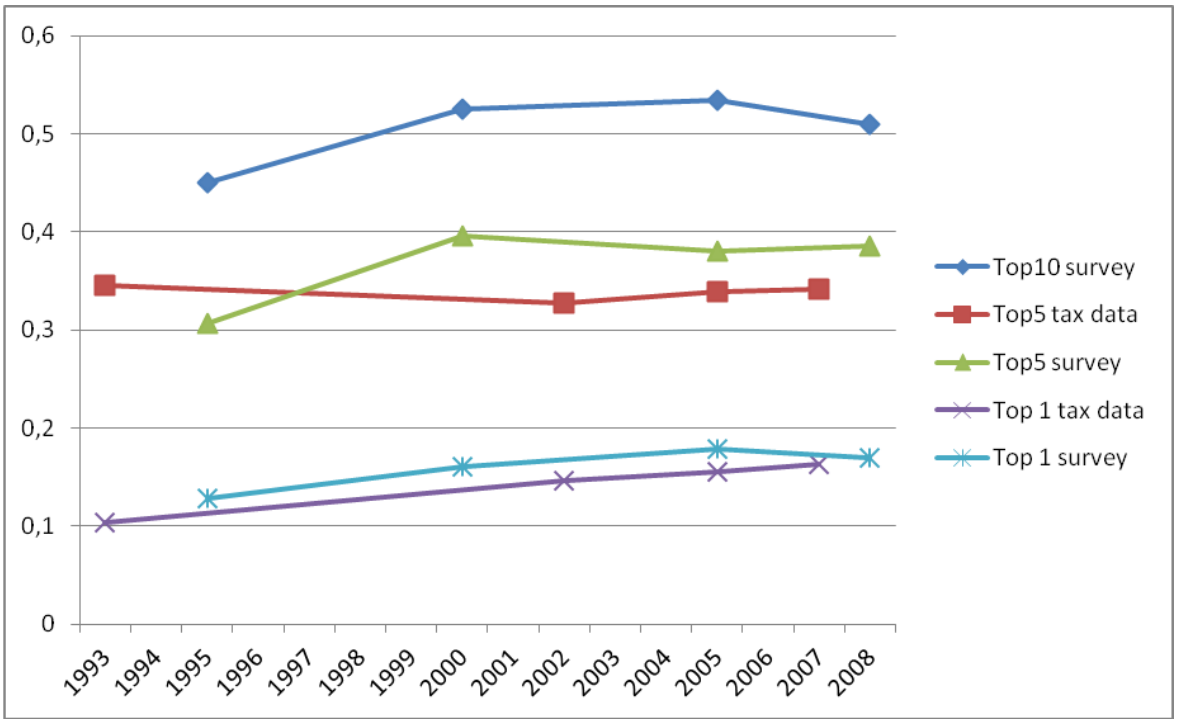
The ratio of coverage for the 1993 PSLDS survey and the 2000 IES are disappointing but not surprising. The 1993 survey was undertaken during a particular context and we compare the 2000 IES with the 2002 tax file. On the contrary, top incomes are well estimated with the 2005 IES and relatively well estimated with the 2008 NIDS. Using the income thresholds computed from survey instead of the income thresholds computed from tax data we managed to overcome the discrepancy found in Table 4. With these different income thresholds we have the same number of individuals in the top income groups. In 2005 and 2008, the differences in top income shares estimates can essentially be explained by the discrepancy between the denominators and not between the numerators. Using the same denominator – total households' income found in National accounts or in household surveys – we found more or less the same top shares estimates. For the 2000 IES the analysis is more complex since we compared two different years. However, we noticed that the top 1 percent and the top 5 percent suffer sometimes more deeply from measurement errors than the total income. If the income growth between 2000 and 2002 was superior for top incomes than on average for total income this can partly explain why the coverage between the 2000 IES and the tax data is lower for top income groups. If the growth was equally shared, this means that top incomes are more underestimated. We decided to keep the 2000 IES in the following section but the results should be analyzed with caution. On the contrary the 1993 PSLDS survey should not be used to study top incomes and we preferred to use the 1995 IES as a proxy for the 1993 tax file.

These methodological sections seem to be a long diversion before analyzing the trend of the top income shares or the racial composition and characteristics of the top income earners. However, it allows us to draw several conclusions. First, some household surveys can be used to study top incomes if they do not significantly underestimate top incomes. In this case, we can better interpret and understand the trend of top income shares. Yet, all household surveys are not comparable and some of them cannot be used to study the entire distribution of income. The issues of noncompliance and inaccuracy in income report are sometimes too high to give a true picture of income inequality.

4.4 Comparing and interpreting the evolution of top incomes in South Africa

In the Figure 4 below, we compared the top income shares found using household surveys and income tax data. The denominator used for tax data is the adjusted total households' income found in National accounts. For household surveys, we used the total households' income reported in each survey. To compute the numerator, we used slightly different income thresholds in the two data in order to have the same number of individuals in each top income group.

Figure 4: Top income shares using tax data and household surveys: 1993 – 2008



Source: own calculations using data from IES 1995, 2000, 2005 and NIDS 2008 data sets and income tax tabulations provided by Alvaredo
Note: the figures are provided in Table F in the appendix.

A comparison between top income share series found with surveys and with tax statistics has already been done. Burkhauser et al. (2009) tried recently to reconcile the Piketty and Saez (2003) top income share series, estimated with tax statistics, with top income shares measured using CPS data. Burkhauser et al. (2009) found close top income share series for the top decile excluding the top percentile. The top 1 percent share measured by the CPS follows the same qualitative trend as the top 1 percent share from tax data but important

qualitative differences remain. This can be explained by several facts: the CPS does not record capital gains or stock option gains which can be an important income sources at the top ; the CPS income are recorded with top code ; income at the top might be underreported. On the contrary in South Africa, we found close but slightly higher estimates for the top 1 percent using household surveys. In 2005, the top 1 percent income share is higher by 2,3% compared to the estimates found in income tax data. The discrepancy is only 1,7% in 2008. We do not face the same issues as Burkhauser et al. (2009) with the CPS. Incomes in IES and NIDS are not recorded with top code but they do report incomes from capital poorly (section 3.2). Burkhauser et al. (2009) also found that the CPS top 1 percent income share increased less than the tax bases top 1 percent income shares. The same conclusion applies itself to our data. Between 1993 and 2007, the share of the top 1 percent increased by 5,9% according to tax data and between 1995 and 2008 the share of the top 1 percent increased by 4,2% according to household surveys. Following the end of the apartheid regime, we find a similar jump in the top 1 percent income share: an increase of 4,3% in income tax data between 1993 and 2002 and an increase of 3,2% in household surveys between 1995 and 2000. We also find higher estimates for the income share of the top 5 percent. The survey top 5 percent income shares increased more than the tax bases top 5 percent income shares. Top income shares exhibit the same trend in the two data. We notice an increase in the top 5 percent income share and an increase in the top 1 percent income share. Yet, in 2008 the income share of the top 5 percent and the income share of the top 1 percent decreased slightly. This can be explained either by a methodological or an economic reason. Compared to the 2005 IES top incomes are more underestimated in the 2008 suvey which can lead to an artificial decrease. The effects of the global financial crisis can also be more significant in 2008 than in 2007 which can explain the decrease in 2008.

