

Measuring Socio-Economic Status: A Discussion and Comparison of Methods

Or

Letting the Gini out of the Bottle

Plus

Some Thoughts on Well-Being

By Neil Higgs

Director

Strategic Business Development Unit

Research Surveys (Pty) Ltd

Johannesburg

Part-time lecturer

School of Economic and Business Science

University of the Witwatersrand

Johannesburg

Abstract

People, communities and countries are all measured in many aspects. One of the most common is the measurement of wealth, or lack of it, and overall socio-economic status.

Marketers and researchers face a frequent dilemma in this regard: “should I use LSMs, or monthly household income, or personal income, or.... or...what?”. This paper has relevance for researchers and marketers who are trying to come to grips with the differences between people in terms of socio-economic status and the various ways of looking at wealth and wealth inequality.

The paper looks at the concepts of socio-economic status and social class and discusses and compares ways of measuring socio-economic status (SES). The discussion centres around several key issues that researchers need to consider before adopting any particular measure of SES. Different issues are then highlighted by examining various datasets, including AMPS, before some interesting across-country comparisons are made using a World Bank dataset and the Gini Coefficient – a measure of wealth inequality.

In the course of this investigation, findings that affect how we consider the relative power of money for different income levels emerge. This turns out to have serious implications for geographically-targetted direct marketing.

Finally, the discussion leads to a wider concept – that of well-being – in which SES is embedded. This new measure may have major implications for marketers in the future if they truly wish to understand how people live their every-day lives.

(V4)

Introduction

Wealth (at an individual or household level) can be approximated by variables such as income, but the approximation is poor as many other factors can intrude. At a suburb or community level, there have been many geo-demographic measures proposed to indicate lifestyle or socio-economic status. At a country level, wealth is often measured via GDP per head (often in US dollars), but this, too, is a poor indicator as it also needs to reflect costs – wealth is relative to the cost environment in which one finds oneself. Other related indicators relate to the distribution of wealth – the Gini Coefficient is but one common measure reflecting the relative inequality in wealth across individuals in a country or community, based on a Lorentz Curve.

At an individual or household level, traditionally, in the UK, socio-economic status (SES) is measured using a sophisticated occupation coding regime (Meier and Moy (1999)). In the US, the Census Bureau uses a scoring regime based on occupation, monthly household income and education – but now tends to speak of a person’s “well-being”. In other countries, very often household durables are included, whilst there are schemes that are based on the area in which one lives – these are the geo-demographic segmentations that began in the UK in the late 1970s with ACORN (A Classification of Regional Neighbourhoods), and now includes PRIZM, PINPOINT, Mosiac and others (Higgs (1983)).

In South Africa, there has always been a very high focus on income. In an attempt to find a broader segmentation basis, in 1987/88, SAARF embarked on an extensive analysis of the AMPS data. A large number of the AMPS demographic and shopping habits questions (20 variables) were selected and scored, and then subjected to a principal components analysis (SAARF, 1998). The first principal factor that emerged turned out to represent primarily a wealth continuum. This, in retrospect, is not too surprising as South Africa has one of the largest Gini coefficients in the world (in the 60s on a scale of 0 to 100) (Deninger and Squire dataset, World Bank (2002)). Hence, the principal components analysis simply reflected the fact that most of the demographic and shopping habit variables in South Africa are correlated with wealth.

A related measure is the degree of urbanisation a person, community or country exhibits – an urban-rural continuum – (Higgs (1987), Higgs (1991) and Boniaszczuk (1992)). Another author (Matison (1991)) has looked at expenditure patterns rather than income to address this perennial issue of socio-economic status.

This paper looks at some of the most relevant literature of this topic over recent years in a number of disciplines and examines some of the most relevant measures (to South Africa) of socio-economic status at an individual level, at a household level, at a community level and at a country level. It shows the construction of some SES measures from other countries that could have relevance here.

The different measures explored use different techniques for their construction, from subjective grouping and indexing, to factor analysis, to correspondence analysis to graphical methods to cluster and related techniques. These techniques are briefly explained.

Some interesting analyses show the relationship between some of the variables discussed, especially when one looks at the relative power of money, and what this means at neighbourhood level.

The discussion then turns to what becomes a wider issue – that of well-being and its relationship to marketing. Well-being could turn out to be one of the most important concepts to arise in understanding people’s lives.

Class and Socio-Economic Status – Conceptual Underpinning

Social class, social grade and socio-economic status are often used almost interchangeably by people in the marketing research world. But, in reality, they all have different origins and, to sociologists, they should not be interchangeable at all. However, there is in the literature something of a lack of a clear consensus on their conceptual meaning, let alone their measurement. Hence, before we begin to address the main topic of this paper, it is necessary to clarify these terms and to define precisely the scope that we cover.

“The number of studies employing class or SES measures is vast and this makes an attempt at reviewing all such works futile.” (Bollen et al, 1999)

However, as one reviews the literature, especially that in the health and related arenas, certain principles do begin to emerge, and it is these that we need to specify.

Social class has perhaps the longest history, having its origins in the unitary concept that was espoused by Marx. This concept distinguished different classes in relation to a country's means of production (Bollen et al, 1999). In an agrarian society, one has land-owners vs peasants. In an industrial country, there are the owners of the means of production, the bourgeoisie, and the workers, the proletariat. Other classes can be acknowledged but Marx felt that they were not material as they were not drivers of change in social organisation. In developing societies, which have features of both, the situation is less clear, but some analysts have attempted a Marxist approach to social class that categorises workers according to –

- ownership;
- authority; and
- expertise.

As Bollen et al (1999) point out, all three of these variables can be regarded as a type of exploitation. Of course, a fundamental problem here too is that only workers can be classified. This is part of a larger issue that we address later.

The term “social class” has become embedded in UK terminology since it has been used by the British Registrar general since the 1911 census (Meier and Moy, 1999). The occupation of the head of the household is placed in one of five social classes: I, professional; II, intermediate; III, skilled (non-manual and manual); IV, partly skilled; and V, unskilled. There is also a “don't work” class. This measure also reflects a unitary construct. A comparable scale for the U.S. is the Edwards Social-Economic Grouping of Occupations. Here, occupations are compiled into major groups based on the average income and education required by the occupation (Bollen et al, 1999). Indeed, very often occupations are synonymous with social classification. All these approaches use a grouping system that does not depend on any kind of scoring regime and so can be distinguished from most socio-economic measures that tend to be more continuum-based.

The main problem with these types of social classification is that, in reality, only workers can be usefully classified (Meier and Moy, 1999). Further, in Britain, it was found that this system was not discriminating enough in terms of commercial goods and services, and especially in terms of media consumption. This led to the development of Social Grading as used on the National Readership Survey (NRS). This system is strictly hierarchical and is more discriminating (it has six very different classes, where manual and non-manual occupations are split (see table 1 overleaf). It allows everyone to be classified because it depends on the Chief Income Earner's occupation (or previous occupation if no longer working), and is applied to everyone in the household. This means

that everyone has a social grade, independently of her or his actual economic status. But a person could also have an individual grade based on her or his actual economic status as well.

And this begins to highlight some of the issues surrounding this whole arena. In brief, what is the unit of measurement? Is it the individual based on individual characteristics, the individual based on household characteristics, the individual based on community characteristics or what? This raises the whole issue of levels of aggregation of key variables and provides a possible framework within which to begin to consider what measures are best for particular purposes. This topic, which is central to any discussion on wealth and wealth inequality, is taken up in the next section.

