DRAFT - DO NOT REFERENCE OR QUOTE WITHOUT THE AUTHOR'S PERMISSION

Subjective well-being and reference groups in post-apartheid South Africa

Marisa von Fintel*

August 24, 2015

Abstract

Previous studies on the determinants of subjective well-being concur on the importance of relative income, acknowledging the fact that individuals evaluate their own well-being in light of the performance of a well-defined reference group. Research using data from South Africa during apartheid has found that reference groups were greatly divided along racial lines, with same-race relative standing being much more important than the relative standing *vis-à-vis* other race groups. Using data from the first wave of the National Income Dynamics Study in 2008, this paper updates and expand these previous findings in order to come to a better understanding of the appropriate definition of reference groups within a divided society such as South Africa. It makes use of a nonlinear model in which various parameters are estimated in order to determine the weight placed on the well-being of individuals in the same race group as the respondent versus all the other race groups living in one of three specified geographic areas. The main methodological innovation of this approach is that it allows the various parameters to interact within the utility function and also takes a more flexible approach in the estimation of the weight placed on others who are of the same race, allowing it to take a range of values. The findings seem to suggest that reference groups have shifted away from a racial delineation to a more inclusive one subsequent to the country's first democratic elections in 1994. Although most of the weight is still placed on same-race relative standing, the estimates suggest that weight is also placed on individuals from other race groups. The paper also examines the spatial variation of reference groups and finds evidence that the relative standing of close others (such as neighbours) enter the utility function positively while individuals who live further away (strangers) enter the utility function negatively. The robustness of these findings is tested with alternative specifications and income measures.

JEL Classification: I30, I31, D60, D63

Key words: Subjective well-being, happiness, reference groups, poverty, post-apartheid, South Africa

*marisa.vonfintel@gmail.com, Department of Economics, University of Stellenbosch, Stellenbosch, South Africa

1 Introduction

The importance of subjective well-being as concept in economics has by now been firmly entrenched (Stutzer and Frey, 2010). Central to the understanding of subjective well-being and its correlates is the role of the individuals' relative well-being, or relative standing, compared to some well-defined reference group.

The relationship between relative standing and subjective well-being are important within economics for various reasons. First, it provides insight into the determinants of subjectively measured individual utility, which in turn challenges the traditional axiomatic revealed-preference models, in which individuals' utility is objective and only observable through their independent, rational, utility maximising choices (Kapteyn, Van Praag and Van Herwaarden, 1976; Frey and Stutzer, 2002; Dolan, Peasgood and White, 2008; Stutzer and Frey, 2010).[1] The relationship between relative income and subjective well-being has also challenged traditional views of utility that "more is always better" - increases in absolute income do not unequivocally lead to higher subjective well-being or utility (Ferrer-i-Carbonell, 2005). Furthermore, research on relative standing provides insight into the definition of poverty and how relative poverty may provide a more comprehensive measure of the number of people who are deprived in a specific country by constructing what Ravallion (2012) calls a "social subjective poverty line"[2] (Pradhan and Ravallion, 2000; Ravallion and Chen, 2011).[3] Last, it provides insights into the impact of inequality on society, and reasons for why inequality is detrimental to society (Van Praag, 2011; Card, Mas, Moretti and Saez, 2012).

Although various studies have been conducted on the importance of relative well-being in developed economies, the literature regarding relative standing and reference groups in emerging and poor economies is still in its infancy. A few studies have focussed on ascertaining the importance of relative versus absolute income in countries such as India, Nepal, and Ethiopia.[4]

In the South African context, most of the research on subjective well-being has focussed on the apartheid era - the period prior to the first democratic elections on 27 April 1994 and subsequent new political dispensation.[5] Kingdon and Knight (2006, 2007) examine the question

[1] Instead, what is proposed by Kapteyn, Van Praag and Van Herwaarden (1976) is the "individual welfare function", which takes cognisance of the well-being of others.

[2] Ravallion (2012) explains how, although subjective well-being measures in themselves cannot be used as poverty lines, the relativeness of the subjective well-being measure may be used to calibrate multidimensional measures of poverty to find the point above which individuals stop feeling poor.

[3] The concept of a relative measure of poverty has come as least as far as Adam Smith, who commented on the differences in perceived deprivation among children in Scotland and England. In Scotland, children were not perceived to be deprived if they did not own a pair of shoes, while in England only the severely deprived did not own a single pair of shoes (Clark and Senik, 2011).

[4] An overview of this literature has been conducted by Stutzer and Frey (2010).

[5] There has been some work using 2008 data, for example Posel and Casale (2011) and Posel (2014).

of reference groups and relative standing in South Africa in 1993. Two main conclusions arise from their studies. First, relative income is identified as a significant predictor of subjective well-being. Kingdon and Knight use two definitions of relative income, namely the mean income of individuals who are the same race as the respondent; and the position of the respondent in the income distribution of the respondent's own race group.[6] This leads to the conclusion that reference groups are divided along racial lines, a conclusion which is not surprising given the country's history of racial segregation. Second, Kingdon and Knight (2007) also find that relative income enters individuals' utility functions positively for individuals who are in the same residential cluster ("close neighbours") and negatively for more far-off individuals ("more distant others").[7] This phenomenon is ascribed to feelings of altruism which exist within small geographic areas such as the neighbourhood or village (of which the residential cluster is a proxy), while feelings of envy exist towards others in the same district, who are seen as being socially more distant.

Since 1994 South Africa has been introduced back into the world economy and has experienced unprecedented economic growth and large-scale racial integration. However, with high and persistent levels of inequality and poverty (both of which have a lingering racial undertone) remaining part of the South African economic landscape (Leibbrandt, Woolard, Finn and Argent, 2010; Finn, Leibbrandt and Levinsohn, 2012), a relevant question at this stage is whether the new political dispensation has caused any shifts in the definition of the reference group used in determining relative well-being. In other words, do individuals still compare their income with others of the same race group? And are these comparisons correlated with higher levels of subjective well-being if the reference group lives in the same cluster? Also, if reference groups are no longer solely divided along racial lines, how much weight is placed on others from different race groups? The answers to these questions require some consideration. Although racial integration has taken place in many spheres of South Africans' everyday lives, interracial contact between individuals remains fragmented. In 2013, 41% of a representative sample of South Africans reported having no conversations with anyone who was not of their own race group (referred to as "interracial talk") on an average day, while only 31% reported taking part in interracial talk on a daily basis (Wale (2013), using the SA Reconciliation Barometer Survey).

The aim of this paper is to attempt to answer questions about the racial and geographic definition of reference groups within South Africa using data from the first wave of the National Income Dynamics Study (NIDS) collected in 2008, 14 years after the first democratic elections. The data offers a unique opportunity to reassess the findings from 1993 and to add to the small body of evidence regarding reference groups in South Africa.

[6]These definitions both assume a restrictive definition of the reference group; something which I will test in this study.

[7]This finding is confirmed by Bookwalter and Dalenberg (2010) for the non-white population.

After providing an overview of the relevant literature, an overview of the data used in this paper is provided. The empirical framework is explained next. In order to test the previous conclusions from the literature on the racial and geographic definition of reference groups in South Africa, I develop a more flexible way of estimating the relevant parameters. This is done by making use of a non-linear model which allows for the estimation of the weight placed on the relative standing of one's own race group compared with other race groups, while simultaneously estimating the weight placed on the geographic distance of others.

The results seem to broadly confirm the findings by Kingdon and Knight (2007) and Bookwalter and Dalenberg (2010) and indicate that households in closer proximity enter the individual's utility function positively while more far-off individuals enter the utility function negatively. More specifically, the spatial parameter is estimated as being positive for others in the same residential cluster, and negative for others living in the same district (although imprecisely estimated and therefore not statistically significant). However, the weight placed on national-level averages, as a proxy of all others in the country, are estimated to be negative and statistically significant. Some interpretations for these findings are explored.

The results also seem to broadly confirm the findings from Kingdon and Knight (2007) in terms of the racial delineation of reference groups. Allowing the parameters to be estimated non-linearly leads to estimates of the size of the own-race parameter that are between 70% and 90% of the total weight placed on others, depending on the specification. This suggests that although some racial integration has taken place, comparisons with own race weigh more than comparisons with other race groups (which, by way of definition, would be 10%-30% of the total weight). In fact, in none of the specifications can the hypothesis be rejected that the own-race parameter is not equal to 100% of the total weight, in other words where all of the weight is placed on individuals of the same race group. The paper concludes by examining the robustness of the main findings by introducing fixed effects as well as alternative income measures.

The contribution of the empirical work is two-fold. First, it revisits the previous studies on reference groups in South Africa in 1993 and updates these findings using data from 2008. However, the main contribution of the study is that it develops a new way of estimating the various parameters which comprise the definition of reference groups in the utility function. This methodological innovation allows for these parameters to be estimated in a more flexible way, which allows for the testing of hypotheses which have formed the basis of previous studies, and which have not been tested before.

2 Subjective well-being and reference groups

2.1 A general overview of the literature

Given the large body of research on the determinants of subjective well-being, certain stylised facts have emerged in the literature. In the first place, there is general consensus that individuals with higher income are on average more likely to report higher levels of subjective well-being. In addition, the causation has been shown to run from income to happiness (Frey and Stutzer, 2002, and more recently Pischke, 2011). However, this positive relationship is limited to cross-sectional, and not time-series data. In explaining this paradox, Easterlin (1995) comes to the conclusion that it is not only absolute, but also relative income that matters - while own income is positively correlated with reported well-being, the income of all relevant others is negatively correlated. This confirmed the relative income hypothesis first proposed by Duesenberry (1949) decades earlier.

A large literature has developed around identifying and characterising the relevant reference group with whom people compare themselves. Definitions of reference groups are mostly centred around individuals who have similar attributes or have frequent social interaction with each other. Examples include parents (McBride, 2001); colleagues (Clark and Oswald (1996) using British data, Clark and Senik (2010) using data on 18 European countries, Schneider (2010) using German data and Card, Mas, Moretti and Saez (2012) using US data); an individual who is similar in age, education and gender (McBride, 2001; Ferrer-i-Carbonell, 2005); neighbours (Fafchamps and Shilpi, 2008; Dittmann and Goebel, 2010); an undefined "representative person" in society (Easterlin, 1995); and the individual at a different time period in their life (Easterlin, 2001, 2006).[8]

Mostly, these studies have found that the income of the reference group enters the utility function negatively (Clark and Oswald, 1996; Clark and Senik, 2010; Schneider, 2010; Card, Mas, Moretti and Saez, 2012), seen as indicative of feelings of envy, rivalry, unfairness or relative deprivation. However, in specific settings there have also been findings that the relative standing enters the utility function positively. For example, Ferrer-i-Carbonell (2005) uses data from the German Socio-Economic Panel and finds that, for the West German sample, the results are as expected - the reference group income is negatively correlated to reported well-being. However, for the East German sample the negative effect of reference income is small (although the coefficient remains negative) and only statistically significant at the 10% level. These results are confirmed by Knies (2012), who finds a positive effect of reference income for the East German sample, although the

[8]Easterlin (2001, 2006) refer to this as life-cycle happiness and adaptation - the fact that individuals compare their present well-being to their well-being in the past.

coefficient is not statistically significant. Ferrer-i-Carbonell (2005) and Knies (2012) provide the explanation that it might be because the East German economy is more unstable, and therefore there is both a negative comparison effect (comparing oneself with better-off others and feeling worse about one's own situation) as well as a positive information effect (if others are doing well, then perhaps I will do better in the future as well). Senik (2004) finds evidence of this positive information effect for Russia, another economically unstable economy. Other reasons for the positive relationship between subjective well-being and relative income include feelings of altruism or, as will be discussed below, risk-sharing within a poor and vulnerable community.