The evidence presented about top incomes in Figure 4 bears out the fact that incomes in South Africa are highly unequally distributed. The share of the top 5 percent in gross income is over a third in the two data and that of the top 1 percent is above 15 percent. Alvaredo and Atkinson (2010) show in their paper that there has been a fall in top income shares in South Africa over much of the twentieth century and that incomes within the top groups have become less concentrated. On the contrary, using income tax data and surveys, evidence for the present century suggests that there may be now an upward trend in top income shares. This is compatible with the facts we presented in introduction. Increasing inequality and stable poverty are consistent with the rising trend in top income shares between

1993 and 2008. These increased shares at the top end of the distribution came at the expense of all other income deciles. The cumulative share of income accruing to the first five deciles stayed almost constant but fell from 11,6% in 1995 to 10,15% in 2008. The cumulative share of income accruing to what we could call the "middle class" (p50-90) decreased by almost 5% between 1995 and 2008. These impoverishments of the first nine deciles can explain the increasing share of total income owned by the last decile. Detailed interrogation of ventile shares shows that the increase in the top decile has been driven by a sharply rising share of the top 5 percent of the distribution. This increase is also observed for the percentile p95-99 whereas the income share owned by the percentile p90-95 is relatively constant over time. In 2008, we observe a small decrease in top income shares, which encourages us to study top incomes in the future to see if this decrease is permanent or only temporary.

5. An empirical description of top incomes over the post-apartheid period: between-race inequality is still the main issue in top incomes

We finally reach the last section of this master's thesis. In this section we propose a full description – age, gender, ethnicity, business sector and level of education - of individuals belonging to the top 10 percent, top 5 percent and top 1 percent. At the same time, we would like to examine the assumption described in the introduction according to which South Africa has become a "class society". Using a decomposition of the Theil coefficient into "within-group" and "between-group" components, all of the census-based empirical work - as the one done by Van der Berg and Louw (2003) - makes the same conclusion: between-group inequality declined over the period 1975 to 1996[16]. In 1975, the "between-group" inequality i.e. the interracial contribution to overall inequality was almost twice as important as the "within-group" inequality i.e. the intraracial contribution to overall inequality. By the late 1990s, this ratio was reversed. In our opinion, this theory must be analyzed with precaution. The Gini coefficients provided in the Table A in the appendix show that the greatest inequality is still within the African population. Moreover, some of the declining between-group inequality is due to the significant increase in the African share of the population over the period. Between 1970 and 2001 the African population share increased from 70 percent to 80 percent. On the contrary, the shares of the white group fell from 17 percent of the population in 1970 to 9 percent of the population in 2001. As noted by Leibbrandt et al. (2010), such demographic change gives increasing importance to the intra-African distribution in driving the aggregate distribution. In this section, we would like to see if this increase in intraracial inequality is due to the access of Blacks to top income groups or if it is only due to the widening gap between the new African middle class and the poor composed mainly by Blacks. The issue of intraracial inequality is crucial in South Africa to determine policy initiatives. Indeed, since 1994, deracialisation was a dominant theme in public policy. It covers labour market policies, public education and social welfare policies. The major emphasis was on the promotion of a black economic elite and middle classes via policies of affirmative action and such as the "black economic empowerment" policy. The main objective of affirmative action was to provide higher-paid occupations to historically

[16] The Theil index and its decomposition into "within-group" and "between-group" components are provided in Table B in the appendix.

disadvantaged people and to expand black entrepreneurial. As a consequence, the changing nature of inequality in South Africa could lead to a change in the focus of policy initiatives: from a focus on redistribution between interracial population groups to a focus on the increase in intraracial inequality.

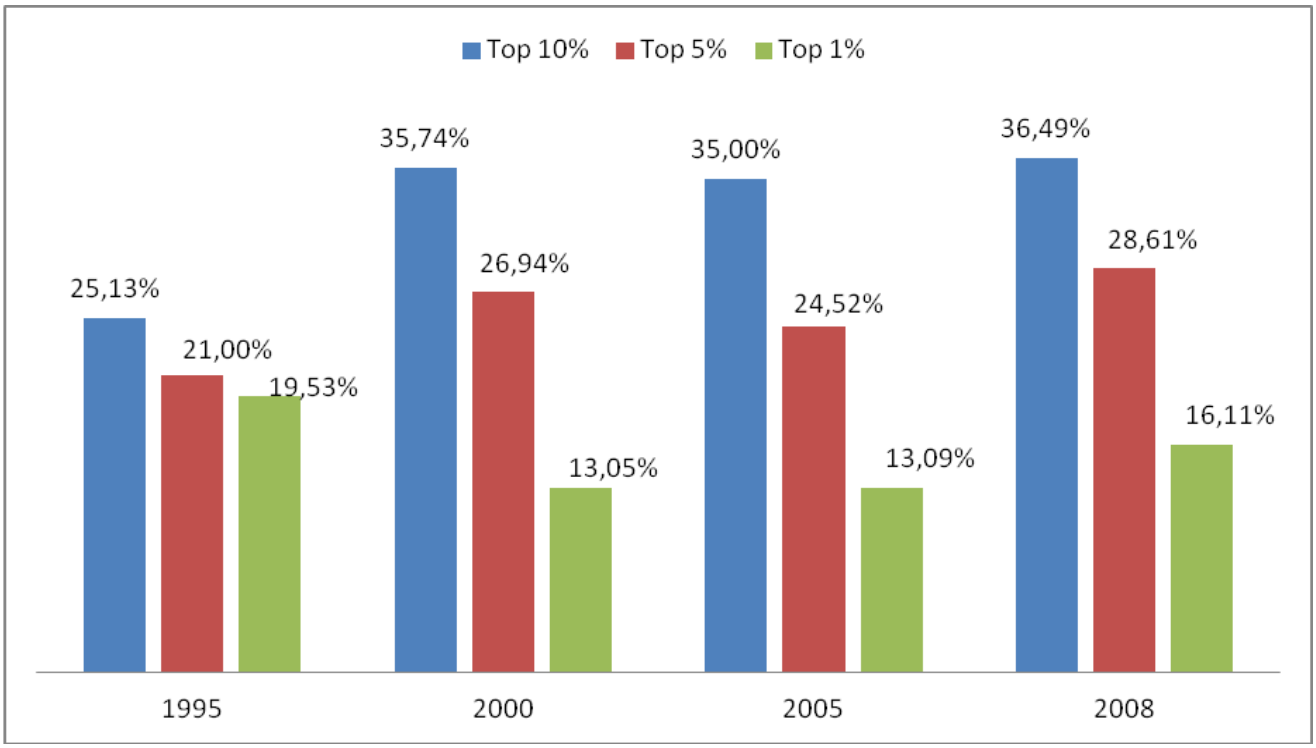
5.1 No room for Blacks in top incomes: evidence of the persistent legacy of apartheid

In 2008, most individuals in the top income groups were between 35 and 60 years old. On average, there were about 43 years old and were slightly older in the top 1 percent than in top 10 percent and in the top 5 percent. The top 10 percent and the top 5 percent are composed of approximately 60% of male whereas the top 1 percent is composed by 84% of males. The age and the gender composition of top income is almost the same between 1995 and 2008. The percentage of women slightly increased – especially between 1995 and 2000 – but the percentage of women in the top 1 percent is never above 20%.