In the 2001 census, the British Government introduced a new Socio-economic Classification (SEC) that groups occupations not on the basis of skills (the old approach) but on the basis of employment conditions and relations (occupation and employment status, where employment status looks at self-employed/employee and size of organisation). This is all based on either the current occupation or the former occupation, allowing everyone to have a class) (see table 1, adapted from Meier and Moy, 1999). It has eight classes but is less hierarchical than the NRS Social Grade (classes four and eight depart from the strict ordering of the other classes, and, indeed, are more heterogeneous than the other classes – hence, the system is less unitary and more multi-dimensional than the other systems). In a comparison of the Social Grade and NEC, Meier and Moy (1999) found that Social Grade is still the more discriminatory for commercial purposes, despite its being based on a “concept of society that seems to belong more to the 19th Century rather than the 20th Century, never mind the 21st” (Meier and Moy, 1999)

Table 1 – Comparing Social Grade, Social Class and Socio-economic Class Systems in Britain (after Meier and Moy, 1999)

Old Social Class	Social Grade	Socio-economic class
I Professional	A Higher managerial, admin, professional	1 Higher managerial and professional
II Intermediate	B Intermediate managerial, admin, professional	2 Lower managerial and professional
III Skilled (manual and non-manual)	C1 Supervisory, clerical, jnr admin and professional	3 Intermediate
IV Partly skilled	C2 Skilled manual	4 Small employers and own account workers
V Unskilled	D Semi skilled and unskilled manual	5 Supervisors/craft and related
	E State dependants, casual & lowest grade workers	6 Semi-routine occupations
		7 Routine occupations
		8 Never worked/long-term unemployed

The European Society for Opinion and Marketing Research (ESOMAR) define social class in terms of occupation (or past occupation) (combined with degree of responsibility in the job function) and education (defined in terms of terminal education age. This is in terms of the Main Income Earner (MIE) of the household so that everyone in the household has the same social class. Again, sometimes allowance is made for the classification of the individual based on that person’s own responses to these variables. The system is quite explicit:

Exhibit 1 - ESOMAR Social Class

		M.I.E OCCUPATION						
		E 1+2	E 3+5	E 4,6+7	E 12	E 8+9	E 11+14	E 15
M.I.E'S TERMINAL EDUCATION AGE	21+	A	A	B	B	C1	C1	D
	17-20	A	B	C1	C1	C2	C2	D
	15-16	B	C1	C2	D	D	D	E1
	14	C1	D	D	E1	E1	E1	E3
	13	D	D	D	E3	E2	E2	E3

The definition of each of the M.I.E Occupation categories on which the matrix is based is as follows.

- E1: General management, director or top management with responsibility for six employees or more;
- E2: Self-employed professional,
- E3: Employed professional,
- E4: General management, director or top management with responsibility for five employees or less;
- E5: General management, director or top management with responsibility for six employees or less;
- E6: Middle management, other management with responsibility for five employees or less;
- E7: Business proprietor, owner (full/partner) of company OR owner of a shop, craftsman, and other self employed person with responsibility for six employees or more;
- E8: Employed position, working mainly at desk;
- E9: Business proprietor, owner of company or owner of a shop, craftsman, and other self employed person with responsibility for five employees or less;
- E10: Student;
- E11: Employed non-manual position, not at a desk but travelling or in a service job;
- E12: Farmer & Fisherman;
- E13: Responsible for ordinary shopping and looking after the home, housewife;
- E14: Supervisor & skilled manual worker;
- E15: Other (unskilled) manual worker, servant;
- E16: Retired or unable to work through illness, unemployment or temporarily not working.

Whilst ESOMAR refers to social class, ESOMAR members seem to refer to this system as social grade, social class and socio-economic status interchangeably. Conceptually, and in terms of the terminology used, it is closest to the British idea of social grade.

As one moves into what are traditionally classed as socio-economic measures (SES), the emphasis on grouping begins to diminish in favour of continua and scoring regimes. Some of these are still based on occupation. The most well-known of these are Duncan's Socio-Economic Index (SEI) – a scoring of the standing of occupations in the US. This index is based on the average educational attainment and income in an occupation. Other use various weightings of occupation based on perceived prestige of the occupation (Bollen, 2001). However, most tend to follow more complex patterns, and the term has tended to become more synonymous with combinations of variables, scored and summed, following the lead of the US Census in 1964 in constructing a index based on occupation, income and education. Another class of multi-variable measures rather focus on expenditure, generally using durable goods as the proxy for this, on the grounds that the

expenditure patterns denoted by these variables represent the longer term or smoothed income of households.

Nowadays, the US Bureau of the Census tends not to look at socio-economic status but rather talks of “well-being”, defined as follows (US Bureau of the Census website, March 2002):

“Personal or household income is generally regarded as the single best measure of the degree to which people are “well off.” But other factors also contribute to people’s well-being. Extended measures of well-being gauge how people are faring at the household level. Included are possession of consumer durables, housing and neighborhood conditions, and the meeting of basic needs.”

This also raises the question of access to as against making use of community and neighbourhood facilities – the question of levels of urbanisation then arises, and its links with socio-economic status – see Higgs (1987). It is clear that the concepts of social class, social grade, socio-economic class, socio-economic status and well-being all have some commonalities but all have some differences.

Further, though, this definition of well-being only goes as far as basic needs – it seems that true well-being must take in higher-order needs as well for it to be really useful. What these higher needs are begin to relate to non-physical needs such as belonging, self-esteem and self-actualisation, no less important in understanding people’s lives,

For our purposes, we will define the following terms:

Table 2 – Definitions of Concepts

Term	Broad definition
Social class	A grouping based on a worker’s relationship to means of production – restricted to workers, and Marxist in conceptual underpinning
Social grade	Manual vs non-manual occupations each grouped into a high/medium/low set of categories based on skills, qualifications and responsibilities – applied to everyone as occupation is that of Chief Wage Earner or MIE in the household, but can be categorised based on individual responses (the NRS social grade and ESOMAR’s social class)
Socio-economic class	Grouping based on occupation and employment status - applied to everyone as occupation is that of Chief Wage Earner in the household, also sometimes an individual social class is given based on individual responses
Socio-economic status	A continuous index based on one or more variables applied at either an individual or higher level
Well-being	Socio-economic status as well as community variables covering the meeting of people’s needs

With this as a background framework, we now define the initial scope of this paper’s investigation.

The definition of social class adopted here means that this concept is not very useful for marketing research purposes (this definition differs from that used in the past by the British Registrar General, which is closer in concept to the one adopted for *socio-economic* class).

Social grade may be a useful concept in a first world country but it appears to be less relevant in other less developed countries due to the often high levels of unemployment, the presence of a large rural subsistence farming population and the existence of a considerable informal sector in the economy. These factors mean that groupings would tend to be rather heterogeneous and probably not equally useful for different uses. The same may be said to apply to socio-economic class – and this has yet to be a proven concept even in Britain.

Well-being has considerable potential and may be a useful broader concept in which socio-economic status is embedded. This in turn suggests that we need to focus first on the measurement of socio-economic status, and to ensure that we understand this concept so that we can choose an appropriate measure to suit a given need.