Comparisons to reference groups are done in various domains, not only income. Frey and Stutzer (2002) and Stutzer and Frey (2010) discuss the literature on the negative impact of the general unemployment rate in a region (the country as a whole, or some smaller geographic area) on the subjective well-being of individuals. Various studies have also found that the negative impact of unemployment on subjective well-being is mitigated by high regional unemployment rates, providing evidence for reference-dependence (Grunow, 2014).

An independent though related body of research looking at the concept of social reference groups in a more structured way has also been developed by the Leyden school (including papers such as Kapteyn, Van Praag and Van Herwaarden, 1976; Van Praag, Kapteyn and Van Herwaarden, 1979; Kapteyn and Van Herwaarden, 1980; Van Praag, Frijters and i Carbonell, 2003). These authors define the concept of individuals' social reference spaces. In other words, the methodology assigns a weight to each individual within the reference group to take into account that certain individuals are more influential than others. Reference weights are estimated along a vector of social characteristics, including race, age, gender, geographic area, education, and employment status. Individuals with the same social characteristics vector are said to be of the same social type. The frequency of observing a specific social type is used along with the assumption that all individuals of the same social type will have the same social reference space, in order to estimate the weight attached to each individual in the reference space.

Apart from regression techniques, a few studies have utilised experiments to estimate the importance of relative well-being. One strand of research has made use of experiments to identify the importance of relative well-being (or "positional concerns") on overall well-being or utility, by asking participants to choose among a list of options, each containing information about the absolute and relative well-being of the participant in hypothetical future scenarios (Johansson-Stenman and Daruvala, 2002; Alpizar, Carlsson and Johansson-Stenman, 2005; Yamada and Sato, 2013). Participants clearly favour more equal societies, but choices indicate that relative standing within society remains important, and participant choices indicate a preference for a higher relative income. This holds true for income, but also consumption, especially for more visible items such as cars and houses (Alpizar, Carlsson and Johansson-Stenman, 2005). In a

randomised experiment, Card, Mas, Moretti and Saez (2012) show how disclosing information on peer's salaries has an impact on workers' job satisfaction, with workers earning below the median in their occupation experiencing a decrease in job satisfaction, and an increased probability of looking for alternative employment subsequent to the information treatment.

Another approach has been to make use of functional magnetic resonance imaging (Fließbach, Weber, Trautner, Dohmen, Sunde, Elger and Falk, 2007). By varying the size of payment for correct answers in an estimation game, Fließbach, Weber, Trautner, Dohmen, Sunde, Elger and Falk (2007) show how participants' brain activity varies as the payment for correct answers changes relative to their comparison participant. This provides evidence that participants make use of social comparison in evaluating their payment.

Although the studies mentioned above have all been within the context of OECD countries, much research has also been conducted on answering similar questions for emerging economies. Clark and Senik (2011) provide an overview of the literature on subjective well-being and relative measures of well-being in various countries in Latin America, Asia and Africa. The results from these studies seem to depend on the context of the country where they were conducted and no consensus has emerged.

Fafchamps and Shilpi (2008) use data from Nepal to ascertain whether relative consumption is as important to individuals in poor, isolated countries as it is in developed countries. They find that, similar to the findings from more developed countries, Nepalese households' subjective assessment of the adequacy of their consumption[9] rises with own consumption but decreases with the higher levels of mean consumption in the ward (a proxy for neighbourhood) where the individuals live. Fafchamps and Shilpi (2008) reject the hypothesis that relative consumption only matters for individuals in developed countries.

Using a questionnaire-based choice experimental method following the design by Johansson-Stenman and Daruvala (2002) and Alpizar, Carlsson and Johansson-Stenman (2005), Carlsson, Gupta and Johansson-Stenman (2008) asks Indian participants to choose between various fictitious future societies on behalf of an imaginary grandchild. Carlsson, Gupta and Johansson-Stenman (2008) find that there is a negative relative income effect overall, i.e. having a lower relative income in comparison to others in the same caste decreases subjective well-being. However, belonging to a caste with higher average income has a positive effect on subjective well-being. The findings suggest that the former effect dominates the latter, i.e. reported subjective well-being decreases as the relative income of the caste increases.

[9]In their paper, self-assessed adequacy of consumption is used instead of subjective well-being. The self-assessed adequacy questions ask household heads whether their household has been consuming adequate amounts of various consumer products. Fafchamps and Shilpi (2008) argue that this is a more accurate measure of utility than subjective well-being.

Ravallion and Lokshin (2010) examine the relative income hypothesis for individuals in Malawi. They find that, in contrast to the negative effect found in OECD countries, relative income of neighbours and friends (measured both objectively by taking the mean income of the enumeration area and subjectively through questions regarding perceived relative standing) enter the subjective well-being function positively for most of the sample. This positive impact may be as a result of risk-sharing agreements that would be more prevalent in poor rural areas where individuals have known their neighbours for many years.

In a related experiment, Akay, Martinsson and Medhin (2011) asks individuals in Ethiopia to choose between two fictitious villages to live in, where their own and others in the village receive different income packages, which are framed as aid packages. Akay, Martinsson and Medhin (2011) find a relatively low degree of positionality, indicating that individuals in their experiment cared less about relative income than what was generally found to be the case in richer countries. This leads Akay, Martinsson and Medhin (2011) to conclude that individuals in poorer countries care less about relative income than individuals in richer countries.

2.2 Subjective well-being and reference groups within the South African context

Within the South African context, quite a few studies have considered the determinants of subjective well-being generally. In earlier research, Møller and Saris (2001) examine the different domains that affect subjective well-being within each race group. Møller and Saris (2001) find that, while income is an important domain for the determination of subjective well-being for black and coloured individuals, white individuals and Asians are more influenced by other domains related to family and relationships. They also find that the determinants of subjective well-being are differentiated between the different provinces.

A similar conclusion is found in the research by Bookwalter and Dalenberg (2004) where they use the Southern African Labour and Development Research Unit (SALDRU) household survey administered in 1993 to examine the determinants of happiness for individuals in and out of poverty. They find that individuals below the poverty line view housing and transportation as the most important determinants of happiness, while those above the poverty line view sanitation, water, energy, education and health as more important. These results have important policy implications.

The structure of subjective well-being equations and the determinants of subjective well-being in South African data have been discussed in depth by Powdthavee (2003, 2005, 2006). In terms

of the structure of subjective well-being equations, Powdthavee (2003) finds correlates with subjective well-being that are similar to those found in developed countries. These include reported as well as relative income, household living conditions as well as individual-level characteristics, such as whether an individual is unemployed as well as the age and race of the individual. Within the South African context, whether the individual has been a victim of a crime is also significantly (negatively) correlated with reported subjective well-being (Powdthavee, 2005).

Kingdon and Knight (2006) also use the 1993 SALDRU household survey to examine the determinants of subjective well-being in South Africa prior to the end of apartheid, and argue that subjective well-being may be used as an alternative measure of poverty. They find that, although reported household income and subjective well-being are positively correlated, the effect of household income on the subjective well-being of the household is not very large. In addition, Kingdon and Knight (2006) find that absolute income seems to matter for individuals in households below the poverty line, while relative income matters for individuals in households above the poverty line. In their study, relative income is calculated using the household's race group as reference and generating race-specific income quintiles from the reported income data.

More recent studies have focussed on the changes in South Africans subjective well-being subsequent to the end of apartheid. In this regard, Møller (2007a,b) provides a detailed overview of the perceptions and attitudes of South Africans ten years subsequent to the 1994 democratic elections. She argues that, within a transitional economy such as South Africa (in which political liberation was introduced before economic reform), a large portion of the population were granted political rights without the necessary economic opportunities. According to Møller, this explains the increase in self-reported well-being among black individuals during the time of the 1994 elections, and the subsequent decrease as basic economic needs were not met, which may be interpreted as a reflection of the economic opportunities available to individuals. This decrease in hope and optimism (as evidenced by a decrease in subjective well-being) has also been ascribed to the increase in violent crime which affected thousands of South Africans in this post-apartheid period (Louw, 2007).

In terms of reference groups within the South African context, Kingdon and Knight (2007) specifically focus on the issue of reference groups within South Africa as a divided society. They find, again looking at 1993 SALDRU data, that although relative education and relative employment levels matter for subjective well-being, relative income is still the most significant determinant of subjective well-being. Relative income to other households in the same neighbourhood cluster is positively associated with subjective well-being, while relative income to more far-off others (i.e. other households in the district) is negatively associated with subjective well-being (Kingdon and Knight, 2007). Testing this hypothesis further, they come to the conclusion that the positive effect of others' income at the cluster level is altruistic. Kingdon and Knight (2007)

use two definitions of relative income, namely the mean reported income of individuals who are the same race as the respondent; and the position of the respondent in the reported income distribution of the respondent's own race group. Finding that these two definitions are both significantly correlated with subjective well-being leads to the conclusion that reference groups also have a racial element in that individuals are more likely to compare themselves to others of the same race group. These definitions both assume a restrictive definition of the reference group, an assumption which I will relax and then test later in this paper.

Bookwalter and Dalenberg (2010) confirm the findings of Kingdon and Knight (2007)[10] by finding that poorer households are more likely to perceive wealthier households as a positive, rather than a negative, impact on well-being. They also expand the definition of the reference group tested by Kingdon and Knight (2007) by also testing the significance of relative income in comparison to one's parents on subjective well-being. Bookwalter and Dalenberg (2010) find that relative income in comparison with one's parents has a large and significant impact on subjective well-being. Those individuals who perceived their own household to be wealthier than their parents were much more likely to report higher levels of subjective well-being.

One criticism against using reported mean income in the geographic area as the measure of relative standing comes from the work of Posel and Casale (2011), in which it was found that perceived relative income, rather than the position of the household in the income distribution based on reported income, influences the subjective well-being level of South Africans, specifically for the black sample. Posel and Casale (2011) show that this is mainly because of the large disconnect between where individuals actually are on the income distribution *versus* where they perceive themselves to be.

Although significant advances have been made in increasing the level of racial integration within South Africa subsequent to 1994, Du Toit and Kotzè (2011) point out the fact that post-apartheid affirmative action may have had the opposite effect, entrenching the racial divide brought about by apartheid legislation.[11] Some evidence of this break-down in society is found in the study by Posel and Hinks (2013) examining the levels of trust in South Africa. They find that South Africans have very low levels of reported trust compared with other countries, even when looking at trust among neighbours. Burns (2012) confirms this finding for high school children. She finds that in trust games played in high schools in South Africa, non-black students still display a greater measure of distrust towards black partners in the game. However, the results from the study do seem to indicate that students from more racially integrated schools are prone to more pro-social behaviour.

In recent work Kaus (2013) approaches the question of reference groups in South Africa from

[10] Using the same data as Kingdon and Knight (2007).

[11] The authors refer to the "re-racialisation of society" in South Africa (Du Toit and Kotzè, 2011, p. 85).

a different angle. Kaus, using Income and Expenditure Survey data from 1995, 2000 and 2005, examines the issue of conspicuous consumption within different race groups in South Africa. The hypothesis is that if conspicuous (visible) consumption is a form of status-seeking behaviour in line with Veblen's signalling model,[12] then the introduction of the relative standing of the individual compared to the reference group into the model should account for all differences in conspicuous spending between race groups. After finding large and significant differences in conspicuous consumption between white and black South Africans, Kaus sets out to test whether these differences can be explained by the signalling model. He does this by introducing a proxy of reference group income, using the mean provincial reported income of each race group as a rough proxy of reference groups. Introducing this proxy diminishes all differences in conspicuous spending in Kaus' models. Indirectly, Kaus' assumption that reference groups in South Africa remain divided along racial lines is therefore confirmed by the results.