One major difference between South Africa and other countries is that we have to pay attention to the racial composition of top income groups to paint a real picture of their evolution. To define the individuals' population group, the household surveys retain the apartheid-based racial classification of South Africans as Black, Coloured, Indian and White. This classification variable is kept in order to monitor change in the life circumstances of those who were disadvantaged in the apartheid era. From 1956 to 1987, the South African income tax statistics were published with a classification by race. The classification is different from the one used in surveys. The classification in tax statistics is White, Coloured, Asian and Bantu. Bantu was used by the apartheid regime to designate Black South Africans. Whites overwhelmingly composed top income groups from 1956 to 1987. In 1956, Whites were 98,38% of the top 5 percent and 98,49% of the top 1 percent. The composition slightly shifts over the following thirty years. In 1987 Whites were 90,55% of the top 5 percent, 96,65% of the top 1 percent, and 97,5% of the top 0,1 percent. In other words, the last of these figures means that of the 15, 600 tax units in the top 0,1 percent only some 400 were non-White (Coloured, Asian or Bantu). If the percentage of White decreased by almost 8% between 1956 and 1987, the percentage of Bantu in the top 5 percent only increased by 2,4%. After 1987, tax statistics stopped being published with a classification by race. We tried to overcome this issue by using household surveys and decompose the top 10 percent, the top 5

percent and the top 1 percent by population group for 1995, 2000, 2005 and 2008. Doing that, we would like to see if the racial composition of top incomes has changed since the end of the apartheid regime. In the Figure 5 below, we report the figures for Blacks only, excluding Whites, Coloureds and Indians/ Asians.

Figure 5: Percentage of Blacks in the top 10 percent, the top 5 percent and the top 1 percent in 1995, 2000, 2005 and 2008.



Source: own calculations using data from IES 1995, 2000, 2005 and NIDS 2008 data sets

In 1995, Blacks composed 72% of the population aged above 15 years old in South Africa. The out-migration of Whites led to an increase in the share of Blacks in the population: Blacks composed 78% of the population aged above 15 years old in 2000 and 2005 and 77% in 2008. Between 1995 and 2008 the share of Blacks in the top 10 percent rose from 25% to 36,5% and the share in the top 5 percent increased from 21% to 28%. On the contrary, the share of Blacks in the top 1 percent decreased from 19,5% to 16%. These percentages are still very low compared to the racial composition of the population. From the figure 5, we can conclude that between-race inequality remains a central issue in South Africa.

Since the apartheid era prevailed until 1994, we could expect a lower percentage of Blacks in the top 10 percent, the top 5 percent and the top 1 percent than those we found in

1995. Two assumptions can explain these figures. We used household surveys to compute theses figures and – as explained in section 4.3 – we found lower income thresholds than with tax data. In 2008, individuals had to earn more than 107,000 rand annualy to belong to the top 5 percent according to tax data whereas they had to earn more than 104,000 rand annualy according to household surveys. If Blacks are more numerous in the bottom of the top incomes groups, the use of lower income thresholds can lead to an overestimation of the number of Blacks. These figures are also a direct consequence of the specific history of South Africa. The struggle over apartheid started before 1994. Following the oil shock of 1973 numerous organizations in black civil society were established including powerful trade union. Popular demonstrations, insurrection and struggles became recurrent. South Africa was excluded from international financial markets and some multinationals decided to boycott South Africa for their importations and exportations. Economic sanctions were taken by the Commonwealth, by the European communities and by the US congress. These economic and social difficulties lead to a true debate over the social organization of the apartheid and racial policies relaxation was granted in the 1980s. This period is now seen as the "transition era". As a consequence, a few Blacks could have access to top income groups before 1995. Moreover, the brain drain among white professionals and managers in the final years of apartheid have freed up some "space at the top" for black professionals and managers (Seekings and Nattrass, 2005). Concerning the evolution, the percentage of Blacks in the top 10 percent and in the top 5 percent has increased since 1994 but not by a tremendous amount. Between 1995 and 2008 the share of Blacks in the top 10 percent rose from 25% to 36,5% and the share in the top 5 percent increased from 21% to 28%. Yet, since 2000, the percentage of Blacks in the top income groups has remained almost constant. The percentage of Blacks in the top 1 percent has decreased from 19,5% in 1995 to 16% in 2008. Individuals belonging to the top 10 percent are not directly comparable to the one belonging to the top 1 percent. In 2008, individuals belonging to the top 10 percent earned on average 156,000 rand annually whereas individuals belonging to the top 1 percent earned on average 523,000 rand annually. Individuals belonging to the top 10 percent and to the top 1 percent are not likely to work in the same business sector or to have the same job position. We have seen that individuals in the top 1 percent are on average older than the one in the top 10 percent and in the top 5 percent. They are also essentially male. They might be at the end of their career and occupy higher job positions than others. Furthermore, they might earn another type of incomes than their wages. These various profiles can explain the different evolution for the top 10 percent,

the top 5 percent and the top 1 percent. The access to the richest groups is more difficult for Blacks than for the others population groups.

A word of caution: the constant shares of Blacks in the top income groups do not mean a constant percentage of Whites. In fact, the percentage of Whites has decreased since 1995 but essentially in the top 10 percent and in the top 5 percent. In 1995, they were 63% of Whites in the top 10 percent and 69% in the top 5 percent against 47% and 57% respectively in 2008. The detailed decomposition between Blacks, Whites, Coloureds, Asians/Indians is provided in the appendix in table C. This evolution can be explained by two phenomena. First, the share of white people in the population aged above 15 years old has decreased since 1995: from 16,5% in 1995 to 11% in 2008. This decline is also due to an increase in the percentage of Coloureds and Asians/Indians[17]. As a consequence, even if the share of Whites was down by 16% in the top 10 percent between 1995 in 2008, the share of Blacks only increased by 11%. As expressed by a journalist of the *The Southern Times* in 2011 "the reality on the ground is that like the rainbow itself, the Rainbow Nation has no room for black"[18]. This sentence reflects the situation of top incomes in South Africa. If couloured, indian or asian people managed to find their places in top income groups, Blacks seem to be always behind: the percentage of Blacks in top income groups is low and the trend is still hesitant.

In the following sections, we would like to better understand why Blacks still compose a small fraction of the richest South Africans. In order to do so, we study below their business sector and their level of education.