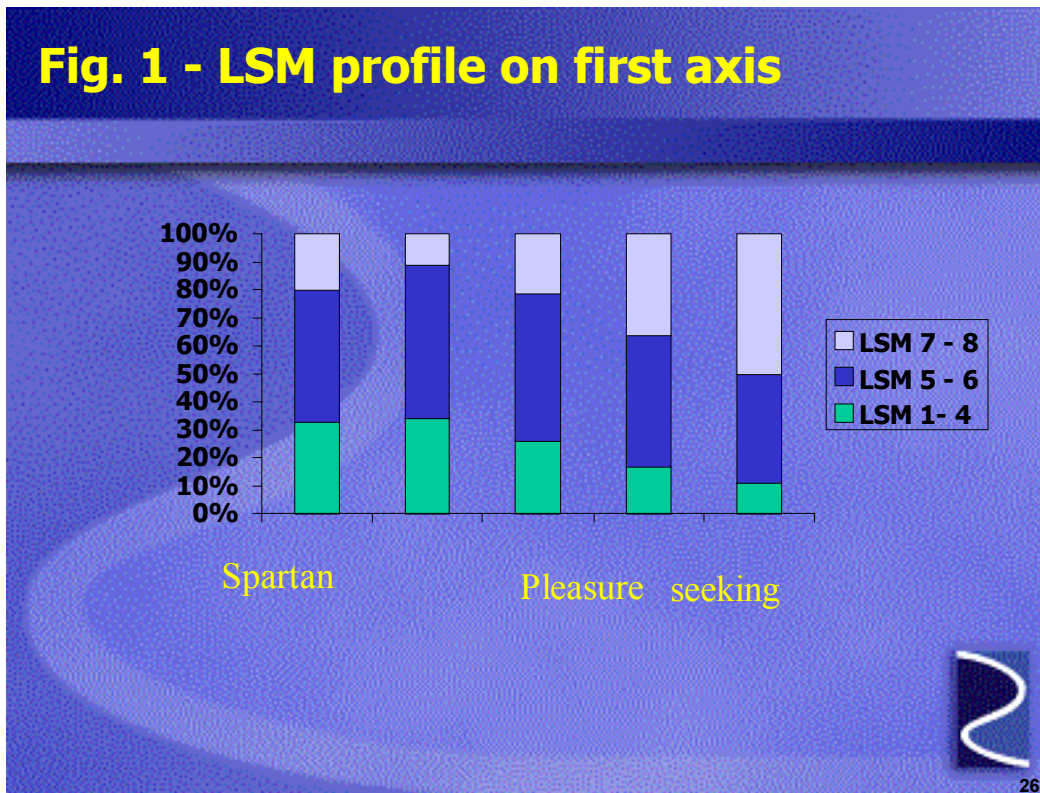
Why Measure Socio-economic Status?

We defined the various terms above, but here, perhaps, we need to pause and ask ourselves “is this a relevant variable”? This is a key question to answer in this 21st Century as we pride ourselves on the concept of the Global Village.

This question is answered at two levels. A review of *some* of the vast array of literature in the field suggests that measures related to wealth and socio-economic status are very important predictors of many outcomes. The bulk of the literature appears to focus on social and health related outcomes that can be quite well predicted by a suitable measure of socio-economic status. For example, many of the papers in the bibliography use socio-economic status (SES) as an explanatory variable for outcomes such as fertility, health (measured both objectively and as self-reported), injury and propensity for injury, mortality, obesity (waist:hip ratio and BMI (Body Mass Index defined as weight in kilograms divided by height in metres squared)), psychological well-being, disability, chronic and degenerative diseases, and nutrition. One review of the literature (Bollen et al, 1999) concluded, “From a sociological perspective, social class and socioeconomic status determine life chances”. Another comparison of SES approaches states that SES has long been a prime predictive variable in epidemiological studies (Deonandan et al, 2000), and as a variable driving health and dietary policies. Its importance in studies in the social sciences is also asserted. Cornish and Denny (1989) state that social grade in Britain is the most effective demographic in common use for a wide range of behaviour and interests.

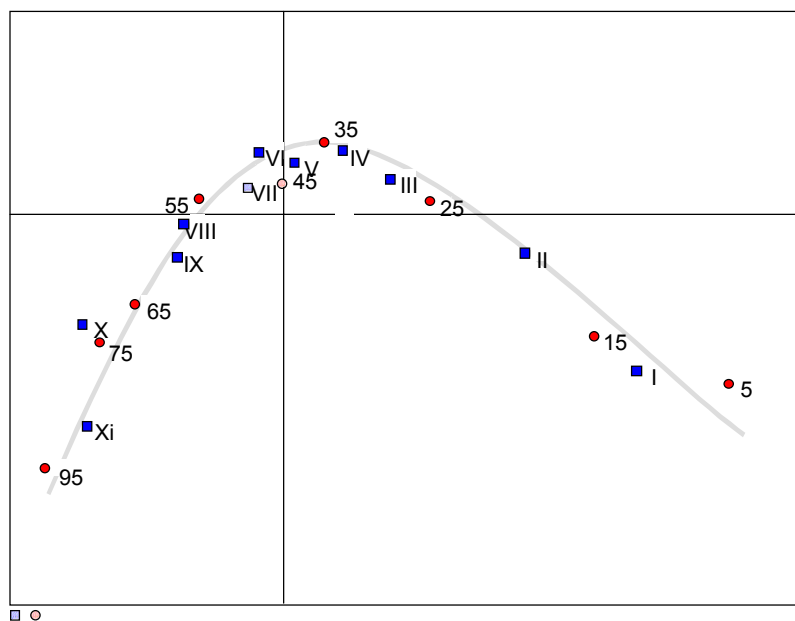
At a second level, empirically, we find here in South Africa that, in many studies, when a segmentation that is not based on an *a priori* theoretical framework is attempted (that is, when the data are subjected to exploratory data analysis techniques), very often the first and strongest differentiator turns out to be a wealth-related index. In the last year, this has been found by the author in such diverse categories as beverages (hot and cold), finances (for a number of financial services) and attitudes to nutrition and exercise, to name just a few.

For example, one report concluded, “Whichever way one segments the category as it is perceived at present, one tends to arrive back at an essentially demographic view of age and wealth”. In another study on beverages, the main differentiator was a continuum that correlated very strongly with LSMs:



In a service category study, the prime segmentation also correlated very strongly with another measure of socio-economic status devised by the author. Here, the segments are numbered in Roman numerals whilst the SES is given by the numbered dots. The SES measure ran from zero to one hundred and the dots represent the mid-points of ten equi-interval classes. The correlation is very strong at 0.9044:

Fig. 2 - Primary segmentation vs Socio-economic status



When SAARF first developed the Living Standards Measure (LSM), the initial set of variables that was selected for consideration were chosen on the basis of its ability to differentiate between consumers:

“After considerable testing of an initial list of 71 characteristics, or variables (from which the population group and media usage were deliberately excluded) thirteen were finally selected for their combined power in differentiating between respondents. Not only did they differentiate well, but they did so in a consistent way when tested on a second set of SAARF AMPS results. This kind of stability was of course a key requirement... . Once the cluster of indicator variables had been chosen, the term ‘Living Standards Measure’ was coined to describe the scale they created.” (SAARF website www.saarf.co.za, 2002).

“Essentially, the LSM is a wealth measure based on standard of living” (Haupt, in MarketPlace, 23 April 2001)

It is clear that the SAARF LSM is one of the measures in the socio-economic status measurement stable. The fact that this variable is such a strong differentiator in the commercial and media world again justifies the examination of ways of measuring socio-economic status.

A Possible Taxonomy of Socio-economic Status Measures

The discussion above, which has provided some essential background and understanding of the whole arena, has highlighted a number of issues. First and foremost, whilst there are a veritable plethora of SES measures in use, there does not appear to be a conceptual framework underpinning them. This means that, practically, when people choose any given SES measure, it is often chosen on that individual measure’s perceived strengths, rather than because it belongs to a particular class of measures. As a result, there appears to be little clear consensus over what types of measure to use in different cases.

As discussed above, it is clear that any measure adopted must enable everyone (not just workers) to be classified or scored. However, what is not clear is what variables to use. It seems that there are two distinctly different sets of choices that must be made:

- The first concerns the level of aggregation issue. For example, does one use data that is respondent-specific (that is, it will vary from individual to individual within a household) or household-specific (that is, everyone in the household will have the same value/classification but different households in a neighbourhood will differ) or neighbourhood-specific or suburb specific or The level of aggregation can continue upwards to the ultimate one of a country, where an average standard of living is given by, say, GDP in US\$ per head. The ESOMAR, British and US Census systems mostly use the values for the main income earner as the values to apply to everyone in a household, but all say in almost a footnote that there will be occasions when a score or classification based on each individual’s responses is required. These cases are seldom explicitly spelt out. And the situation is further complicated by the growth in geo-demographic classification schemes as outlined in the introduction. In a seminal paper at the 1979 Market Research Society Conference in Britain, Baker et al (1979) concludes that “A Classification of Residential Neighbourhoods extends social class as a measure of status” and that “its importance lies in its ability to show clear life-style differences between neighbourhoods leading to many differences in consumer behaviour”. For many researchers, especially in the social and health sciences, there is

not the luxury of data at respondent level. However, there is often data at an area-level, based on census or other data. These researchers use these area-level variables to estimate people's socio-economic status (Deonandan et al, 2000). A considerable number of papers have been written in the social and health science fields comparing area-based methods of socio-economic status determination with respondent-level results, but conclusions are usually too case-specific to be readily generalisable, especially into the marketing arena.