However, Du Toit and Kotzè (2011) highlight the fact that recent data from the World Values Survey (2006) seem to signal an increased racial tolerance and inter-personal trust. Indications of racial integration seem to be borne out by the results of Hinks (2012) on the impact of fractionalisation within the South African context. Hinks (2012) finds that a higher level of ethnic and linguistic fractionalisation within the residential cluster is positively correlated with subjective well-being, which seems to indicate that there has been an increase in racial tolerance. Hinks (2012) even goes so far as describing these results as being "... *consistent with a nation that enjoys diversity*" (p. 261).

As indicated in the introduction, the question is therefore whether the new political dispensation had any effect on the way South Africans view their lives. In other words, did subjective well-being and the reference groups against which individuals compare themselves changed since 1993? The remainder of this paper is aimed at answering this question.

3 Methodology

In this study, I test two hypotheses. First, whether the influence of reference groups is heterogeneous across space. More specifically, I test the earlier finding of Kingdon and Knight (2006, 2007) in which it was found that in apartheid South Africa, neighbours (i.e. others that live in close proximity) have a positive impact on subjective well-being (which is ascribed to altruism towards others living in close proximity) whereas others who live further away enter the utility function negatively (reflecting feelings of envy). This has also been referred to by Fafchamps and

[12] In other words individuals purchase certain assets purely to indicate their relative standing in the income distribution of their reference group. Therefore, the higher up one is in the reference group income distribution, the more conspicuous one's consumption will be.

Shilpi (2008) as the convivial *versus* invidious village hypothesis. Second, I test whether earlier findings by Kingdon and Knight (2006, 2007) remain valid and test whether in post-apartheid South Africa reference groups are still racially delineated as they were found to be in 1993.

I conduct my analysis in two stages. I start by roughly replicating the approach followed by Kingdon and Knight (2007) to test whether these prior findings still hold. This approach involves the estimation of the utility function with the assumption that the parameters all enter the function linearly, and that the parameter which represents the weight placed on individuals of one's own race fully captures the relevant reference group; i.e. individuals from other race groups do not enter the utility function. This approach is restrictive in the sense that it does not allow the parameter capturing the weight placed on different race groups to vary, but rather assumes that all the weight is placed on others within the same race group, and none of the weight is placed on others from a different race group.

Therefore, to be able to get a real sense of whether the results from these previous studies still hold, it is necessary to relax these restrictions and to estimate a more flexible model, so that I am able to test both the size of the spatial parameters as well as the race parameter. For this reason, in the second stage of the analysis, I develop a more flexible model which allows the weighting parameters to enter non-linearly. With this model, I am able to estimate the weight placed on others within the same race group versus those of another race, while simultaneously estimating the spatial weights that individuals place on others in order to get to a better understanding of the correct definition of the appropriate reference group. I set out the two stages of my analysis below.

3.1 Testing spatial and racial variations in the reference group as per Kingdon and Knight (2007)

I start with the basic model used by Fafchamps and Shilpi (2008) in terms of which only own income, x_{ik}, and relative income, $\overline{x}_k = \frac{1}{n}\sum_{i=1}^{n} x_i$, which is just the mean income in the geographic area, are considered. The individual utility function U of individual i living in the geographic region k then takes the form:

$$U_{ik} = \alpha log(x_{ik}) + \beta log(\frac{x_{ik}}{\overline{x}_k}) + \gamma z_{ik}$$

Here utility is measured as the subjective well-being of the individual. It is modelled as a function of individual income, x_{ik} (defined as the monthly *per capita* household income); the income of the reference group, $\overline{x}_k$ (defined as the mean monthly *per capita* household income of others in

the geographic area k), as well as a set of individual and household controls, z_{ik}. Typically, this is then estimated as:

$$U_{ik} = \kappa log(x_{ik}) - \beta log(\bar{x}_k) + \gamma z_{ik} \tag{1}$$

where $\kappa = \alpha + \beta$. The expectation is that the sign associated with the coefficient on the mean income of all others in the relevant geographic area would be negative, while the sign on the own income coefficient would be positive, reflecting the way in which the two variables enter the utility function.

In order to test the two hypotheses, I start by estimating ordered probit regressions. I include two definitions of the geographic area in which the mean income is measured: the cluster and the district. The choice of these two geographic areas is motivated by the aim of testing whether the conclusions by Kingdon and Knight (2007) remain valid. In Kingdon and Knight (2007), the residential cluster is taken as a proxy of those living in close proximity, who might be referred to as neighbours. The district is seen as a proxy of all others living a further distance away, who might be seen as strangers. In order to remain consistent with the approach by Kingdon and Knight (2007) I repeat these ordered probit regressions, not only for relative income, but also for relative education and relative employment.

I then estimate Equation 1, but now I use the average mean income of all others of the same race group in the cluster and then the district as the relative income measure, thereby testing the hypothesis of whether the correct reference group in South Africa is still defined in terms of race.

The robustness of this finding is tested by making use of a different measure of the reference group by including the position of the individual's household on the race-specific income distribution. This repeats the approach of Kingdon and Knight (2007), however the results can of course not be directly comparable to the results where the mean same-race income is used as a proxy of relative income. The main reason for this is that the use of the position in the race-specific income distribution makes the additional assumption that the individuals are aware of their position in relation to others of the same race, something which has been shown to rarely be true in South Africa, as discussed earlier (Posel and Casale, 2011). On the other hand, the use of mean income of all others in the same race group in the cluster and district acts merely as a proxy of the well-being of all others in the geographic area and does not assume any knowledge from the individual about where he or she ranks in the income distribution.

3.2 Testing spatial and racial variations in the reference group taking a more flexible approach

As indicated above, the approach above imposes some restrictions on how the parameters enter the utility function. In order to relax these assumptions, I make use of the concept of social reference spaces developed by the Leyden school (Kapteyn, Van Praag and Van Herwaarden, 1976), in the sense that I make provision for the inclusion of weights in the utility function, in order to estimate what preference is given to individuals of the same race *vis-à-vis* others of a different race in defining the relevant reference group with whom an individual compares themselves in evaluating their subjective well-being or utility.

For this purpose I define a race parameter, μ, representing the weight placed on an individual's own race group in the utility function. I define the weight placed on the other race groups in the individual's utility function as $1 - \mu$, in order to capture the idea of a proportion - i.e. the sum of the two weights should be equal to one. [13]

This allows Equation 1 to be re-written as:

$$U_{ik} = \kappa log(x_{ik}) - \beta log\left(\mu \bar{x}_{1k} + (1 - \mu)\bar{x}_{2k}\right) + \gamma z_{ik} \tag{2}$$

where $\bar{x}_{1k}$ is the mean income of all others in the geographic area k who are the same race as individual i and $\bar{x}_{2k}$ is the mean income of all others in the geographic area k who are a different race as individual i.

The implicit assumption made by Kingdon and Knight (2007) is that $\mu = 1$, in other words all of the weight is placed on individuals of the same race, and none on individuals of other race groups. In other words, Equation 2 above just simplifies back to:

$$U_{ik} = \kappa log(x_{ik}) - \beta log(\bar{x}_{1k}) + \gamma z_{ik}$$

with the only relevant comparison income being that of others in the same race group in area k.

In order to combine this race parameter with the hypothesis about spatial variation, I expand Equation 2 into a non-linear model which is able to test the two hypotheses described above;

[13] However, in the estimations in the next section, the race parameter μ is not constrained to be between zero and one, which essentially allows it to take any value.

i.e. whether there is spatial variation in the reference groups as well as estimating the weight or preference placed on own *versus* other race.

$$\begin{aligned} U_{icdn} &= \lambda_y log(x_{icdn}) + \lambda_y \theta_c log(\mu \bar{x}_{1cdn} + (1-\mu)\bar{x}_{2cdn}) \\ &\quad -\lambda_y \theta_d log(\mu \bar{x}_{1dn} + (1-\mu)\bar{x}_{2dn}) - \lambda_y \theta_n log(\mu \bar{x}_{1n} + (1-\mu)\bar{x}_{2n}) + \gamma z_{icdn} \end{aligned} \tag{3}$$

To be clear, Equation 3 sets out the utility function of individual i, living in residential cluster c, in district d, in nation n. The utility function includes essentially four sets of parameters or weights:

1. First, a set of preference parameters capturing the weight placed on or preference given to income λ_y in the utility function.

2. Second, a set of spatial parameters, capturing the weight or preference placed on the geographical reference group, is included. These are represented as θ's. There are 3 spatial parameters, namely θ_c (the weight placed on others living in the same residential cluster), θ_d (the weight placed on others living in the same district), and θ_n (the weight placed on all other South Africans).

   Following the conclusions of Kingdon and Knight (2007), I expect the signs of the spatial parameter associated with the cluster, θ_c, to be positive, as it captures the positive influence of the well-being of others in the same neighbourhood or village (the convivial village hypothesis). However, the signs of the spatial parameters associated with the district, θ_d and nation, θ_n, are expected to be negative in line with the invidious village hypothesis (i.e. others living further away invoke envy and have a negative impact on the individual's subjective well-being).

3. Third, the race parameter, μ, which captures the weight placed on the race group of the individual, with $1-\mu$ capturing the weight or preference placed on other race groups within the specific geographic area (cluster, district or nationally). The race parameter μ enters the utility function through the mean income variables for each of the geographic layers. In other words, μ is used to weight the mean income of the individual's own race group in the relevant geographic area, whereas $1-\mu$ is used to weight the mean income of all the other race groups in the relevant geographic area.

   An intuitive way of thinking about the interpretation of the race parameter μ is by considering the case where there is no preference given to any other individual on the basis of their race. In other words, where individual i is agnostic about the preference or weight placed on every individual in the cluster. In this instance, the size of μ would merely reflect

the population weight attributed to a specific race group, and the second term in Equation 3 could be re-written as:

$$\begin{aligned}
\lambda_y\theta_c log\left(\mu\bar{x}_{1cdn} + (1-\mu)\bar{x}_{2cdn}\right) &= \lambda_y\theta_c log\left(\mu(\frac{1}{n_1}\sum_{i=1}^{n_1} x_{icdn}) + (1-\mu)(\frac{1}{n_2}\sum_{i=1}^{n_2} x_{icdn})\right) \\
&= \lambda_y\theta_c log\left(\frac{n_1}{n_1+n_2}(\frac{1}{n_1}\sum_{i=1}^{n_1} x_{icdn}) + \frac{n_2}{n_1+n_2}(\frac{1}{n_2}\sum_{i=1}^{n_2} x_{icdn})\right) \\
&= \lambda_y\theta_c log(\overline{x}_{cdn})
\end{aligned}$$

In other words, it would simplify back to the simplest case where the mean income of all individuals in the cluster, irrespective of their race, is taken as the correct measure of relative income.

4. Last, γ captures the weight placed on the individual and household controls included in the estimations as a way in which to ensure that other influences on the individual's subjective well-being has been taken into account.