5.2 Business sector of top income earners: is it necessary to deracialise capital ownership?

According to Alvaredo and Atkinson (2010) "the rising trend in top shares in recent years could be associated to the favorable conditions in the world market for agricultural commodities, the increase in the value of minerals other than gold, and the developments in financial sector, as these are the main activities of the richest South Africans in the Sunday Times rich list, but better data are needed to establish a more formal link"[19]. The predominance of the agricultural and the mining sectors could explain the low percentage of

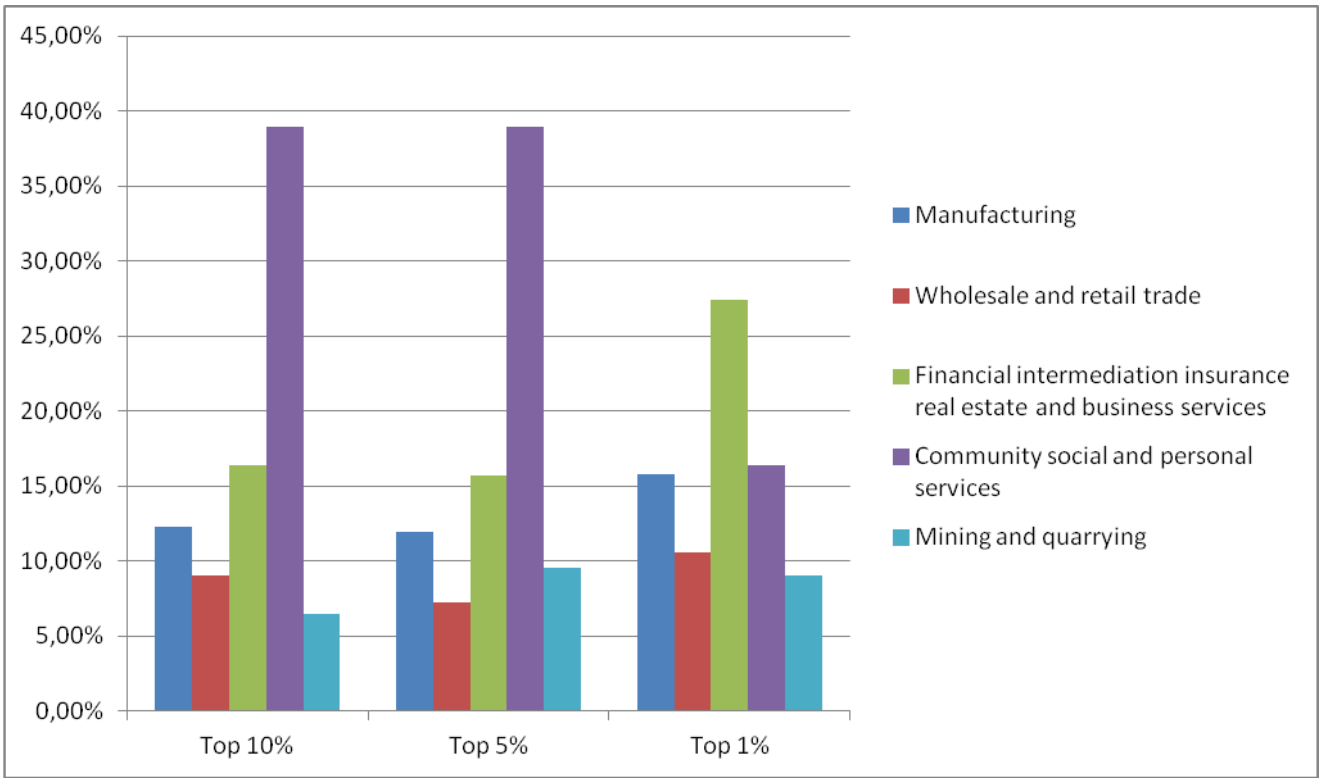
[17] Coloured, asian or indian people constitute a very small sample in household surveys. As a result, they are hard to study and we do not provide an interpretation of their evolutions in the top income groups.
[18] Gabriel Manyati**,** May 2011 "No Black on the rainbow", The Southern Times.
[19] Alvaredo and Atkinson (2010), p20.

Blacks in the top income groups. Indeed, the agricultural and the mining sectors are two sectors where the legacy of apartheid is predominant. The Natives Land Act of 1913 restricted land ownership for Africans to certain specified areas, about 8% of the country's land area. According to the World Bank, in the early 1990s, 67 000 white farmers owned 86% of agricultural land and white commercial agriculture produced 90% of agricultural value-added[20]. In 1995, a land reform started to overcome this unequal distribution. Three plans were started: land redistribution, land restitution and tenure reform. According to Aliber and Mokoena (2003) over the target of 30% of farmland to be redistributed by 1999, only 1% had been transferred by 2002. This can be explained because the land redistribution was based on market transactions and required a willing seller. Restitution concerned only those who were forcibly removed under the apartheid regime. The necessity to have a documentary proof of ownership to obtain a land restitution resulted in slow delivery and in a few rural cases (Gelb, 2004). As a consequence, profits in the agricultural sector still go mainly to Whites. Ownerships in the Mining and quarrying sector are also concentrated among very few people in South Africa. In 1990, six conglomerates centered on mining and finance controlled companies with 80% of the market capitalization on the Johannesburg Stock Exchange (Gelb, 2004). Under the apartheid regime, ownership was restricted for land, houses but also for firms. The 1950 Group Areas Act restricted firm ownership by Blacks: firm ownership was restricted to specified areas; black entrepreneurs could own no more than one business and African firms were restricted to certain markets - mainly retail supply of food and fuel. These restrictions on property ownership prevented Blacks from having access to credit markets. As a result, there were very few black South African firms and basically no firm of a medium-size. In 2002, in the context of the "black economic empowerment" policy, a mining charter was signed for the transformation of the mining sector. The broad-based socio-economic empowerment charter for the South African mining industry and scorecard that accompanies it set two targets for ownership: 15% participation by historically disadvantaged South Africans in terms of equity ownership or attributable units of production within five years and 26% in 10 years. Yet, in 2010, the gross value of black shareholdings within the top 25 mining companies was 5,27% of the total market capitalization[21]. As a result, if the rising trend in top income shares could indeed be associated to the favourable conditions in the world market for agricultural commodities or to the increase in the value of minerals other

[20] World Bank, 1994, "South African Agriculture: Structure, performance and options for the future", World Bank informal discussion papers on aspects of the South African economy, 6, 1994.
[21] *Black ownership in the mining industry: facts & figures 2010*, Conference "A roadmap to development and sustainable growth"

than gold this could explain the low percentage of Blacks in the top income groups. In fact, ownerships in these sectors are highly concentrated due to an absence of redistribution in the years following the end of apartheid. To corroborate this assumption and to obtain more information on the jobs occupied by top income earners, we studied their business sectors.

Figure 6: Business sector for the top 10 percent, the top 5 percent and the top 1 percent in 2008



Source: own calculations using 2008 NIDS data set.

In the 2008 survey, 15,7% of the labor force was employed in the manufacturing sector, 13% in the "wholesale and retail trade", 11% in "the financial intermediation, insurance, real estate and business services", 22% in "community, social and personal services" and finally 5% in mining and quarrying. Following the assumption of Alvaredo and Atkinson (2010) we should find that most rich South Africans are working in "wholesale and retail trade" (which include wholesale and retail trade in agricultural raw materials, livestock, food, beverages and tobacco), in "mining and quarrying" and in "financial intermediation, insurance, real estate, and business services". Most of top income earners work in the manufacturing sector, in "financial intermediation, insurance, real estate and business services", and in "community, social and personal services". Yet, compared to the entire population, individuals in the top 10 percent and in the top 5 percent are overrepresented in

“financial intermediation, insurance, real estate and business services”, in “mining and quarrying” and in “community, social and personal services”.