- The second area of variable choice concerns the actual variables. These appear to lie in two broad domains, although the situation is by no means as clear-cut as in the level of aggregation arena. These two domains are –
 - income or income-related variables; and
 - expenditure or expenditure-related variables.

An example of the first set of variables is the income, education and occupations set of variables that feature very strongly in most discussions of SES, because of the US Bureau of the Census's lead in 1963. Education correlates with income, as does occupation. An example of the second set is ownership of durables, shopping patterns and the like, all of which can be scored in many diverse ways. There is potentially a third set of variables that are often used: these refer to parents' education and occupation on the grounds that the previous generation has the most significant effect on a person's future. Very often, these variables are, however, combined with some "present" measures, such as family income. Finally here, the reason that this is less of a clear-cut issue than the aggregation issue is that many measures of SES combine both types of variable. Bollen et al (1999) cites a measure based on family income, mother's and father's education, father's occupation, and material and educational possessions owned by the household.

All of the above suggests, however, at least a consideration framework for socio-economic status measures that might look as follows:

Exhibit 2 – A Possible SES Measure taxonomy

		Level of Aggregation					
		Individual	Household	Neighbourhood	Suburb	Town/City	Country
Income vs Expenditure	Income						
	Expenditure						

Whilst it is fairly common for measures to cut across the income/expenditure divide, it is less common for different levels of aggregation to be included in one measure. We suggest, however, that one needs to consider this explicitly, so that the implications can be clearly assessed. So, for example, if a measure includes individual and household responses, then different individuals in a household will have different scores or classifications. This is not necessarily bad, but one needs to consider if this outcome is desirable in the context of the arena of study. In one of the LSM scoring frameworks, questions about household durables were included along with questions about personal shopping habits and personal banking issues. Hence, a husband and wife could conceivably have different LSM groups. This is not a flaw but a characteristic of the measure. Is it what a marketer or researcher needs for a particular market?

For example, a bank might prefer to look at an individual's income, occupation and education in granting them a current account, for these indicate both current and future potential. But where people live a life together and share leisure activities, major purchases, eating patterns and living standards, perhaps the combined household income and the occupation and education of everyone (except children) needs to be used in an SES measure.

The most recent version of the LSM scoring regime has mostly household variables but there are few which are at community level (water in home, hot running water and, potentially, Telkom phone), at province level (living in Gauteng, Western Cape) and wider (living in non-urban area outside Gauteng and Western Cape). Again, this is not a criticism, but a characteristic of the measure. It almost starts to move LSMs in the direction of "well-being" as defined by the US Bureau of the Census, as this includes community and wider variables.

Each time one includes these wider variables, to some extent, one is applying averages to a group of households and reducing somewhat the variability of the measure. For example, in Zimbabwe, whilst a version of LSMs is in use, another (older) system used the following (from personal correspondence with Probe Market Research (Pvt) Ltd, Harare, February 2002):

Exhibit 3 – Variable used in Socio-economic Status Measurement in Zimbabwe

Car
Radio
TV - B/W or colour
Hifi
Refrigerator
Microwave
Dishwasher
Electricity
Highest education in household
Occupation of CWE
Suburb score

This measure contains a variety of measures, including a suburb score that, in itself, is a measurement of SES at community level. This drop in variability may be of little moment in widely diverse societies but may be more restricting in more homogeneous markets. Hence, again, one needs to consider the use to which the measure will be put. Clearly, as one moves to higher levels of aggregation, one shifts from SES to level of urbanisation to "well-being".

This in turn suggests that researchers and marketers need to have a variety of measures in their arsenal, to apply as the particular case demands. This, indeed, becomes a central idea in this field, and suggests that, here in South Africa with our very diverse society, we need to broaden our thinking as regards SES measures considerably.

In a comparison of income-based vs expenditure-based measures, Bollen et al (2001) concluded that expenditure-based measures often seemed more robust, reliable and predictive where many health and medical issues were concerned. However, the income-based measures do well in fields where new and high expenditure is to be predicted, such as where a couple might seek in-vitro fertilisation (which is very expensive). Hence, again, it seems that we need to consider more than one measure of socio-economic status, and the above framework is presented as a basis for variable selection.

A Review of Some Socio-economic Status Variables and Measures

As noted above, there is a vast literature on SES in many fields. It cannot all be reviewed but certain patterns emerge as discussed above. This means that there are certain variables that are commonly used, and these need to be detailed so that people can know what is a suitable subset of variables and what is not.

What is clear is that SES is generally approached empirically, without the use of any clear framework. As Bollen et al (2001) points out:

“Empirical strategies range from selecting a single variable to proxy economic status to constructing one or more proxies based on a composite of different factors with potentially equal or variable weightings. The proxies themselves are sometimes included as simple dummy variables while at other times they enter as aggregations of multiple assets with estimated or reported values attached to each asset.”

What follows is a partial listing of variables at various levels. This list will help researchers and marketers to open their minds to new possibilities in the socio-economic status measurement field.

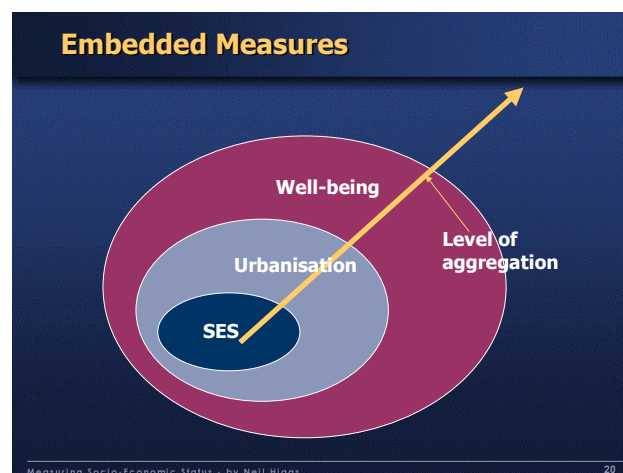
Table 2 – Partial List of Variables at Different Levels, all applied to an individual to obtain an individual score/class

Level	Variable	Comments
Individual	Occupation	With level of responsibility – there are also several occupational prestige scales, and some that have used median income (and median education of given occupations) to obtain the scores
	Education	Qualifications or terminal year (sometimes termed “pedagogic wealth”!)
	Income	Personal
	Personal shopping patterns	
	Personal financial holdings	
	Hobbies/leisure activities	
	Media	
Household	Durables	Various sets (country/society specific including vehicle), scored either as simple counts, as indicator variables, by value (not recommended) and via principal components or similar analysis a la LSMs
	Parent’s occupations	Same comment as above
	Occupation of main wage earner	Same comment as above
	Parent’s education	Mother’s education is an important variable in many arenas
	Household Income	
	Income of main wage earner	
	Type of housing	
	Housing quality	
Neighbourhood/	Area scores	Already a measure of SES

Community/ Suburb	Availability of basic services	Electricity, water at various levels, communications (electronic, electric, roads etc)
	Availability of infrastructure	Different types of retail outlet (approaching the geographical concept of “thresholds” and “relative space”, a key concept in urbanisation (Higgs, 1987)
	Size of community, health, schools	
	Proportion of students	
	Proportion of unemployed	
	Proportion of people under a certain age	Normally picked as the average age of entry into the work force
	Proportion of two-car households	
	Proportion of immigrants	
	Level of overcrowding	
	Type of tenure	
	Type of housing	
	Employment structure	
City/Province	Many of the above	
	GDP per head	
	Human Development Index	(HDI) – incorporates life expectancy and health measures as well as income and education
Country	GDP per head in US\$	
	Gini coefficient	A measure of inequality
	HDI	

Very often, choice of variables is limited by what is available in a dataset.