4 Data

The data used in this analysis are from the first wave of NIDS. NIDS is conducted by SALDRU at the University of Cape Town. The survey, which was conducted by SALDRU and based on the initial 1993 SALDRU survey, was completed during 2008.[14] It incorporates data from just over 7 000 households, containing approximately 28 000 household members as well as data on approximately 18 600 "adults", defined as all individuals aged 15 years and older.[15]

The level of subjective well-being of all adults in the NIDS dataset is recorded by the inclusion of a variable measuring, on a scale from 1 to 10, the level of satisfaction with life experienced by each adult (with 1 signalling extreme dissatisfaction and 10 signalling extreme satisfaction). This differs from the SALDRU data discussed above where the question was posed to ascertain

[14] Although the second and third waves of NIDS are available, for this study it was decided to focus only on the first wave. The primary reason for this decision is the fact that there was such high levels of attrition among the white population. More specifically, attrition within the white population was 50.3% over the three waves, and was mostly as a result of refusal to complete a questionnaire. On the other hand, attrition in the black population group was much lower at 13.4%, mostly attributable to loss of contact. (De Villiers, Brown, Woolard, Daniels and Leibbrandt, 2013, p. 21-22)

[15] It should be noted that the sample worked with here is limited to individuals who were included in the adult questionnaire, which was aimed at individuals aged 15 years and older. However, as a result of inaccurate birth dates, 50 individuals aged 14 years were accidentally included in the adult questionnaire and accordingly also in the sample used in the current study.

the household's subjective well-being (Posel, 2012, 2014). Approximately 74% of adults who completed a questionnaire provided a response to this question.[16]

The distribution of the level of subjective well-being in the data is graphically depicted in Figure 1, while the summary statistics of the variable are broken down by race in Table 1 (all figures and tables are to be found in the appendix to the paper). Although the mean level of subjective well-being for the entire sample is approximately 5.5, marked differences in the subjective well-being between black and white individuals are observed in the data. While the mean subjective well-being for the black population in the sample is just above 5 (with a standard deviation of 2.5), the mean for the white sample is much higher at a subjective well-being level of almost 7 (with a standard deviation of 1.8). This is in line with the differences in mean per capita income which is observed between the two race groups: the average *per capita* household income for white individuals in the data was approximately R6 448, while the average *per capita* household income for black individuals was approximately R878 (both in 2008 Rands). In addition, the distribution of subjective well-being for the white sample is much more skewed, indicating the higher levels of subjective well-being generally observed amongst white respondents (Hinks, 2012; Posel and Casale, 2011).[17]

For the current analysis, there are essentially four sets of variables used in the empirical analysis. Each of these will be discussed in turn below to provide an overview of the variables used in the empirical analysis.

In the first place, the analysis makes use of geographic variables. The NIDS data include 400 residential clusters that are all in the same district and urban or rural area. This is the smallest geographic unit of analysis within NIDS. These clusters together comprise the district councils or district municipalities, of which there are 53 in South Africa.[18] Within the district councils, households from different geographical areas are included. Since districts include a larger geographic area than the residential clusters, the district is therefore seen as a proxy for more distant others, while the cluster is seen as a proxy for closer others.

Second, the analysis relies on various individual and household-level controls which are included in the utility function. These include mean household monthly *per capita* income;[19] household

[16]This calculation and all subsequent analyses exclude all adults for whom only a proxy questionnaire was completed.

[17]Although the significance of this difference disappears in some specifications, as indicated in the results set out in the next section.

[18]In this paper, I refer to districts or district councils interchangeably. To be clear, I make use of the demarcation of the district and provincial boundaries at the time of the 2001 census. Although some changes have been made since 2001 in terms of the official district and provincial boundaries, these changes were small and should not have any impact on the results reported here. I have, however, also tested the robustness of my results using the boundaries of the 2011 census and there is no significant difference in the results.

[19]It should be noted that the income variable used includes the variable generated by SALDRU which includes

size; the mean education of the household for all individuals aged 18 or older; and the proportion of individuals aged between 16 and 64 (both inclusive) in the household who are employed. This employment variable is more preferable than the unemployment rate of the household, since a large proportion of households in the sample do not include any individuals who are economically active.[20] In calculating the mean unemployment rate, it is therefore not clear to know how to treat these cases. Setting the unemployment rate in these households equal to zero distorts the impact of the variable in the regressions. It was therefore decided to rather include the variable as an employment rate rather than an unemployment rate.

The individual-level controls included in the utility function are as follows: race; age in years; years of education; whether male or female; employment status; and marital status. These have all been shown to influence reported well-being in South Africa (see, for example, Powdthavee, 2003; Hinks and Gruen, 2007).

Differences in the subjective well-being of the four main race groups in South Africa should be seen in the light of the large differences in these covariates described above. Table 2 contains the descriptive statistics of variables describing the living conditions of individuals in the estimation sample, by race. It is clear that white individuals reside in smaller, wealthier households with more educated and employed adults than any of the other races, on average. As expected, individuals in the black population group remain more likely to be in poverty, live in households where individuals are less educated and are more likely to not be employed, on average. At the level of the individual, the descriptive statistics in Table 2 provides more evidence of the inequality that still remains between the four race groups, with much larger proportions of black individuals in the sample being discouraged or unemployed, and with a lower level of education, than white individuals, on average. Interestingly, in the sample we also see differences in the age of the individuals by race group (white individuals included in the sample are older on average), and in the marital patterns (black individuals are far less likely to be married than any of the other race groups), a phenomenon which has been researched by Posel, Rudwick and Casale (2011).

Third, the analysis also includes variables capturing the mean income, mean education and mean employment of individuals living in the same cluster and district who are from the same race group as the respondent or who are from a different race group. These variables are essentially weighted averages of the reported income (derived at the level of the household as described above) for various geographic levels - the residential cluster and the district - as the weighted mean of the respondent's own race group and as a weighted mean of all the other race groups

all income brackets and a limited amount of imputations on the level of the household, but does not include a full set of imputations at the individual level.

[20]This makes sense in the context of the findings of Leibbrandt, Woolard, Finn and Argent (2010), i.e. that approximately 30% of households reported that their main source of income was social grants.

residing in the geographic area. For individuals living in areas with complete racial homogeneity, the "other race group" variables take the value of zero. Table 3 sets out how the "other race group" variables differ per geographic level per race. From the table, it is clear that black individuals, when residing in neighbourhoods and villages (captured by the residential cluster) where there is racial heterogeneity (the likelihood of this occurring is discussed below), generally reside in poorer areas. This is in stark contrast to the other three race groups, but especially the white sample, who typically reside in more affluent neighbourhoods. Table 3 generally paints a picture of pockets of affluence and poverty which are highly correlated with the race of the households living there. This economic division raises the question whether there would be any racial integration in terms of the reference groups that individuals use to compare themselves to.

Finally, the fourth set of variables are included in order to control for the racial concentration in the various regions in South Africa. More specifically, since the utility function specified in Equation 3 above makes use of the division between own race and other race groups, it is necessary to control for the fact that the different geographic regions used in the analysis are not all equal in terms of racial concentration. In order to capture the extent of racial concentration a Herfindahl-Hirschman index (HHI) is calculated, in line with Hinks (2012), as follows: The HHI in area (cluster, district or province) j is calculated as follows: $H_j = 1 - \sum_{i=1}^{N=4} s_{ij}$ where s_{ij} is the proportion of race i residing in area j. More racially fractionalised or heterogeneous clusters or districts would therefore have a higher HHI.[21]

The distribution of the HHI for each race group in the cluster, district and province is reported in Table 4. Here it is clear to see that approximately 82% of all black individuals included in the NIDS data reside in residential clusters where there are no other race groups, i.e. neighbourhoods or villages of complete racial homogeneity. In order to control for this phenomenon, I include a control variable capturing the degree of racial concentration in the residential cluster.

5 Empirical analysis

As discussed above, I conduct the empirical analysis by first implementing the approach followed by Kingdon and Knight (2006, 2007) and thereafter relaxing some of the restrictions of their approach so as to estimate the parameters in a more flexible way.

[21] According to this formulation, a HHI of zero is equal to having only a single race group in the geographic area, whereas if the four race groups were distributed equally, the HHI within the geographic area would be 0.75.

5.1 Spatial reference groups

The first hypothesis I wish to test is whether there are variations in how reference groups who live various distances from the individual enters the utility function, in line with what Kingdon and Knight (2007) have found. For this purpose, I estimate Equation 1 separately for each of income, education, and employment at both the cluster and district level. I include all individual and household controls described in Table 2 as well as provincial fixed effects. The results from these ordered probit regressions are reported in Table 5.

The regression output seems to provide evidence of a convivial village or neighbourhood, since the coefficient on the average well-being at the level of the residential cluster enters positively in each of the regressions (although not statistically significant in the case of the employment parameter). This confirms the findings by Kingdon and Knight (2007) where the positive coefficient was ascribed to the fact that the ties people have with their neighbours (or others in close proximity) mimic the ties with family members and that this "extended family" effect encourages feelings of altruism towards these close others rather than envy.

However, again in line with Kingdon and Knight (2007), the positive coefficients make way for negative coefficients on the relative income, education and employment of the district (although the coefficient is not significant in any of the specifications except when looking at relative education). This seems to provide evidence of a invidious district effect - i.e. the feelings of altruism that existed towards others in the same neighbourhood or village have been replaced by envy towards peers in further away areas such as the district.

5.2 Racial reference groups

Next, I test the finding of Kingdon and Knight (2007), namely that preference is given to the race group of the individual and less weight is placed on other race groups in the utility function when considering relative income. I follow their approach and make use of three specifications. The results from these estimations are reported in Table 6. All three specifications include individual and household controls as set out in Table 2 as well as provincial fixed effects. I estimate the results using an ordered probit model and report these coefficients.

In the first place, I create a variable which captures the position of the individual's household in the race-specific income distribution. More specifically, I specify the race-specific income quintile. Column (10) sets out the results from this specification. Unlike the results reported by Kingdon and Knight (2007) for 1993, none of these race-specific income variables are significant.

However, I also use the log of the mean income of all others residing in the same residential cluster who are the same race as the respondent to estimate the importance of same-race reference groups in the cluster. The results from this estimation are set out in Column (11). The coefficient on mean income of others of the same race is positive and significant in the regression, indicating that the income of others of the same race do seem to make a difference to the subjective well-being of individuals in a way which is consistent with the convivial village hypothesis set out above. However, although the sign of the coefficient on the mean income of others of the same race residing in the same district is negative, it is not significant.

Given the theoretical framework set out in Section 3, it is not clear exactly how these results should be interpreted. What is captured by the coefficients on the mean income of others of the same race (within the cluster and district) is a combination of the following three weights, which all enter the utility function: the weight placed on income, the weight placed on others within the same race group, and the weight placed on others within the cluster or district. It is not possible to separate the size of these three weights without allowing for a more flexible model, as discussed in Section 3.

As a first step, I allow μ to not be equal to one, by introducing variables which capture the mean income of others who are from different race groups alongside the mean income of others within the same race group, for the cluster, district and nationally. These results are reported in Table 7. Whereas the regression output in Table 6 assumed that $\mu = 1$, the approach in Table 7 allow μ to not be equal to one, however the approach does not allow for a separation of the weight associated with geographic distance from the weight associated with income or the weight associated with race. However, the results do provide some indication that both individuals within the race group as well as individuals from other race groups belong in the utility function.

5.3 Non-linear estimates

Next, I allow the parameters on income, geographic location and race to enter the utility function non-linearly.

To this end, I estimate Equation 3. Since the preference parameters enter the utility function in a non-linear way, I make use of a non-linear least squares estimation. This estimation technique differs from the ordered probit or logit model that is generally used to model subjective well-being. Ordered probit or logit regressions allow an unobserved latent variable with unevenly spaced cut-points, in line with an ordinal utility function, whereas least squares assumes a cardinal utility function with evenly spaced categories. However, as indicated by Ferrer-i-Carbonell

and Frijters (2004); Senik (2004) and Geishecker and Riedl (2012), least squares estimation remains useful in research regarding subjective well-being as an alternative to ordered probit or logit models where individual fixed effects are included, since including individual fixed effects to control for individual heterogeneity leads to biased ordered probit or logit estimates. In general, these authors have found their findings from least squares estimations to be very similar to the results where more flexible techniques were employed (for further detail, the reader is referred to Geishecker and Riedl, 2012). In addition, since I will not be using the fitted values from the regression, the objection of out-of-bound (i.e. below 1 or above 10) predictions is not a big concern (Angrist and Pischke, 2009).