Most individuals in the top 10 percent and in the top 5 percent are working in the sector coded “community, social and personal services”. Unfortunately, it is also a broad and imprecise industry code. This sector is composed by several divisions: “public administration and defense activities” which includes the general activities of the Central Government and provincial governments, the activities of the armed forces, police and legislative, judicial and administrative departments and offices; “education” which includes public and private education of all types (primary and secondary schools, universities…) provided by institutions and by private teachers; “heath and social work” which includes all types of human health activities such as hospital or clinics activities and medical and dental practice activities and “activities of membership organizations” which includes activities of business, employers’ and professional organizations and activities of trade unions. This vagueness prevents us from drawing some conclusions out of these percentages. The others sectors where top income earners are overrepresented are clearer: “mining and quarrying” and “financial intermediation, insurance, real estate and business services”. 11% of the population work in “financial intermediation, insurance, real estate and business services” against 16,5% of individuals belonging to the top 10 percent, 16% of individuals belonging to the top 5 percent and 27,5% of individuals belonging to the top 1 percent. Concerning the sector “mining and quarrying”, only 5% of the population work in this sector against 6,5% of individuals belonging to the top 10 percent, 9,5% of individuals belonging to the top 5 percent and 9% of individuals belonging to the top 1 percent. The overrepresentation of the richest South Africans in “mining and quarrying” and in “financial intermediation, insurance, real estate and business services” is consistent with the assumption made by Alvaredo and Atkinson (2010).

The high level of employment in “financial intermediation, insurance, real estate and business services” should be linked to a study of capital income. Indeed, finance and business services sectors are jobs where capital incomes might compose a large part of total income. However, we have seen – in section 3.2 comparing household surveys and National accounts – that capital incomes are poorly reported and underestimated in surveys. Using the 2008 survey to study the top 1 percent we found that capital incomes contribute for 13,8% of their total income but we have good reason to think that the real contribution is higher. Incomes generated from capital are always small in surveys but increase across the years among the last decile: its contribution rose from 4% in 1993 to almost 12% in 2008. The lack of information on capital incomes prevents us to precisely study the role of this component in the

increase of top shares. The tax data tabulations cannot be used because they only provide the total income assessed by taxpayers. However, we know that registered individual taxpayers employed in the "finance, insurance, real estate and business services" account for about 20 percent of the total income assessed for the period 2003 to 2008[22]. The number of individuals registered as taxpayers within the sector grew from 545 024 in 2003 to 674 769 in 2006. Moreover, the article IV of the IMF also adds that financial institutions have enjoyed good profitability, capitalization levels and reserves financial assets in 2008. In sum, we can assume that the increase in capital incomes has contributed to the rise in top income shares but how much cannot be measured exactly. This increase in capital incomes has probably benefited mostly to the top 1 percent since they are more numerous to work in the financial sector.

In contrast with the agricultural and the mining sector, "financial intermediation, insurance, real estate and business services" might be less affected by the laws implemented under the apartheid regime. In that case, the business sector is insufficient to explain why there are only 16,1% of Blacks in the top 1 percent while almost 28% of individuals belonging to the top 1 percent work in this sector. The level of education of top income earners might help us to better understand the Figure 5.

5.3 The level of education of top income earners: the role of human capital in wealth accumulation

In South Africa, education is organized as a three tier system starting with primary school called "general education and training", followed by high school classified as "further education and training" and tertiary education, "higher education and training", in the form of academic universities and universities of technology. Primary schools comprise Grade 0 to 9 and High schools Grade 10 to 12. Grade 12 is the year of matriculation. Since 1996, education is compulsory for all South Africans until the age 15, or the completion of grade 9. The tertiary education, above grade 12, includes education for undergraduate and postgraduate degrees, certificates and diplomas, up to the level of the doctoral degree. A matric endorsement is required for the study of university degrees.

[22] 2008 Tax Statistics, National Treasury and the South African Revenue Service

Education is one of the first priorities in South Africa and is seen as a solution to overcome the inequalities inherited from the apartheid regime. Under the Bantu Education Act of 1953, white children received a quality schooling while black children have access to a "Bantu education" which provided only the basic knowledge. This education was focused essentially on limited technical and vocational skills. The Extension of University Education Act of 1959 restricted the admission of black students to white institutions. Blacks were excluded from English-language universities and admitted only to segregated "bush colleges" set up in the 1960s. A direct consequence is the high rate of illiteracy in South Africa: 18% of adults over 15 years old are unable to read or write in 2010. The matric pass rate remains low, it was 40% in the 1990s and 60% in 2008. Despite the improvements, the legacy is still present: in 2009, 65% of Whites over 20 years old have a high school or higher qualification, while this is only the case for 14% of Blacks and 17% of Coloured[23].

The level of education among top income earners can help us to analyze the low percentages of Blacks in the top income groups. Individuals belonging to the top income groups are on average 40 years old in 2008, which means that they were about 26 years old in 1994. They went to school under the apartheid regime. If the richest South Africans are all highly skilled, this could explain why Blacks are less numerous. In the Table 5 we provide the highest level of education for individuals belonging to the top 10 percent, the top 5 percent and the top 1 percent in 2008.

[23] South African Institute of Race Relations, *South African survey*, 2009

Table 5: Human capital: highest level of education attained for individuals belonging to the top 10 percent, the top 5 percent and the top 1% in 2008.

Highest level of education	Top 10%	Top 5%	Top 1%
Grade 12 or below			
Below Grade 12	14,2%	10%	1,2%
Grade 12	61,8%	56%	46,2%
Total	**76%**	**66%**	**47,4%**
Above Grade 12: Higher education and training			
Bachelor degree	10,3%	15,3%	11,9%
Honour degree	5%	6,3%	7,4%
Higher degree (Masters Doctorate)	7,7%	12,4%	33,3%
Total	**23%**	**34%**	**52,6%**

Source: own calculations using 2008 NIDS data set.

Individuals in the top income groups have a relatively high level of education. In the last decile and in the top 5 percent, 10% of the individuals did not finish high school but the percentage is reduced to 1,2% in the top 1 percent. The matric pass rate was low in the 1990s and people who have only Grade 12 finished at least the final year of high school. In the top 1 percent, almost 35% have either a Master or a Doctorate. The level of education seems to increase with the level of wealth since individuals belonging to the top 1 percent are more educated than the one belonging to the top 5 percent who are also more educated than the one belonging to the top 10 percent. Our assumption is that these highly skilled workers benefited from the strong increase in labor force participation following the end of apartheid. This increase was essentially driven by a low-skilled and undereducated black population. Their scarce skills and qualifications might allow them to obtain higher wages relatively to low-skill workers. Unemployment in South Africa was over 23 percent in 2007and the unemployment rate is higher for Blacks (26,8%) than for Whites (3,9%)[24]. The OECD Economic survey (2009) highlights signs of a growing dualism in labour markets, with large union wage premia and rising real wages in large formal sector firms, but broadly stagnant economy-wide real wages. Some part of the union wage premium might be linked to the existence of substantial

[24] OECD Economic surveys (2009)

product market rents, which allows firms to pay wages above the competitive level. Expanding opportunities in high-paying occupations and at the same time growing unemployment resulted in an increase in top incomes shares over the last fifteen years. A high level of education seems to be a common characteristic of top income earners. Their levels of education probably offer them job opportunities and high wages that Blacks cannot obtain given their lower levels of education. Individuals lacking social and human capital are more likely to be shut out of the labour market.