It is clear that, as one adds in community level data, the measure produced begins to look more and more like a measure of urbanisation (Higgs, 1987). For example, the Research Surveys urbanisation continuum uses access to types of retail outlet, time to nearest food shop, access to basic services, access to print media and questions about one’s mental horizons. Hence, again, it is important to begin to understand the context of these measures so that one can really understand the underlying construct that the measure represents. In parenthesis, too, it is also worthwhile noting that none of the variables above begin seriously to address well-being, and this suggests an initial model linking these three measures might be conceptualised as follows:



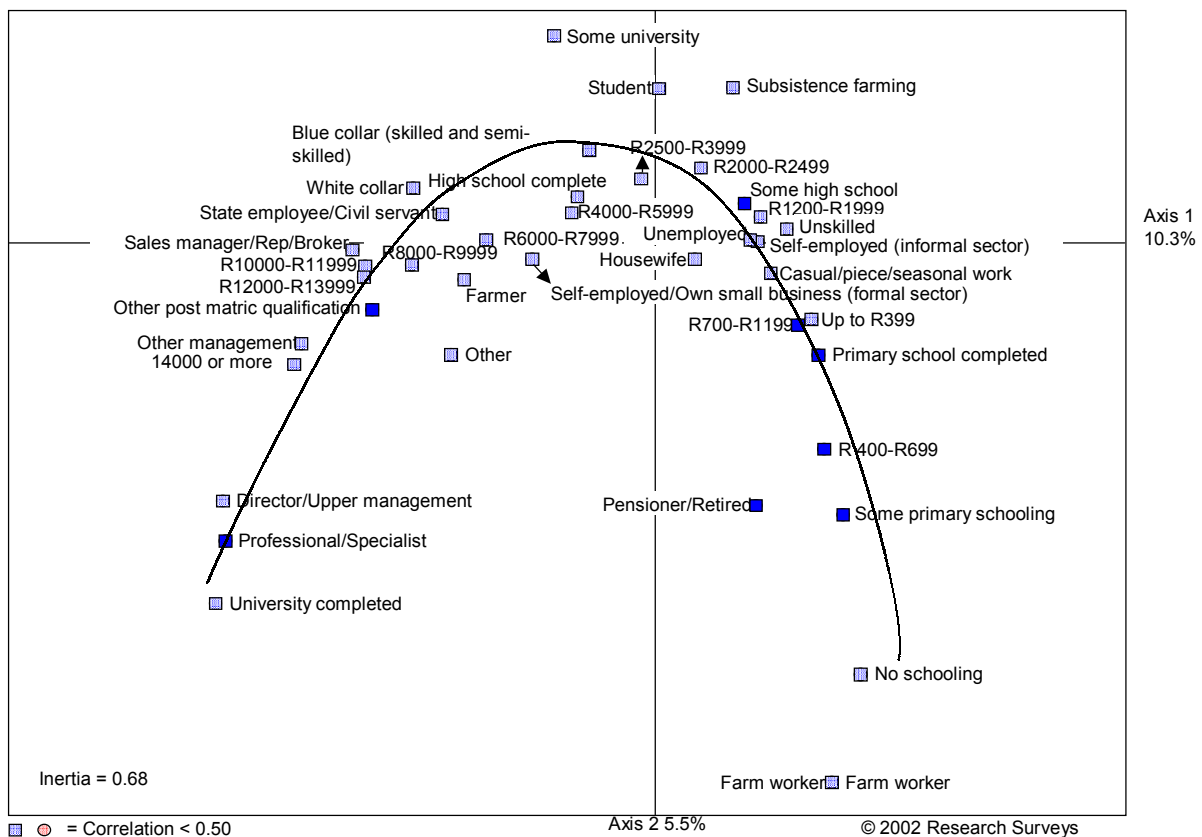
Looking at Some Data

With all the above in mind, it is useful to begin to explore how different measures of socio-economic status compare. This needs far more exploration and what follows just scratches the surface. It is quite crude at this stage but there is a rich field of endeavour waiting here.

A large study covering metro and large urban areas was selected. It was a tracker and the period selected for review covers two years with a sample of approximately 7 000.

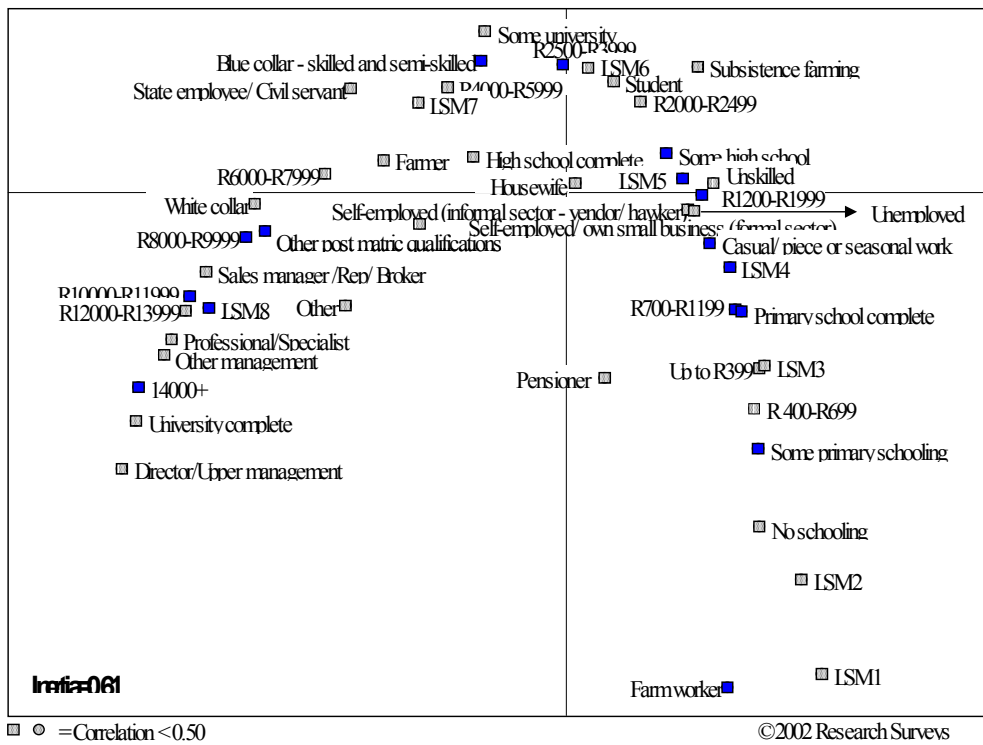
In each case, a large matrix was constructed with its columns being household income, personal education and personal occupation (this choice was dictated by what was on the study – given the earlier discussion, it might have been better to have chosen either personal income or education and occupation of the main wage earner; what we have here is rather disparate set of variables). These same columns were used also as the rows of the matrix, forming a square matrix (Burt matrix). When such a matrix is subjected to correspondence analysis, if a clear continuum emerges linking these variables, then a horseshoe shape often appears (Higgs, 1994), especially where a particularly single-minded continuum is at play. This was indeed the case and shows that these three variables can be usefully combined into a continuum that we will call socio-economic status in the light of the above history.