The results from the non-linear regressions are set out in Table 8. Columns (16), (17) and (18) report the results from three different specifications. In the first specification in Column (16), only race is included as an additional control along with the parameters of interest. In Column (17), all of the the individual controls set out in Table 2 are included, whereas Column (18) report results from a regression which also includes the also includes the household-level controls described in Table 2. In the first specification, μ is estimated to be approximately one, indicating that individuals place all of the weight in making comparisons on their own race group, and almost nothing on other race groups. However with the inclusion of the other individual controls, the race parameter is estimated to be approximately 0.8. These point estimates seem to provide evidence that although most of the weight is still placed on individuals of the same race group, there is some comparison to others who are from the other race groups. However, importantly, none of the estimated race parameters are statistically different from 1 and I therefore cannot reject the conclusions made by Kingdon and Knight (2006, 2007); i.e. that the reference group for South Africans comprises only individuals from the same race group.

Another result from Table 8 is the fact that the weight placed on income is large and significant. As far as the spatial parameters are concerned, the signs of the parameters seem to confirm the findings from the ordered probit regressions - i.e. that closer others enter as positives. However, as with the ordered probit results, it is not clear what the influence of the district parameter is, since it is imprecicesly estimated and therefore statistically insignificant, although the sign of the coefficient is negative, in line with expectations. The national parameter, θ_n, is estimated to be negative and statistically significant. This seems to provide clearer evidence of the hypothesis that others living further away enter the utility function as negatives.

In order to try and ascertain the reason why the cluster parameter is positive, I test two hypotheses. First, the reason for the positive coefficient may be that the cluster-level income is a proxy for the improved service delivery which exists in richer neighbourhoods. In other words, one of the benefits of living in a wealthier neighbourhood is the fact that neighbours are more concerned about the quality of the service delivery in the neighbourhood and would make sure

that basic amenities are in working condition. To test this hypothesis, I include three variables as proxies for service delivery, namely access to piped water; access to working electricity and access to a flush toilet. The results are set out in Table 9. Although the inclusion of these variables diminishes the size of θ_c, it remains positive and significant.

Next, I test whether the hypothesis of Kingdon and Knight (2007), namely that the positive sign of the cluster parameter is as a result of altruism existing among neighbours in the cluster. For this I re-estimate the main model for two sub-samples: the sub-sample of smaller clusters where there are less than 100 sampled individuals residing in the cluster and a separate regression for all of the larger samples where there are 100 or more sampled individuals residing in the cluster. These are reported in Table 9. If there are signs of altruism in the residential cluster, θ_c should be more significant for smaller clusters where individuals know each other better and the social distance between individuals is small. This is exactly what is seen from these regressions, with the estimated cluster parameter being positive and significant for the sub-sample of smaller clusters, but not for the larger clusters. This may be because the sample of larger clusters is too small to obtain an accurate estimate of θ_c. I therefore vary the cut-off for small and large clusters to be 80 and 90 individuals, and for both of these definitions, the same pattern remains. It therefore seems to indicate that there is evidence of altruism within the residential cluster.

I also examine variation in the size of the parameters. The most likely source of heterogeneity in these parameters would potentially be race. Indeed, previous studies have argued for the estimation of separate subjective well-being equations for the white and black (or non-white) population, given the large differences in die reported well-being between the race groups (Posel, 2012; Bookwalter and Dalenberg, 2010). In addition, one would expect the size of μ for the black population group to be different to that of the other population group, since (as has been shown in Table 4) black individuals in the sample are much more likely to live in residential clusters and districts where there is complete racial homogeneity. More precisely, 82% of black individuals in the sample live in these racially homogeneous residential clusters, while almost 19% of black individuals live in districts with complete racial homogeneity.[22] Generally, one would expect this to influence the weight that these individuals would attach to the relative well-being of individuals from other race groups.

I therefore start by estimating the preference parameters using the non-linear model for two separate samples. First, I use only the black individuals in the sample. Then, I combine the sample of non-black individuals, as these three groups are all much more likely to be living in areas where there is racial heterogeneity. The findings are set out in Table 10. The results seem to indicate that the black sample places a greater weight on individuals in their own racial group

[22]As is clear from Table 4, the other race groups are living in areas that are much more racially integrated, and none of the other race groups live in completely homogenous districts.

than on individuals of another race, with an estimated size of 0.7 for μ for the black sample, whereas μ is estimated to be 1 for the non-black population. The results from the black sample also suggest a larger weight on income,[23] in comparison to the non-black sample.

These findings seem at first to be in contradiction to what was found by Posel and Casale (2011), i.e. that perceived relative standing is more important to the black sample than to the white sample. Posel and Casale (2011) suggest that this might be because white individuals are more likely to have access to information and would not rely solely on their own perceptions of the well-being of others in their immediate environment in order to make conclusions on their relative position in the income distribution. In contrast, many black individuals are in poverty, have less years of education on average, more likely to live in rural areas and would accordingly be less likely to have access to accurate information regarding their relative standing. In other words, these black individuals are more likely to compare themselves to others of the same race, who live in close proximity. However, the results could also just be indicative of the fact that individuals always aspire to have more and therefore compare themselves to those who are on the same economic rank or on a higher rank than they are. This would explain why for black individuals this includes their own race as well as others of other race groups, while for the other race groups who are wealthier, this comparison is centred more around their own race group.

At this point it should be acknowledged that there has been criticism against estimating the impact of different spatial reference groups as was done above. Posel and Casale (2011) have shown, specifically with reference to the NIDS data, that perceived relative standing is a more important correlate of subjective well-being than relative standing based on reported income. However, given the framework used in this study, it would not be possible to use perceived relative standing for all of the hypotheses (most importantly, it would not have been possible to include perceived relative standing in the model used to estimate the race parameter or any of the parameters not related to income). This, along with the fact that it has been shown that relative standing based on repoted income remains an important input into the utility function, even after controlling for perceived relative standing (Posel and Casale, 2011), are the reasons why reported relative income is used in this study and not perceived relative income.

6 Alternative income measures and specifications

This section tests the robustness of the findings in the previous section in two ways. First by introducing geographic fixed effects and then by making use of three variations in the definition of income used in the utility function.

[23] In splitting the sample, Posel and Casale (2011) also found larger coefficients on the own income variable for the black sub-sample than for the white sub-sample.

First, I extend the main specification set out in Equation 3 by introducing fixed effects at the level of the province, as well as at the level of the district council. The results from the fixed effects regressions are reported in Table 11. The introduction of the provincial-level fixed effects in the first column does not have any influence on the size or significance of the race parameter, while the estimation of the weight placed on others in South Africa remains negative (and somewhat larger than estimated in the baseline regressions). The cluster parameter also remains positive and significant. When the district controls are included, as reported in the second column of Table 11, the size of μ is increased from approximately 0.8 to 0.9. In addition, only the cluster parameter remains significant (and positive), as was the case with the introduction of the provincial fixed effects. The estimate of the weight placed on others within South Africa is diminished and the coefficient becomes insignificant. The increase in the estimated size of μ is not substantial and neither of the estimates of μ in Table 11 has a confidence interval which excludes one. It is therefore not possible to reject the hypothesis that μ is not equal to one. These results seem to confirm the main conclusion; i.e. that individuals place the bulk of the weight on others of their race group rather than those who are of a different race group in the construction of the reference group in their utility function. The relative standing enters the utility function positively for everyone residing in the same cluster and negatively for others who live further away - i.e. in the rest of South Africa.

The second set of robustness checks I conduct is to vary the way in which income enters the utility function. In the main specification, I have followed the convention of previous studies and have included relative income by taking the logarithm of the mean of the income in the relevant geographic area. However, this approach is susceptible to the influence of outliers. This is especially true in this particular study since I have only made use of the data available in NIDS in calculating the mean income of the various geographic areas.[24] For small geographic areas such as the residential cluster, the inclusion of outliers will almost definitely influence the size of the mean income. This could influence the size of the cluster parameter as well as the parameter associated with income. In order to ascertain whether this is the case, I propose two different ways in which income may be specified in the utility function, both of which have been suggested by Deaton (1997: 121).

In the first place, I amend Equation 3 so that income is specified as the mean of the logarithm of individual incomes, in other words, where utility of individual i in cluster c, district d and country n was previously defined as in Equation 3, now it is defined as

[24]The GIS data in NIDS which would allow one to obtain census data for the geographic means became available too late to be included in this study. However, it is definitely an avenue worth exploring in the future.

$$
\begin{aligned}
U_{icdn} &= \lambda_y log(x_{icdn}) + \lambda_y \theta_c(\mu \bar{x}_{1cdn} + (1-\mu)\bar{x}_{2cdn}) \\
&\quad -\lambda_y \theta_d(\mu \bar{x}_{1dn} + (1-\mu)\bar{x}_{2dn}) - \lambda_y \theta_n(\mu \bar{x}_{1n} + (1-\mu)\bar{x}_{2n}) + \gamma z_{icdn} \qquad (4)
\end{aligned}
$$

where $\bar{x}_{1cdn} = \frac{1}{n_1} \sum_{i=1}^{n_1} log(x_{icdn})$.

The results from the specification using the mean of the logarithms of individual incomes is set out in Table 12. The income and cluster parameters do not change substantially from the main specification, seeming to validate the conclusion from the main results that the cluster income enters the utility function positively. The race and South African parameters are estimated to be somewhat larger as in the main specification. This does not however invalidate the main conclusion that the size of μ is no longer equal to one, and that some weight is also placed on other race groups. The negative sign of θ_n remains and is somewhat larger using the mean of the log income. This again confirms the result from the main specification that others who live further away, and can be seen as strangers rather than neighbours, enter the utility function as a negative.

The second alteration to the definition of the income variable again makes use of the specification in Equation 4, however instead of mean income, the median is used. The results from this regression is reported in the second column in Table 12. The income parameter λ_y and the cluster parameter θ_c again do not significantly differ in size or significance from the main regression, seeming to again confirm the conclusion that the income in the cluster enters the utility function positively. However, μ is estimated to be slightly larger, at 0.92. The parameter associated with national income, θ_n is however statistically insignificant (and positive) in this specification.

One final robustness check which I conduct involves amending the composition of the income variable in order to remove the potential of double counting individuals in the utility function. More specifically, up to this point I have defined the own race cluster, district and national income averages to include individuals to whom these averages pertain. In the final robustness check, I amend this by creating the own race average cluster income variable by excluding own income; I create the own race average district income variable by excluding the average income of the individual's own cluster and I create the own race average national income by excluding the average income of the individual's district. The results from this regression are reported in the third column of Table 12. The estimated size of μ is slightly diminished, but confirms the main conclusion, as it is not possible to reject the hypothesis that it is equal to one. The size and significance of the income and cluster parameters remain, however the national parameter becomes insignificant.

7 Conclusion

This aim of this work was to update and expand the findings from previous authors, most notably Kingdon and Knight (2006, 2007) and Bookwalter and Dalenberg (2010) regarding the definition of reference groups for the purposes of subjective well-being or utility functions in the South African context. Being a country which remains highly unequal, and where this inequality has a lingering racial undertone, South Africa makes for an interesting case study of the formation of reference groups. Since 1994, much has been done to improve this inequality and the racial segregation which remains part of the South African landscape.