An encouraging conclusion would be that the only criterion necessary to be rich in South Africa is to be highly educated. Indeed, Blacks are progressively catching up their education backwardness due to the apartheid regime. In 2008, out of a total number of 80,803 degrees awarded in South Africa, 45.8% of these were awarded to Blacks, 39% to Whites, 8.5% to Indians and 6.5% to Coloureds[25]. If the new generation of Blacks becomes as educated as the White, we should expect the discrimination in the top income groups to slowly fade overtime. If equal opportunities are provided for each ethnicity, racial imbalance could be eliminated. Over the longer term, improvements in education might be the key to reduce the excess supply of less-skilled workers and to improve the racial diversity among the top income groups. Nevertheless, if discrimination persists this would mean that education and human capital were not the only criteria. According to the Commission for Employment Equity's and their annual report for 2008-2009 there are not the only criteria. The Commission for Employment Equity is a statutory body established to advise the Minister and is required to submit an annual report to the Minister of Labour on the implementation of employment equity. Employers with 150 or more employees (i.e. large employers) are required to submit reports to the Department on an annual basis. Employers with fewer than 150 employees (i.e. small employers) are expected to report every two years. All employers, i.e. both large employers and small employers were expected to report in 2008. 10 580 reports were received in 2008 and 7 229 were analyzed covering about three millions employees. In the table 6 below we show the representation of employees for all employers in each occupational level by gender and by population group.

[25] South African Institute of Race Relations , South African survey, 2009

Table 6: Representation of employees (including people with disabilities) for all employers in each occupational level

Occupational level	Male				Female			
	Black	Coloured	Indian	White	Black	Coloured	Indian	White
Top management	9.8%	3.5%	4.8%	61.1%	3.8%	1.2%	1.1%	11.7%
Senior management	11.9%	4.6%	6.0%	47.4%	5.4%	2.3%	2.3%	17.8%
Professionally qualified and experienced specialists and midmanagement	16.5%	4.9%	5.6%	33.2%	11.5%	4.5%	3.3%	18.4%
Skilled technical and lower management	30.9%	6.5%	3.7%	18.5%	18.9%	5.1%	2.7%	12.5%
Semi-skilled and discretionary decision making	46.3%	6.3%	2.0%	4.0%	23.0%	7.0%	1.9%	5.5%
Unskilled and defined decision making	56.0%	5.3%	0.6%	0.8%	27.3%	5.0%	0.4%	0.4%

Source: The Commission for Employment Equity's annual report for 2008 – 2009

The economically active population in South Africa is constituted mainly by 74.1% of Blacks, followed by 12,1% of Whites, 10,8% of Coloureds and 3% of Indians. In terms of gender, males and females are relatively evenly distributed at 54% and 46% respectively. Blacks are the only population group for which the economically active population lags behind their national population distribution. From Table 6, we notice that for "Top Management positions", Black males, Black females and Coloured females show the largest deficit gap compared to their representation in the active population. On the contrary, White

males are overrepresented at this level, followed by White females and Indian males. We also observe that at the skilled level – "skilled technical and lower management" – Blacks constitutes the major part of employees. Yet at the next level – "Professionally qualified and experienced specialists and mid-management" – their representation is extremely lower. According to the Commission's report, "this indicates that the opportunity to move up is disproportionately favouring the White group"[26]. Even if Black people constitute the major part of the professionally qualified and skilled employees, they do attain higher levels. The Commission's report also shows that even in the disability group White people are still being disproportionately preferred. As a result, the different levels of human capital might be not sufficient to explain the low percentage of Blacks in top income groups. Social networks and social capital – as presented by Burt (1992) – might play a significant role in distributing job opportunities where wages are particularly high. Moreover, keeping only the private sector and excluding the State owned enterprises, the number of Black is smaller in top echelons. The inclusion of the State owned enterprises data positively influenced the final result. The sector "community, social and personal services" studied in section 5.2 contains all individuals employed in the public administration and in the education sector. Black individuals employed in the public sector are more likely to access top management positions but there are less likely to belong to the top 1 percent since wages offered in the public sector are often lower than in the private sector. Most individuals in the top 1 percent work in the private sector in the "financial intermediation, insurance, real estate and business services", and part of their wages comes in the form of capital income which can explain their higher income. In this business sector Blacks are less likely to have access to top management positions. This table is consistent with our Figure 5: White people, man or woman, continue to dominate the top echelons.

The different levels of human and social capital can partly explain the low percentage of Blacks in top income groups. We should not neglect also the role of capital ownership in specific sectors such as the mining and quarrying sector, the agricultural sector or even the financial sector. The absence of redistribution in these sectors after the apartheid regime tends to amplify the role of capital and inherited wealth in determining the belonging to top income groups. Promotion of black owners and managers could be intensified with eventually a "de-conglomeration" of South African business (Gelb, 2004).

[26] The Commission for Employment Equity's annual report for 2008 – 2009, p6.

6. Summary and conclusion

This paper shows that South Africa's inequality levels are still among the highest in the world and that inequality continues to bear a persistent racial undertone. The share of the top 10 percent in gross income is over a half, that of the top 5 percent is over a third and that of the top 1 percent is above fifteen percent. Focusing on top incomes, we highlighted the fact that racial inequalities are still very important in the post-apartheid South Africa. The richest South Africans are mainly white and only a few Blacks belong to the top income groups.

In order to draw these conclusions, we used five household surveys to compute the top income shares and to provide a full description of top income earners. We also used those household surveys, instead of tax statistics, to better understand the evolution of top income shares. This choice required a rigorous methodological approach. Household surveys are often considered as not reliable to study top incomes. The level of quality of each survey has to be assessed before using it. In order to do so, we have compared the top income shares obtained using household surveys and tax data. We have found a similar trend but the top income shares were larger with surveys. The goal of section 3 and 4 was to understand why we found larger top income shares. More precisely, we wanted to know if the overestimation of top income shares was due to an underestimation of top incomes in surveys. To compute the top income shares with tax statistics, Alvaredo and Atkinson (2010) use as a denominator the total households' income reported in National accounts. In section 3, we have compared this denominator with the one we are using: the total households' income declared in household surveys. Doing so, we found that the total households' income declared in surveys is underestimated compared to the one reported in National accounts. We can explain these discrepancies between the two estimates by the biases existing in our surveys – noncompliance and income underestimation – and by the significant differences between income concepts used in National accounts and in household surveys. Discrepancies between National accounts and household surveys are expected and can be tolerated in our paper because they are not too important. The coverage between the denominator found with household surveys and the denominator used by Alvaredo and Atkinson (2010) varies from 87% in 1995 to 78% in 2008. The lowest ratio of coverage is 61% in 1993 and the highest is attained in 1995. These results were the first sign of the relative good quality of the surveys we used. Our ratios of coverage with National accounts estimates were higher than the one found by others authors for many developing countries. In order to strengthen this conclusion, the section 4 displays the entire distribution of income – by income brackets – in tax statistics