Fig.3 - Household Income / Education / Occupation



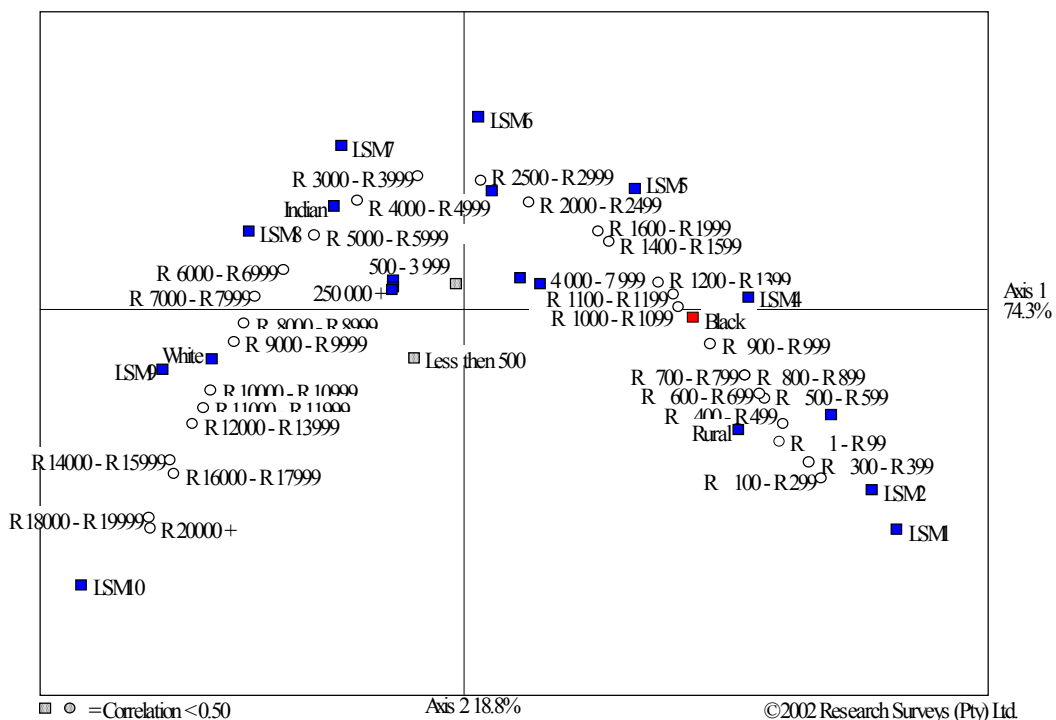
We then overlaid (old) LSMs on to this, with interesting results. There is a clear correlation, and it is evident that LSMs are an excellent differentiator at the lower end but less so at the upper end, justifying the new upper-end extension to LSMs (see Fig. 4).

Fig. 4 - Adding LSMs



[In parenthesis, when one runs the new LSMs against household income, the better differentiating power of the new LSMs becomes immediately apparent:

Household income by demographics – AMPS 2001A



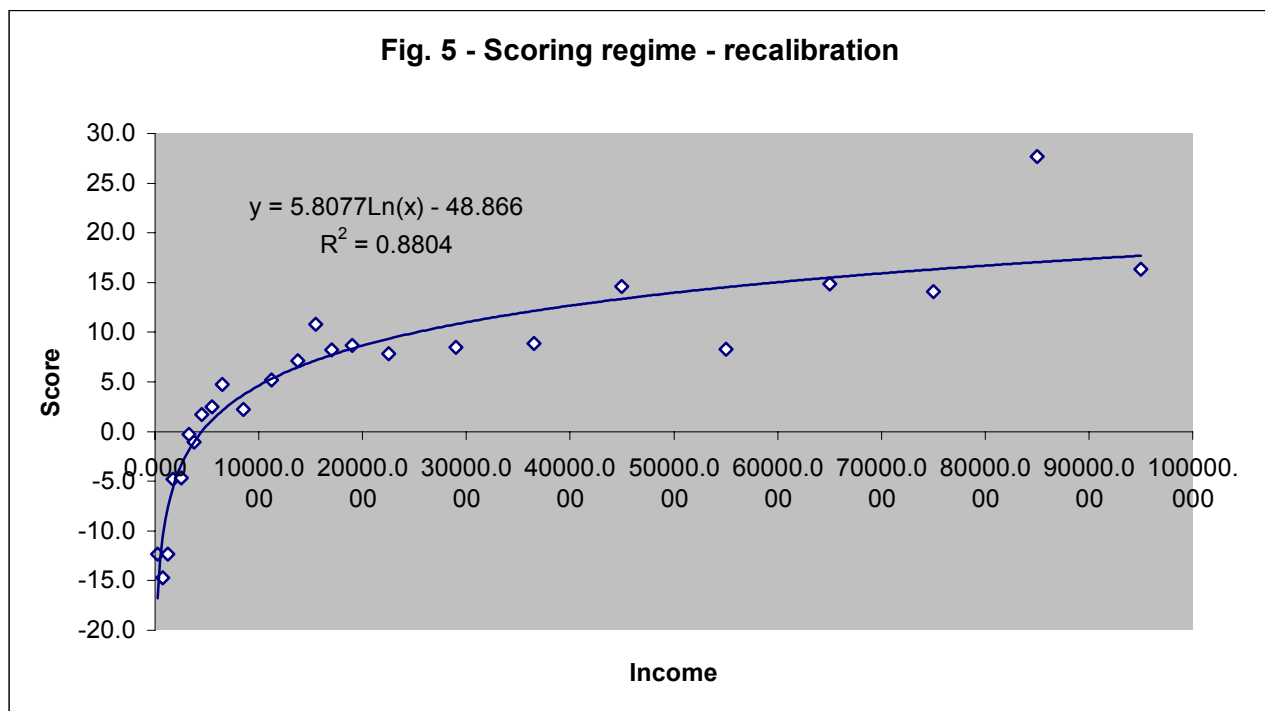
Of course, one might say “why bother?” if there is such a good correlation with LSMs. In South Africa, where there is such a diversity of income levels, any two measures that reach the extremes will show a good correlation. The point is that this is not so for all markets and for all subsets of people, especially within more homogenous subsets. The intention here is to show that there is more than one SES measure – indeed there are literally hundreds. And the need is for people to select the most appropriate one for the purpose at hand and for the target market of interest. Working at country level is one thing: working within smaller target markets is a very different thing altogether. LSMs are indeed very good – especially when considering the whole country together. But it is time to consider a wider array of options for other circumstances. There is a need to begin to develop such measures ready for use, or to develop guidelines for the construction of such measures. This paper represents just a small start in this direction.

From the maps depicted above, one can calculate the contribution that each level of each variable makes to a scored continuum, and so construct an SES measure, either by using the co-ordinates of the levels on the first principal axis (an approach similar to the principal components approach described earlier), or by unfolding the horseshoe. For simplicity and illustration, the first approach was used here. This yielded a table of scores, which are scaled so that the maximum is 100 and the minimum is zero (see table 3 below).

Table 3 - Early Scoring Regime for an Income-based SES measure

H/hold income	Up to R399	-8.7
	R 400-R699	-9.4
	R700-R1199	-7.9
	R1200-R1999	-5.8
	R2000-R2499	-2.5
	R2500-R3999	0.8
	R4000-R5999	4.7
	R6000-R7999	9.5
	R8000-R9999	13.6
	R10000-R11999	16.3
	R12000-R13999	16.4
	14000 or more	20.3
Education	No schooling	-11.4
	Some primary schooling	-10.4
	Primary school completed	-9.1
	Some high school	-4.9
	High school complete	4.4
	Some university	5.7
	University completed	24.6
	Other post matric qualification	15.9
Occupation	Director/Upper management	24.2
	Other management	19.8
	Sales manager/Rep/Broker	17.0
	Professional/Specialist	24.1
	Self-employed/Own small business (formal sector)	6.9
	Self-employed (informal sector)	-5.6
	Blue collar (skilled and semi-skilled)	3.8
	Unskilled	-7.2
	White collar	13.6
	State employee/Civil servant	12.0
	Farmer	10.7
	Housewife	-2.2
	Pensioner/Retired	-5.6
	Unemployed	-5.2
	Student	-0.1
	Other	11.5
	Farm worker	-9.8
	Subsistence farming	-4.3
Casual/piece/seasonal work	-6.4	
Constant to add		30.6

The scoring regime is crude, because a simple approach was used and because this particular study does not go far enough into rural areas. Hence, the final distribution is skewed, as can be seen by the fact that the constant to add is well below the midpoint of the 0 – 100 scale. Further, examination of the table reveals that there are some inconsistencies especially on the household income levels, where, in a few instances, the order followed is illogical. This is probably due to sample error, and can be corrected by plotting income against score and reading off better values. The income set is rather short here, but an illustration of the procedure is given below from another study where a similar exercise was conducted.