Using data from the NIDS survey in 2008, the study aims to investigate spatial and racial variation in the determination of the relevant reference group. It revisits previous findings by Bookwalter and Dalenberg (2010) and Kingdon and Knight (2007, 2006) for South Africa in 1993. First, the empirical methodology implemented in these previous studies is repeated in order to confirm the validity of the results on the 2008 data. However, in addition to this, I also expand these previous findings by relaxing the assumptions used in the estimation of the results by Kingdon and Knight (2007). More specifically, while Kingdon and Knight (2007) made use of a restrictive model in which the spatial parameters entered the utility function linearly, and the race parameter was not allowed to vary between different races, I relax these restrictions. For this purpose, I make use of a more flexible non-linear model in which the various spatial parameters are allowed to enter on-linearly and interact with the race and preference parameters. In addition, the race parameter is allowed to vary between zero and one, and may be interpreted as the weight placed on individuals of the same race group in the determination of the reference group.

The results for various specifications of the more flexible non-linear model suggest that for the sample, the relative well-being of others in the same residential cluster (a proxy for neighbours) enter the utility function positively whereas the relative well-being of others who live further away, more specifically in the rest of the country, enter the utility function negatively. This seems to confirm the hypothesis that while individuals are altruistic towards their neighbours or individuals living in the close vicinity, they view those living further away with envy. Similar conclusions by Fafchamps and Shilpi (2008) label the first as the convivial village hypothesis whereas the second is referred to as the invidious village hypothesis.

Furthermore, as far as the race parameter is concerned, the results seem to indicate that at least some racial integration has taken place in the 14 years subsequent to the end of apartheid, with reference groups shifting from being solely based on own race, to also include others from different races. However, the size of the own-race parameter remains larger than the weight placed on

other race groups, and is estimated to be between 0.7-0.9,[25] depending on the specification. This suggests that although some racial integration has taken place, comparisons with own race weigh more than comparisons with other race groups.

The robustness of these results is tested by introducing provincial and district fixed effects. In addition, three alternative definitions of income are used to test for the effect of outliers. These robustness checks seem to confirm the main results.

[25] In none of the specifications could I reject the hypothesis that the race parameter is equal to one.

References

Akay, A., P. Martinsson and H. Medhin (2011): "Does Positional Concern Matter in Poor Societies? Evidence from a Survey Experiment in Rural Ethiopia," *World Development*, 40(2), 428–435.

Alpizar, F., F. Carlsson and O. Johansson-Stenman (2005): "How much do we care about absolute versus relative income and concumption?," *Journal of Economic Behavior and Organization*, 56, 405–421.

Angrist, J. and J. Pischke (2009): *Mostly Harmless Econometrics*. Princeton University Press, Pinceton.

Bookwalter, J. and D. Dalenberg (2004): "Subjective well-being and household factors in South Africa," *Social Indicators Research*, 65(3), 333–353.

——— (2010): "Relative to What or Whom? The Importance of Norms and Relative Standing to Well-Being in South Africa," *World Development*, 38(3), 345–355.

Burns, J. (2012): "Race, diversity and pro-social behaviour in a segmented society," *Journal of Economic Behavior and Organization*, 81, 366–378.

Card, D., A. Mas, E. Moretti and E. Saez (2012): "Inequality at Work: The Effect of Peer Salaries on Job Satisfaction," *American Economic Review*, 102(6), 2981–3003.

Carlsson, F., G. Gupta and O. Johansson-Stenman (2008): "Keeping up with the Vaishyas? Caste and relative standing in India," *Oxford Economic Papers*, 61, 52–73.

Clark, A. and C. Senik (2010): "Who Compares to Whom? The Anatomy of Income Comparisons in Europe," *The Economic Journal*, 120(May), 573–594.

——— (2011): "Will GDP Growth Increase Happiness in Developing Countries?," *IZA Discussion paper 5595*.

Clark, A. E. and A. Oswald (1996): "Satisfaction and comparison income," *Journal of Public Economics*, 61, 359–381.

De Villiers, L., M. Brown, I. Woolard, R. Daniels and M. Leibbrandt (2013): "NIDS Income Dynamics Study Wave 3 User Manual," Available at: http://www.nids.uct.ac.za/documents/wave-3-documents-and-questionnaires/200-nids-wave-3-user-manual-updated/file.

Deaton, A. (1997): *The Analysis of Household SSurvey: A mIcroeconometric Approach to Development Policy*. John Hopkins University Press.

Dittmann, J. and J. Goebel (2010): "Your House, Your Car, Your Education: The Socioeconomic Situation of the Neighbourhood and its Impact on Life Satisfaction in Germany," *Social Indicators Research*, 96, 497–513.

Dolan, P., T. Peasgood and M. White (2008): "Do we know what makes us happy? A review of the economic literature on the factors associated with subjective well-being," *Journal of Economic Psychology*, 29(1), 94–122.

Du Toit, P. and H. Kotzè (2011): *Liberal Democracy and Peace in South Africa - The Pursuit of Freedom as Dignity*. Palgrave, Macmillan.

Duesenberry, J. (1949): *Income, Savings and the Theory of Consumer Behavior*. Harvard University Press, Cambridge, MA.

Easterlin, R. (1995): "Will raising the Income of all increase the Happiness of all?," *Journal of Economic Behavior and Organization*, 27, 35–47.

——— (2001): "Income and Happiness: Towards a Unified Theory," *The Economic Journal*, 111, 465–484.

Easterlin, R. (2006): "Life cycle Happiness and its Sources: Intersections of Psychology, Economics and Demography," *Journal of Economic Psychology*, 27(4), 463–482.

Fafchamps, M. and F. Shilpi (2008): "Subjective welfare, isolation, and relative consumption," *Journal of Development Economics*, 86, 43–60.

Ferrer-i-Carbonell, A. (2005): "Income and well-being: an empirical analysis of the comparison income effect," *Journal or Public Economics*, 89, 997–1019.

Ferrer-i-Carbonell, A. and P. Frijters (2004): "How Important is Methodology for the Estimates of the Determinants of Happiness?," *The Economic Journal*, 114, 641–659.

Finn, A., M. Leibbrandt and J. Levinsohn (2012): "Income Mobility in South Africa: Evidence from the first two waves of the Income Dynamics Study," *Southern Africa Labour and Development Research Unit*, Working Paper 82.

Fliessbach, K., B. Weber, P. Trautner, T. Dohmen, U. Sunde, C. Elger and A. Falk (2007): "Social Comparison Affects Reward-Related Brain Activity in the Human Ventral Striatum," *Science*, 318, 1305–1308.

Frey, B. and A. Stutzer (2002): "What can Economists learn from Happiness Research?," *Journal of Economic Literature*, XL, 402–435.

Geishecker, I. and M. Riedl (2012): "Ordered data response models and non-random personality traits: Monte Carlo simulations and a practical guide," *Center for European Governance and Economic Development Research Discussion paper number 116*.

Grunow, M. (2014): "Reference-Dependent Effect of Unemployment on Mental Well-Being," *University of Augsburg, Department of Economics Working Paper Number 323*.

Hinks, T. (2012): "Fractionalization and well-being: Evidence from a new South African data set," *New Zealand Economic papers*, 46(3), 253–271.

Hinks, T. and C. Gruen (2007): "The Structure of South Africa's Happiness Equations," *Social Indicators Research*, 82, 311–336.

Johansson-Stenman, O. and F. C. C. Daruvala (2002): "Measuring Future Grandparents' Preferences for Equality and Relative Standing," *The Economic Journal*, 112(April), 362–383.

Kapteyn, A. and F. Van Herwaarden (1980): "Independent welfare functions and optimal income distribution," *Journal of Public Economics*, 14, 375–397.

Kapteyn, A., B. Van Praag and F. Van Herwaarden (1976): "Individual Welfare Functions and Social Reference Spaces," *Report 76.0, Leyden University, Centre for Research in Public Economics, Leyden*.

Kaus, W. (2013): "Conspicious consumption and "race": Evidence from South Africa," *Journal of Development Economics*, 100, 63–73.

Kingdon, G. and J. Knight (2006): "Subjective Well-Being Poverty vs. Income Poverty and Capabilities Poverty?," *Journal of Development Studies*, 42(7), 119–1224.

——— (2007): "Community, Comparisons and subjective well-being in a divided society," *Journal of Economic Behavior and Organization*, 64, 69–90.

Knies, G. (2012): "Income Comparisons Among Neighbours and Satisfaction in East and West Germany," *Social Indicators Research*, 106, 471–489.

Leibbrandt, M., I. Woolard, A. Finn and J. Argent (2010): "Trends in South African Income Distribution and Poverty since the Fall of Apartheid," *OECD Social, Employment and Migration Working Papers No. 101*.

Louw, A. (2007): "Crime and Perceptions after a Decade of Democracy," *Social Indicators Research*, 81(2), 235–255.

McBride, M. (2001): "Relative-income eeffect on subjective well-being in the cross-section," *Journal of Economic Behavior and Organization*, 45, 251–278.

Møller, V. (2007a): "Quality of Life in South Africa: The first ten years of democracy," *Social Indicators Research*, 81(2), 183–201.

——— (2007b): "Satisfied and Dissatisfied South Africans: Results from the General Household Survey in International Comparison," *Social Indicators Research*, 81(2), 389–415.

Møller, V. and W. Saris (2001): "The Relationship between Subjective Well-Being and Domain Satisfactions in South Africa," *Social Indicators Research*, 55(1), 97–114.

Pischke, J. (2011): "Money and Happiness: Evidence from the Industry Wage Structure," *National Bureau of Economic Research Working Paper 17056*.

Posel, D. (2012): "Self-assessed well-being: Analysis of the NIDS Wave 1 and 2 Datasets," *SALDRU, University of Cape Town. (SALDRU Working Paper Number 79 / NIDS Discussion Paper 2012/2)*.

——— (2014): "Self-assessed well-being and economic rank in South Africa," *Development Southern Africa*, 31(1), 51–64.

Posel, D. and D. Casale (2011): "Relative Standing and Subjective Well-Being in South Africa: The Role of Perceptions, Expectations and Income Mobility," *Social Indicators Research*, 104(2), 195–223.

Posel, D. and T. Hinks (2013): "Trusting Neighbours or Strangers in a Racially Divided Society: Insights from Survey Data in South Africa," *Jurnal of African Economies*, 22(1), 136–162.

Posel, D., S. Rudwick and D. Casale (2011): "Is marriage a dying institution in South Africa? Exploring changes in marriage in the context of ilobolo payments," *Agenda: Empowering women for gender equity*, 25(1), 102–111.

Powdthavee, N. (2003): "Is the structure of Happiness Equations the same in Poor and Rich Countries? The case of South Africa," *University of Warwick Economic Research Paper Number 675*.

——— (2005): "Unhappiness and Crime: Evidence from South Africa," *Economica*, 72, 531–547.

——— (2006): "Are there Geographical Variations in the Psychological Cost of Unemployment in South Africa?," *Social Indicators Research*, 80, 629–652.

Pradhan, M. and M. Ravallion (2000): "Measuring Poverty using Qualitative Perceptions of Consumption Adequacy," *The Review of Economics and Statistics*, 82(3), 461–471.

Ravallion, M. (2012): "Poor, or Just Feeling Poor?," *The World Bank Policy Research Working Paper Number 5968*.

Ravallion, M. and S. Chen (2011): "Weakly Relative Poverty," *The Review of Economics and Statistics*, 93(4), 1251–1261.

Ravallion, M. and M. Lokshin (2010): "Who cares about relative deprivation?," *Journal of Economic Behavior and Organization*, 73, 171–185.

Schneider, S. (2010): "The Selection of Pay Referents: Potential Patterns and Impacts on Life Satisfaction," *SOEP Papers on Multidisciplinary Panel Data Research, DIW Working Paper Number 333*.

Senik, C. (2004): "When information dominates comparion: Learning from Russia subjective panel data," *Journal of Public Economics*, 88, 2099–2123.

Stutzer, A. and B. Frey (2010): "Recent advances in the Economics of Individual Subjective Well-Being," *Social Research*, 77(2), 679–714.