and in surveys. The aim was to evaluate if top incomes were underestimated in surveys. We showed that top incomes are underestimated in surveys compared to tax data and that the accuracy of the distribution can fluctuate over income brackets. Yet, the total incomes declared by the top 10 percent and above are more or less similar in surveys and tax statistics. As a consequence, we argued that the discrepancies in top incomes shares in 2005 and 2008 are mainly due to the choice of a lower denominator and not to the underestimation of top incomes in surveys. In 2005, the total income declared by the top 10 percent is equal to 99% of the one reported in tax statistics. In 2008, this ratio is equal to 98% but the total income declared by the top 1% is only equal to 89% of the one reported in tax data. The impossibility to compare the other years with tax files prevented us from generalizing this conclusion. Nevertheless, it seems that the other surveys for 1993 and 2000 seriously underestimate top incomes. This can be explained because in 1993, the survey was undertaken during the democratic transition of South Africa, a specific context of violence and political instability and because between 2000 and 2005, true improvements have been done to enhance the quality of the data collected. These methodological sections seem to be a long diversion before analyzing the trend of the top income shares or the racial composition and characteristics of the top income earners. However, it allows us to conclude that household surveys can be used to study top incomes if one should previously verify that the surveys do not significantly underestimate top incomes. In order to do that, a comparison with tax statistics is necessary. If household surveys have some drawbacks, using this type of data allows for a broader range of research than is possible with tax data. In this case, we can better interpret and understand the trend of top income shares. Nevertheless, all household surveys are not comparable and some of them cannot be used to study the entire distribution of income because the issues of non compliance and inaccuracy in income report are too high to give a true picture of income inequality.

After these methodological sections, we were able to provide an empirical description of top income earners for 4 years: 1995, 2000, 2005 and 2008. We showed that between-race inequality remains the main issue for top incomes. Indeed, despite an increase in the share of Blacks in the population - from 76,7% in 1995 to 79,2% in 2008 – top incomes are still mainly composed of Whites. Between 1995 and 2008 the share of Blacks in the top 10 percent rose from 25% to 36,5% and the share in the top 5 percent increased from 21% to 28%. On the contrary, the share of Blacks in the top 1 percent decreased from 19,5% to 16%. These percentages are still very low compared to the racial composition of the population and their trends provide evidence of the legacies of apartheid. Capital ownership is still concentrated

among few people in South Africa and the policies established under the Apartheid regime such as jobs reservation, spatial and property restrictions and discrimination in the education, health and social services have created a workforce with racially skewed human, social and economic capital. Studying top incomes we showed that such economic and human capital legacies leave a very long-run footprint. These processes are hard to reverse. Individuals belonging to the top income groups were on average 40 years old in 2008 which means that they were about 26 years old in 1994. They went to school under the apartheid regime. We showed= that the richest South Africans are all highly skilled: about 53% of individuals belonging to the top 1% have a university degree. We explained the low percentage of Blacks in the top income groups by three factors: ownership concentration in business sectors where top income earners are overrepresented, human capital with the high level of education of top income earners, and social network and social capital which could explain why top management positions are essentially reserved to Whites. These three factors seem to be more powerful in the richest category – the top 1 percent – where the financial sector is prevailing. The access to the top 10 percent and to the top 5 percent for historically disadvantaged racial groups have been promoted in the public sector and in State owned enterprises. An extension of this paper could use econometric tools such as quantile regressions to measure exactly the role of each factor – gender, age, ethnicity, level of education and business sectors - in determining income. Doing that, our conclusion could be improved in the future.

Even if we observe a decline in the importance of between-race inequality and an increase in within-race inequality we insist on the importance of between-race inequality in South Africa and particularly among the Rich. The between-race component remains high and its decline has slowed since the mid 1990s. The bottom deciles of the income distribution and the poverty profile are still dominated by Blacks (Leibbrandt et al., 2010) whereas top incomes are still overwhelmingly Whites. As a consequence, we conclude that we are far from the new "class society" described by some economists. Given the decrease in between-race inequality and the increase in the number of Blacks in the total population – Blacks accounts for almost 80% of the population and this share is rising – the policy focus on race-based redistribution is progressively replaced by a focus on the increasing inequality within each race group. Yet, this study of top incomes has shown that between-race inequality is still an issue in South Africa. As a consequence, policies such as the Broad-based Black Economic Empowerment should be intensified and not relaxed to support the access of Black to top incomes through the promotion of black owners and managers.

APPENDIX

1. Inequality in South Africa: Gini coefficient and Theil index

Table A: Gini coefficients for households' direct income by race, 1993 - 2008

Gini coefficient	1993	1995	2000	2005	2008
All households	**0,59**	**0,57**	**0,62**	**0,64**	**0,60**
African	**0,51**	**0,55**	**0,57**	**0,56**	**0,55**
White	**0,43**	**0,42**	**0,46**	**0,50**	**0,46**

Source: own calculations using data from 1993 PSLSD, IES 1995, 2000, 2005 and NIDS 2008 data sets.
Notes: The Gini coefficient is computed using the direct annual households' income and taking into account only positive values.

Table B: Decomposition of the Theil-T-index for income inequality, 1975 - 2008

Theil coefficient	1975 (%)	1993 (%)	1995 (%)	2000 (%)	2005 (%)	2008 (%)
Within-group inequality	**38**	**57**	**56**	**54**	**56**	**57**
Between-group inequality	**62**	**43**	**44**	**46**	**44**	**43**
Total	**100**	**100**	**100**	**100**	**100**	**100**

Source: Seeking and Natrass (2005) and own calculations

Table C: Top incomes' decomposition by race: figures used in Figure 5

Ethnicity decomposition of top incomes in %	Top 10%	Top 5%	Top 1%	Total individuals aged above 15
1995				
Blacks	25,1	21	19,5	72,3
Coloureds	5	4,2	2,5	8,5
Indians/ Asians	7,1	6,56	6,6	2,7
Whites	62,75	68,2	71,5	16,4
2000				
Blacks	35,7	26,9	13	78,5
Coloureds	10,1	7,9	4,7	8,8
Indians/ Asians	6,2	6	5	2,7
Whites	47,2	58	76,1	9,7
2005				
Blacks	35	24,5	13,1	77,4
Coloureds	9,3	7,9	3,5	8,8
Indians/ Asians	5,5	5,96	3,1	2,7
Whites	49,8	61,5	80,3	11
2008				
Blacks	36,6	28,6	16,1	76,8
Coloureds	9,7	7,2	3,63	9,2
Indians/ Asians	6,6	7,3	13,85	2,9
Whites	47	56,76	66,4	11

2. Computation of top income shares

Table D: Tax income threshold, level of annual income below which no income tax is payable, current prices, rands

Tax threshold	1993/1994	2002/2003	2005/2006	2007/2008
Tax threshold below age 65	10 714	27 000	35 000	43 000
Tax threshold age 65 and over	11 285	42 640	60 000	69 000

Source: South African Revenue Service

Table E: Income thresholds corresponding to the P95 and P99 used to define the top income groups in tax statistics and household surveys. Computed among individuals above aged 15. Nominal rands

Year of the household survey / fiscal year	Income threshold according to tax data		Income threshold according to household surveys	
	Top 5%	Top 1%	Top 5%	Top 1%
1993	43 266	88 972	22 800	61 200
2000			66 000	156 000
2002	81 513	207 240		
2005	109 704	284 108	106 561	265 356
2008 / 2007	107 634	317 459	104 000	279 855

Source: own calculations for household surveys using data from 1993 PSLSD, IES 2000, 2005 and NIDS 2008 data sets. Computations from Alvaredo for tax statistics.
Notes: For example, individuals belonging to the top 5 percent in 1993 earn more than 43 266 rand annually.