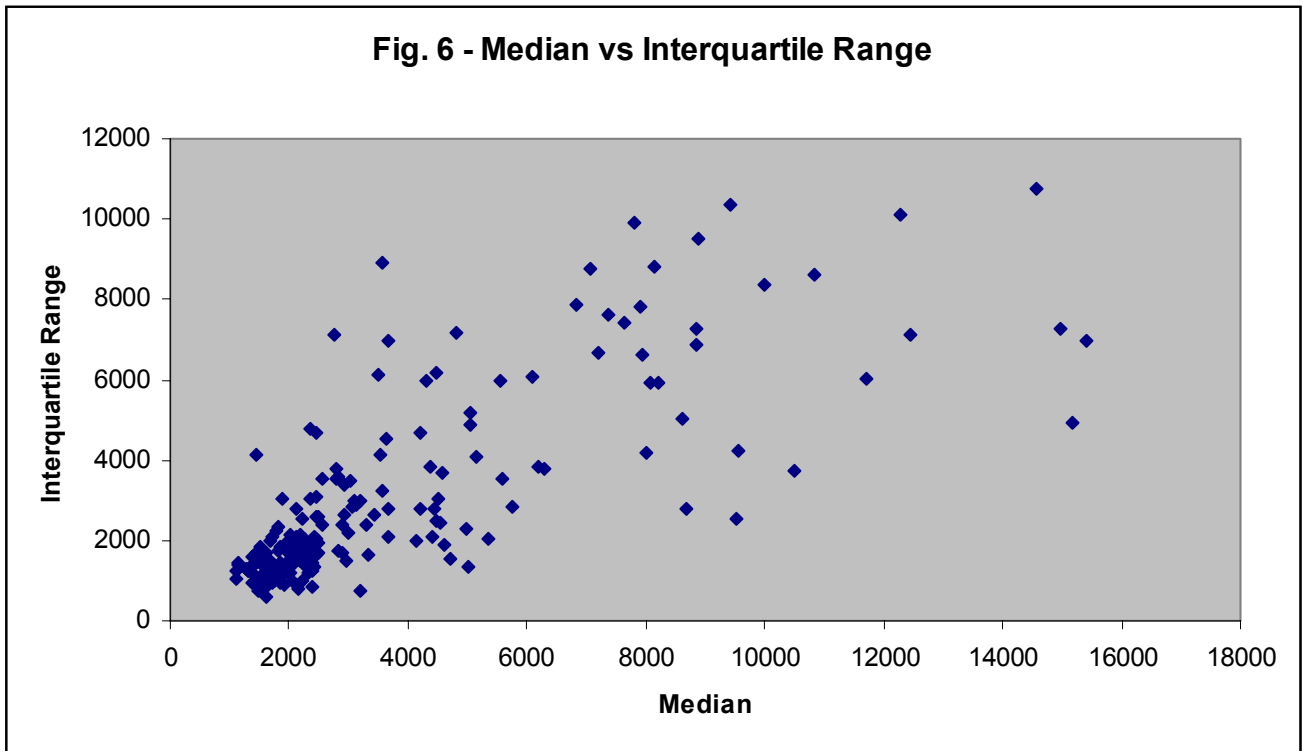
The above calibration graph provides an interesting and important view on the relative utility of money – an increase in income at the lower end of the income scale improves one's socio-economic levels much faster than at higher levels. Finally, the scoring regime in respect of occupation provides an interesting view of the relative worth of jobs. This occupation set was not particularly good (it is not a very relevant variable in the market concerned), but similar work on a comprehensive occupation set will, essentially, provide some sort of prestige rating of occupations.

Geodemographics

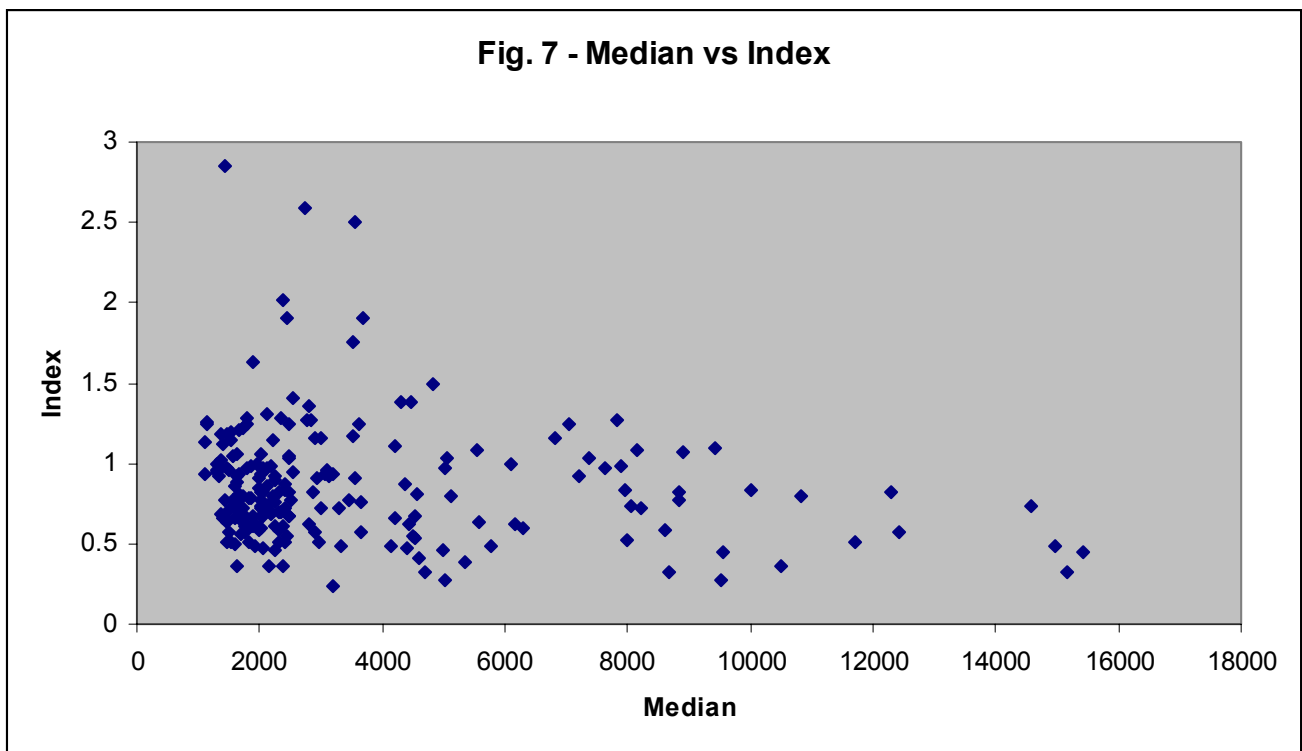
Once one begins to move upwards from household level variables, the question arises as to the extent to which area-based classifications or scores predict individual-level scores. Of course, there is a huge averaging effect, and much of the variability in the real world will be lost. Nonetheless, geodemographics are an important tool for targetting overseas and increasingly here. Deoandan et al (2000) concluded that, in a medical category, the degree of agreement between individually-based SES measures and post-code measures was only moderate, although within each of these groupings, there was good correlation. This is somewhat to be expected, and will be more so if where a person lives is not a matter of free-will in conjunction with economics.