Van Praag, B. (2011): "Well-Being Inequality and Reference Groups - An Agenda for New Research," *Journal of Economic Inequality*, 9(1), 111–127.

Van Praag, B., P. Frijters and A. F. i Carbonell (2003): "The anatomy of subjective well-being," *Journal of Economic Behavior and Organization*, 51, 29–49.

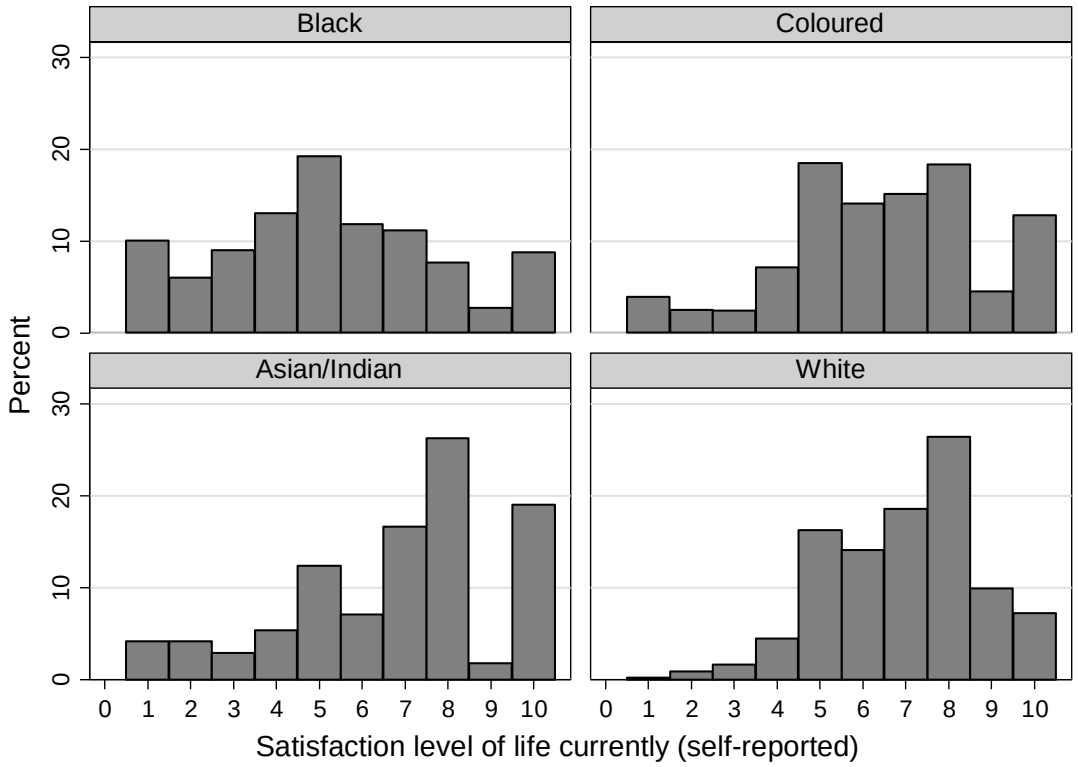
Van Praag, B., A. Kapteyn and F. Van Herwaarden (1979): "The definition and measurement of social reference spaces," *The Netherlands' Journal of Sociology*, 15, 13–25.

Wale, K. (2013): "South Africa Reconciliation Barometer Survey: 2013 Report," Discussion paper, Institute of Justice and Reconciliation.

Yamada, K. and M. Sato (2013): "Another avenue for anatomy of income comparisons: Evidence from hypothetical choice experiments," *Journal of Economic Behavior and Organization*, 89, 35–57.

Appendix

Figure 1: Subjective well-being level by race



Source: NIDS data (2008).

Notes: Weighted satisfaction levels of 13 777 adults who completed the question on subjective well-being, measuring, on a scale from 1 to 10, the level of satisfaction with life experienced by each adult (with 1 signalling extreme dissatisfaction and 10 signalling extreme satisfaction).

Table 1: Summary statistics of subjective well-being by race

	Mean	Standard deviation	Median	Mode	Number of observations
Black	5.229	2.553	5	5	10 690
Coloured	6.464	2.294	7	5	1 991
Asian & Indian	6.809	2.474	7	8	218
White	6.950	1.766	7	8	878
Total	5.567	2.536	5	5	13 777

Source: NIDS data (2008).

Notes: Weighted satisfaction levels of 13 777 adults who completed the question on subjective well-being, measuring, on a scale from 1 to 10, the level of satisfaction with life experienced by each adult (with 1 signalling extreme dissatisfaction and 10 signalling extreme satisfaction).

Table 2: Descriptive statistics of characteristics of estimation sample

Variable	Mean (standard deviation)				
	Black	Coloured	Asian & Indian	White	Entire sample
Mean household:					
Monthly pc hh inc (2008 Rands)	877.909 (2166.175)	2231.740 (9459.668)	3456.610 (4204.452)	6448.484 (7886.163)	1651.319 (4613.861)
Size	4.864 (3.370)	4.631 (2.611)	4.886 (2.695)	3.079 (1.236)	4.656 (3.181)
Education in years°	8.484 (3.161)	9.025 (3.123)	10.438 (2.581)	12.164 (1.620)	8.972 (3.229)
Proportion of adults employed°°	0.427 (0.383)	0.537 (0.370)	0.585 (0.360)	0.659 (0.383)	0.465 (0.389)
HHI in the cluster	0.043 (0.127)	0.234 (0.224)	0.327 (0.214)	0.325 (0.208)	0.0966 (0.183)
Mean individual:					
Age in years	34.300 (14.913)	36.994 (14.702)	37.528 (15.178)	42.604 (15.178)	35.494 (15.153)
Education in years	8.591 (3.833)	8.969 (3.594)	10.269 (3.202)	12.128 (1.959)	9.043 (3.805)
Proportion male	0.448 (0.497)	0.423 (0.494)	0.417 (0.494)	0.444 (0.497)	0.444 (0.467)
Not economically active	0.364 (0.481)	0.289 (0.453)	0.330 (0.471)	0.265 (0.442)	0.346 (0.476)
Unemployed (discouraged)	0.051 (0.220)	0.071 (0.257)	0.064 (0.245)	0.034 (0.182)	0.051 (0.221)
Unemployed (strict)	0.160 (0.367)	0.112 (0.315)	0.038 (0.192)	0.073 (0.260)	0.144 (0.351)
Employed	0.425 (0.494)	0.528 (0.499)	0.568 (0.497)	0.628 (0.484)	0.459 (0.498)
Married	0.251 (0.434)	0.404 (0.491)	0.586 (0.494)	0.614 (0.487)	0.312 (0.463)
Living with partner	0.102 (0.302)	0.111 (0.314)	0.007 (0.085)	0.040 (0.196)	0.093 (0.291)
Widowed	0.063 (0.243)	0.044 (0.206)	0.076 (0.265)	0.053 (0.223)	0.060 (0.238)
Divorced or separated	0.021 (0.143)	0.059 (0.236)	0.062 (0.242)	0.107 (0.310)	0.034 (0.182)
Never married	0.563 (0.496)	0.382 (0.486)	0.268 (0.444)	0.186 (0.310)	0.500 (0.500)
Number of observations	9 774	1 802	202	727	12 505

Source: NIDS data (2008).

Notes: Descriptive statistics of 12 505 adults (aged 15 years and older) who are included in the estimation sample.°Education in the household is calculated only for individuals aged 18 years and older.°°Proportion of individuals aged between 16 and 64 (both inclusive) in the household who are employed.

Table 3: Distribution of income, education and employment in the residential cluster, district and province of the estimation sample

Variable	Mean (standard deviation)				
	Black	Coloured	Asian/Indian	White	Entire sample
Mean income of other race groups in:					
Cluster	1860.54	2540.55	2825.29	5235.01	2988.28
	(2551.06)	(3093.17)	(2672.61)	(4513.00)	(3602.68)
District	3491.15	2159.26	1799.05	1367.84	3050.15
	(2625.65)	(1220.45)	(774.52)	(554.99)	(2446.72)
Province	4801.13	1981.19	1096.17	1206.52	4082.93
	(1915.29)	(544.82)	(462.74)	(411.50)	(2193.88)
Mean education of other race groups in:					
Cluster	8.114	10.029	10.575	11.398	9.546
	(3.958)	(2.593)	(1.680)	(1.888)	(3.438)
District	10.247	9.883	8.964	9.212	10.046
	(2.428)	(1.208)	(1.083)	(0.971)	(2.206)
Province	11.330	9.734	8.159	8.886	10.851
	(0.947)	(1.072)	(0.975)	(0.789)	(1.334)
Mean employment of other race groups in:					
Cluster	0.482	0.575	0.595	0.723	0.574
	(0.292)	(0.270)	(0.133)	(0.244)	(0.285)
District	0.511	0.495	0.456	0.481	0.504
	(0.191)	(0.102)	(0.063)	(0.081)	(0.171)
Province	0.550	0.467	0.433	0.449	0.529
	(0.093)	(0.078)	(0.032)	(0.060)	(0.096)
Sample size	9 774	1 802	202	727	12 505

Source: NIDS data (2008).

*Note*s: Mean income of other race group is calculated as the weighted average of the per capita monthly household income for households in the geographic area. Mean education of other race group is calculated as the weighted average of the mean education in each household for all individuals aged 18 or older in the geographic area. Mean employment of other race group is calculated as weighted proportion of individuals aged between 16 and 64 (both inclusive) in the household who are employed within each geographic area.

Table 4: Distribution of concentration of race groups

	Black	Coloured	Asian & Indian	White	Entire sample
% of sample living in residential cluster where:					
HHI=0°	82.38	29.91	16.06	13.07	69.37
0<HHI≤0.25	11.89	20.30	18.14	34.78	14.89
0.25<HHI≤0.5	3.92	34.11	40.90	24.43	9.54
0.5<HHI≤0.75	1.81	15.67	24.90	27.72	6.11
% of sample living in district where:					
HHI=0°	18.85	0	0	0	14.87
0<HHI≤0.25	41.14	8.05	7.77	13.75	34.70
0.25<HHI≤0.5	29.84	37.36	63.18	41.84	32.50
0.5<HHI≤0.75	10.17	54.59	29.06	44.42	17.93
% of sample living in province where:					
HHI=0°	0	0	0	0	0
0<HHI≤0.25	52.45	10.69	11.35	27.55	45.26
0.25<HHI≤0.5	42.62	18.84	85.41	44.79	41.74
0.5<HHI≤0.75	4.92	70.47	3.25	27.66	13.00

Source: NIDS data (2008).

Notes: °Regions where there is only a single race group. Statistics for entire sample included in NIDS (28 226 individuals). The HHI in area (cluster, district or province) j is calculated as follows: $H_j = 1 - \sum_{i=1}^{N=4} s_{ij}$where s_{ij} is the proportion of race$i$ residing in area j.