Table F: Top income shares find using income tax data and household surveys, 1993 – 2008

Year of the household survey / fiscal year	Top income shares using income tax data			Top income shares using household surveys		
	Top10%	Top5%	Top1%	Top10%	Top5%	Top1%
1993	n.a	35%	10,3%	*47%*	*36%*	*15%*
1995	n.a	n.a	n.a	45,05%	30,6%	12,83%
2000	n.a	n.a	n.a	52,55%	38,1%	16%
2002	n.a	32,7%	14,6%	n.a	n.a	n.a
2005	n.a	33,86%	15,50%	53,4%	39,6%	17,8%
2008	n.a	34,10%	16,2%	51%	38,6%	17%

Source: own calculations for household surveys using data from 1993 PSLSD, IES 1995, 2000, 2005 and NIDS 2008 data sets. Alvaredo and Atkinson for the income tax data's results.

3. Computation of weights in surveys

Two weights are available in the data: the design weight and the post stratification weights. Two sets of calculations were necessary in deriving the design weights. First there is a calculation of the probability of sampling each primary sample unit (PSU) and, second, there is a calculation about the probability of including each specific household in each PSU. The latter corrects for household non-response. The PSU inclusion probability is given by $P_{PSU} = \frac{n_{PSU}}{N_{PSU}}.n_S$, where n_{PSU} is the number of households constituting the selected PSU during census fieldwork, n_S is the number of PSUs per stratum, and N_{PSU} is the number of households constituting the selected stratum during census fieldwork. The household

inclusion probability per PSU is given by $P_{HH} = \frac{n_{HH}}{H_{HH}}$, where n_{HH} is the number of selected dwelling units per PSU, and H_{HH} is the number of dwelling units in the PSU in question at a particular time different from the census time. The non-response adjustment factor is given by $\frac{1}{r_{HH}}$, where r_{HH} is the response rate and is given by $r_{HH} = \frac{n_{RESP}}{n_T}$ where n_{RESP} is the number of responding households and n_T is the total number of visited households (in the sampled dwelling units) per PSU. The design weights adjusted for non-response are then given by $W_{HH} = \frac{1}{P_{PSU}.P_{HH}.r_{HH}}$

REFERENCES

Aaberge, Rolf, and Atkinson Anthony B. 2010, "Top Incomes in Norway." In *Top Incomes: A Global Perspective*, ed. Anthony B. Atkinson and Thomas Piketty, 448–81. Oxford and New York: Oxford University Press

Aliber, M and Mokoena R, 2003, "The land Question in contemporary South Africa". In J Daniel, A Habib & R Southall (eds), *The State of the Nation: South Africa, 2003-2004*. HSRC Press, Pretoria

Alvaredo Facundo and Anthony B Atkinson, 2010, "Colonial Rule, Apartheid and Natural Resources: Top Incomes in South Africa 1903 – 2005", Oxcarre Research Paper 46

Alvaredo Facundo, 2007, "The rich in Argentina over the twentieth century: From the conservative republic to the Peronist experience and beyond 1932-2004", Working paper N° 2007 – 02

Atkinson, Piketty and Saez, 2011, "Top Incomes in the Long Run of History", *Journal of Economic Literature,* 49:1, 3 – 71.

Burkhauser Richard, Feng Shuaizhang, Jenkins Stephen P, Larrimore Jeff, 2009, "Recent Trends in Top Income Shares in the USA: Reconciling Estimates from March CPS and IRS Tax Return Data" NBER Working Paper No. 15320

Burt Ronald Stuart, 1992 *Structural Holes: The Social Structure of Competition*, Harvard University Press

Deaton, August, 2005, "Measuring poverty in a growing world (or measuring growth in a poor world)", *The review of Economics and Statistics*, February, 87(1): 1 -19.

Dinkleman, Taryn. 2009. "The Effects of Rural Electrification on Employment: New Evidence from South Africa." Princeton University, mime

Gelb Stephen, 2004, "Inequality in South Africa: Nature, Causes and Responses", Policy Initiative on Addressing Inequality in Middle-income Countries, The EDGE Institute, Johannesburg

Glaser, D., 2007, "Should an egalitarian support Black Economic Empowerment?", *Politikon,* vol 34, n°2, pages 105 – 123

Groves and Cooper, 1998, *Nonresponse in Household Interview Surveys*, New York: Wiley

Johnson Barry and Moore Kevin, 2008, "Differences in Income Estimates derived from survey and tax data", paper presented at the American Statistical Association's 2008

Kuznets, Simon, 1953, *Shares of Upper Income Groups in Income and Savings*. Cambridge, Mass.: National Bureau of Economic Research.

Leibbrandt, M. *et al.*,2010, "Trends in South African Income Distribution and Poverty since the Fall of Apartheid", *OECD Social, Employment and Migration Working Papers*, No. 101, OECD Publishing.

Ravallion Martin, 2001, "Measuring aggregate welfare in developing countries: how well do national accounts and surveys agree" *The review of economics and statistics*, 85 (3), 645-652

Roine, J and Waldenström, D, 2005, "Top incomes in Sweden over the Twentieth Century", Working Paper, Stockholm School of Economics.

Seeking and Natrass, 2005, *Class, Race and Inequality in South Africa*, University of Kwazulu – Natal Press

Simkins Charles, 2004, "What happened to the distribution of income in South Africa between 1995 and 2001?", University of the Witwatersrand

Statistics South Africa, Pali Lehohla Statistician-General, 2008, *Income and expenditure of households 2005/2006: Analysis of results*, Report No. 01-00-01, ISBN No. 978-0-621 37635-7

Surjit Bhalla, 2002, "Surveys and National Accounts: Can a Choice Be Made?" In *Imagine There's No Country: Poverty, Inequality, and Growth in the Era of Globalization*, 288 pp. ISBN paper

Székeley, M. and Hilgert, M., 1999, "What's behind the inequality we measure: an investigation using Latin American data". Research Department Working, Paper 409. Inter-American Development Bank.

Sampie Terreblanche, 2004, *A history of Inequality in South Africa, 1652 – 2002*, University of Natal Press

Van der Berg and Louw, 2003, "Changing patterns of South African income distribution: towards time series estimates of distribution and poverty", *South African Journal of Economics*, 72(3): 546-572.