Table 5: Subjective well-being and spatial reference groups (ordered probit model)

Dependent variable: SWB	Relative income			Relative education			Relative employment		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log pc monthly hh income	0.103*** (0.026)	0.131*** (0.024)	0.104*** (0.026)	0.112*** (0.025)	0.128*** (0.024)	0.113*** (0.025)	0.124*** (0.024)	0.127*** (0.024)	0.124*** (0.024)
Mean education in hh	0.024*** (0.007)	0.027*** (0.007)	0.025*** (0.007)	0.014* (0.008)	0.027*** (0.007)	0.013* (0.008)	0.026*** (0.007)	0.026*** (0.007)	0.026*** (0.007)
Employment rate in hh	0.074 (0.059)	0.090 (0.059)	0.078 (0.059)	0.089 (0.060)	0.086 (0.059)	0.094 (0.060)	0.068 (0.059)	0.092 (0.059)	0.068 (0.059)
Black	-0.179** (0.075)	-0.282*** (0.077)	-0.178** (0.074)	-0.197*** (0.070)	-0.274*** (0.075)	-0.196*** (0.070)	-0.263*** (0.074)	-0.275*** (0.075)	-0.270*** (0.074)
Coloured	0.181* (0.083)	0.113 (0.089)	0.181** (0.082)	0.180* (0.081)	0.117 (0.089)	0.180** (0.080)	0.123 (0.087)	0.122 (0.088)	0.123 (0.087)
Asian & Indian	0.332*** (0.128)	0.322** (0.126)	0.368*** (0.135)	0.320** (0.127)	0.307** (0.124)	0.345** (0.133)	0.301** (0.123)	0.299** (0.122)	0.302** (0.123)
Average in cluster	0.091** (0.038)		0.055*** (0.017)	0.055*** (0.017)		0.065*** (0.020)	0.088 (0.132)		0.148 (0.145)
Average in district		-0.071 (0.046)	-0.019 (0.025)		-0.019 (0.025)	-0.053* (0.029)		-0.225 (0.278)	-0.329 (0.306)
Number of observations	12 505	12 505		12 505	12 505	12 505	12 505	12 505	12 505
Number of clusters	400	400		400	400	400	400	400	400
Number of districts	53	53		53	53	53	53	53	53
F-statistic	23.238	23.120		23.788	22.705	23.338	22.468	22.774	21.890

Source: NIDS data (2008).

Notes: Ordered probit regression coefficients with standard errors (clustered at the district level). All controls described in Table 2 have been included, but not all reported. All specifications also include provincial fixed effects. * Significant at the 10% level **Significant at the 5% level ***Significant at the 1% level

Table 6: Subjective well-being and racial reference groups (ordered probit model)

Dependent variable: SWB	Specification		
	(10)	(11)	(12)
Log pc monthly hh income	0.081	0.097***	0.127***
	(0.052)	(0.026)	(0.024)
Mean education in hh	0.025***	0.023***	0.026***
	(0.007)	(0.007)	(0.007)
Employment rate in hh	0.076	0.076	0.085
	(0.059)	(0.059)	(0.059)
Black	-0.379***	-0.164**	-0.283**
	(0.133)	(0.075)	(0.127)
Coloured	0.047	0.196**	0.114
	(0.120)	(0.085)	(0.110)
Asian & Indian	0.218	0.340***	0.298**
	(0.135)	(0.121)	(0.124)
Position in racial income distribution in SA^			
Quintile 2	-0.013		
	(0.059)		
Quintile 3	0.057		
	(0.082)		
Quintile 4	0.096		
	(0.109)		
Quintile 5	0.173		
	(0.164)		
Log of mean own race income in cluster		0.102***	
		(0.038)	
Log of mean own race income in district			-0.009
			(0.057)
Number of observations	12 505	12 505	12 505
Number of clusters	400	400	400
Number of districts	53	53	53
F-statistic	20.559	23.141	23.615

Source: NIDS data (2008).

Notes: ° As per reported income. Ordered probit regression coefficients with standard errors (clustered at the district level). All controls described in Table 2 have been included, but not all reported. All specifications also include provincial fixed effects. * Significant at the 10% level. **Significant at the 5% level ***Significant at the 1% level

Table 7: OLS estimates of preference parameters

Dependent variable: SWB	Specification		
	(13)	(14)	(15)
Log pc monthly hh income	0.213***	0.215***	0.217***
	(0.063)	(0.064)	(0.064)
Mean education in hh	0.059***	0.059***	0.059***
	(0.018)	(0.018)	(0.018)
Employment rate in hh	0.097	0.101	0.110
	(0.142)	(0.143)	(0.143)
Black	-0.261	-0.124	-
	(0.176)	(0.280)	
Coloured	0.667***	0.739***	-
	(0.197)	(0.237)	
Asian & Indian	0.315	0.341	-
	(0.240)	(0.244)	
Log of mean own race income in cluster	0.329***	0.291***	0.283***
	(0.075)	(0.112)	(0.112)
Log of mean other race income in cluster	0.014	0.014	0.014
	(0.026)	(0.026)	(0.027)
Log of mean own race income in district	-	0.104	0.109
		(0.169)	(0.169)
Log of mean other race income in district	-	-0.006	-0.006
		(0.021)	(0.021)
Log of mean own race income in SA	-	-	-1.361***
			(0.348)
Log of mean other race income in SA	-	-	-2.095***
			(0.409)
Number of observations	12 505	12 505	12 505
Number of clusters	400	400	400
Number of districts	53	53	53
R-squared	0.123	0.123	0.122

Source: NIDS data (2008).

Notes: OLS coefficients with standard errors (clustered at the district level). All controls described in Table 2 have been included, but not all reported. Race dummies for third specification not included because of perfect multicollinearity with mean income per race at national level. * Significant at the 10% level. **Significant at the 5% level ***Significant at the 1% level

Table 8: Non-linear estimation of preference parameters

Dependent variable: SWB	Specification		
	(16)	(17)	(18)
Weight on own race (μ)	1.022***	0.832***	0.757***
	(0.019)	(0.163)	(0.185)
Weight on income (λ_y)	0.193***	-0.166***	0.214***
	(0.053)	(0.053)	(0.063)
Weight on others in cluster (θ_c)	1.574*	2.183*	1.788**
	(0.860)	(1.185)	(0.896)
Weight on others in the district (θ_d)	0.953	-0.541	-0.523
	(0.787)	(0.967)	(0.684)
Weight on others in the country (θ_n)	-1.917***	-0.918	-0.806*
	(0.633)	(0.609)	(0.433)
African	-1.000***	-0.602***	-0.558***
	(0.113)	(0.179)	(0.195)
Coloured	0.225	0.395**	0.420**
	(0.178)	(0.160)	(0.165)
Asian/Indian	0.132	0.234	0.210
	(0.255)	(0.243)	(0.249)
Number of observations	12 506	12 505	12 506
Number of clusters	400	400	400
Number of districts	53	53	53
R squared	0.095	0.112	0.119
P-score from Wald test: $\mu = 1$	0.221	0.303	0.191
Additional individual controls	N	Y	Y
Household controls	N	N	Y

Source: NIDS data (2008).

Notes: Non-linear least squares regression coefficients with standard errors (clustered at the district level). * Significant at the 10% level **Significant at the 5% level ***Significant at the 1% level

Table 9: Non-linear estimation of preference parameters - testing for altruism

Dependent variable: SWB	Specification		
	Including access to services	Cluster size >= 100	Cluster size < 100
Weight on own race (μ)	0.684***	0.709***	1.022***
	(0.260)	(0.258)	(0.028)
Weight on income (λ_y)	0.210***	0.137	0.264***
	(0.061)	(0.176)	(0.060)
Weight on others in cluster (θ_c)	1.298*	9.069	0.750*
	(0.732)	(15.056)	(0.440)
Weight on others in the district (θ_d)	-0.820	-1.861	0.904
	(0.779)	(3.822)	(0.563)
Weight on others in the country (θ_n)	0.849	-2.428	-0.154
	(0.668)	(3.760)	(0.386)
Black	-0.271*	-0.908	-0.121
	(0.149)	(0.765)	(0.148)
Coloured	0.515***	-0.645	0.798***
	(0.175)	(0.642)	(0.172)
Asian & Indian	0.293	-0.528	0.045
	(0.255)	(0.430)	(0.434)
Access to piped water	0.152		
	(0.223)		
Access to working electricity	0.334**		
	(0.147)		
Access to flush toilet	0.478***		
	(0.160)		
Number of observations	12 505	3 374	9 131
Number of clusters	400	68	332
Number of districts	53	30	53
R squared	0.130	0.120	0.133

Source: NIDS data (2008).

Notes: Non-linear least squares regression coefficients with standard errors (clustered at the district level). All specifications include individual controls and household controls as set out in Table 2. * Significant at the 10% level **Significant at the 5% level ***Significant at the 1% level

Table 10: Non-linear estimation of preference parameters by race

Dependent variable: SWB	Black sample	Non-black sample
Weight on own race (μ)	0.692***	0.994***
	(0.281)	(0.038)
Weight on income (λ_y)	0.218***	0.208*
	(0.073)	(0.123)
Weight on others in cluster (θ_c)	1.792*	2.212
	(1.030)	(1.875)
Weight on others in the district (θ_d)	-0.506	-1.485
	(0.705)	(1.147)
Weight on others in the country (θ_n)	-4.155	-1.492
	(1.294)	(1.136)
Coloured		-0.161
		(0.199)
Asian/Indian		-0.078
		(0.273)
Number of observations	9 774	2 731
Number of clusters	400	322
Number of districts	53	41
R squared	0.066	0.096
P-score from Wald test: $\mu = 1$	0.273	0.871
Individual controls	Y	Y
Household controls	Y	Y

Source: NIDS data (2008).

Notes: Non-linear ordinary least squares regression coefficients with standard errors (clustered at the district level). * Significant at the 10% level **Significant at the 5% level ***Significant at the 1% level

Table 11: Non-linear estimation of preference parameters with fixed effects

Dependent variable: SWB	Provincial level controls	District level controls
Weight on own race (μ)	0.751***	0.918***
	(0.188)	(0.167)
Weight on income (λ_y)	0.228***	0.216***
	(0.059)	(0.059)
Weight on others in cluster (θ_{ct})	1.323*	1.389*
	(0.686)	(0.712)
Weight on others in the district (θ_d)	-0.955	0.221
	(0.699)	(0.820)
Weight on others in the country (θ_n)	-2.981***	-0.739
	(0.772)	(0.575)
Number of observations	12 505	12 505
Number of clusters	400	400
Number of districts	53	53
R squared	0.140	0.168
P-score from Wald test: $\mu = 1$	0.186	0.624
Individual controls	Y	Y
Household controls	Y	Y
Fixed-effects	Y (provincial)	Y (district)

Source: NIDS data (2008).

Notes: Non-linear least squares regression coefficients with standard errors (clustered at the district level). * Significant at the 10% level **Significant at the 5% level ***Significant at the 1% level

Table 12: Non-linear estimation of preference parameters using alternative income measures

Dependent variable: SWB	Specification		
	Mean of log income	Median of log income	Income excluding own income°
Weight on own race (μ)	0.946***	0.922***	0.618**
	(0.095)	(0.105)	(0.262)
Weight on income (λ_y)	0.222***	0.233***	0.257***
	(0.062)	(0.065)	(0.063)
Weight on others in cluster (θ_c)	1.146*	1.133*	1.050*
	(0.677)	(0.661)	(0.563)
Weight on others in the district (θ_d)	0.526	0.119	-0.372
	(0.896)	(0.813)	(0.284)
Weight on others in the country (θ_n)	-2.888***	0.756	-0.072
	(0.900)	(0.745)	(2.625)
Number of observations	12 505	12 505	12 505
Number of clusters	400	400	400
Number of districts	53	53	53
R squared	0.117	0.116	0.117
P-score from Wald test: $\mu = 1$	0.573	0.461	0.145
Individual controls	Y	Y	Y
Household controls	Y	Y	Y

Source: NIDS data (2008).

Notes: Non-linear least squares regression coefficients with standard errors (clustered at the district level). ° Income excluding own income refers to the definition of income where: (i) own individual income is excluded in the calculation of the own race mean income in the cluster; (ii) own cluster income is excluded in the calculation of the own race mean income of the district; and (iii) own district income is excluded in the calculation of the own race mean income in South Africa. * Significant at the 10% level **Significant at the 5% level ***Significant at the 1% level