This was examined differently using the same dataset where suburbs were also recorded. For each suburb, the median monthly household income was calculated, as well as the inter-quartile range as

a measure of variability. These were plotted against each other. The result is given below in Fig. 6.



What this shows is that, generally, for low incomes, the range of incomes found in an area is, not unexpectedly, low. That range increases as income rises. Partly, that is because the possible range must increase as income increase. This can be factored out by dividing the inter-quartile range by the median, to index it on income. This is given below:



This shows a different picture. Relative income range drops in higher income areas but is relatively greater for lower income areas. Taking this in conjunction with the income calibration curve of Fig. 5, we have an important finding: differences in income are much more important in low income areas and for lower income earners than is the case for high income areas or earners. Hence, for anyone using area as a predictor of socio-economic status, targetting will be poorer in lower-income areas in the urban areas of South Africa than in higher income areas. Marketers need to establish the income variability of individual suburbs when conducting such area-based targeting operations.

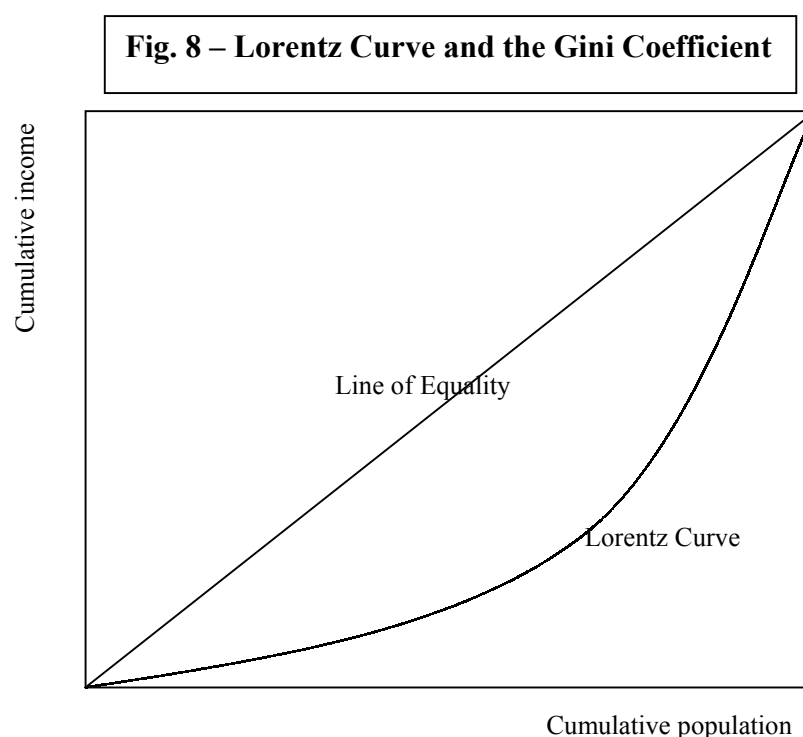
The Gini Coefficient

Measuring socio-economic status is really about measuring differences between people. This, and the findings above, lead into a discussion of inequality: the Gini coefficient is the main measure here.

The Gini coefficient was developed by Italian statistician Corrado Gini to provide a mathematical expression of the degree of concentration of wealth or income. While it has been criticized over the years, it continues to be used by social scientists describing inequality or comparing inequality among nations. A Gini coefficient of approximately 40 is normal for most developed economies.

The coefficient is based on the construction of a Lorenz curve, developed by Max O. Lorenz in order to describe inequality in a society. Imagine a graph in which the cumulated income (expressed as a percentage) is placed on the vertical axis and the cumulated number of households (expressed as a percentage) is placed on the horizontal axis. If there were perfect equality (so that the first 10% of the households received 10% of the income and 20% of the households received 20% of the income, and so on), a diagonal line would be drawn across the graph. When actual income distributions are depicted on this graph, the curve departs from the line of perfect equality. For example, the bottom 20% of households may receive only 4.5% of the total income.

This line is the Lorenz curve. The Gini coefficient is an expression of the ratio of the amount of the graph located between the line of perfect equality and the Lorenz curve to the total area of the graph below the line of equality. This is illustrated below:



Below are some selected Gini coefficients taken from the Deininger and Squire Dataset (World Bank, 2002) that illustrate the issue. Since this is essentially a measure of inequality, it will also be interesting to compare these coefficients with measures such as the inter-quartile range. In theory, it is possible to calculate the Gini coefficient for any subset of people. Given the wide range of incomes seen within suburbs of South Africa, it seems that it is important to watch such measures when looking at socio-economic status measures at that or higher levels of aggregation.

For example, South Africa has an approximate GDP per head of US\$2 800 per head, compared with the whole of Africa at US\$600. But the high Gini coefficient associated with South Africa means that that figure is a very poor reflection of our overall socio-economic status. The same will be true at a smaller level when we talk of suburbs.

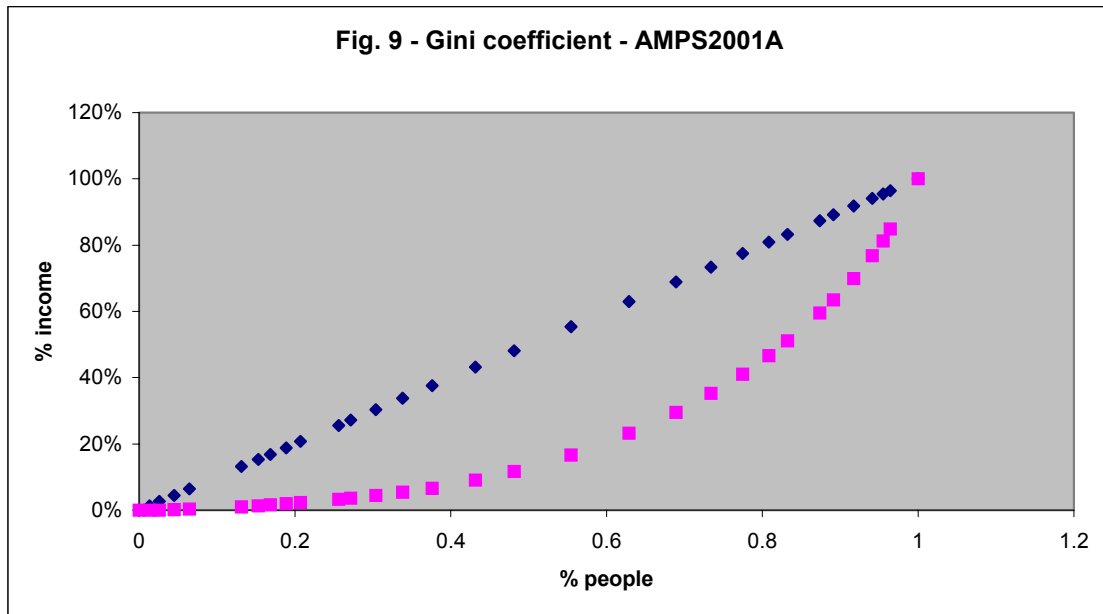
The suburb data given above also reflects that the fact that most of our suburbs are rather large. Most overseas geo-demographic systems work at post-code or enumeration area (EA) level. Using such small neighbourhoods (between 15 and 100 households) greatly improves the ability of area-based socio-economic status measures in predicting individual socio-economic status. All this means that South African marketers need to use geo-demographic classifications wisely by asking for information on the variability within the classifications given. Measures such as the Gini coefficient and other standard measures of dispersion such as the inter-quartile range need to be employed here.

Table 5 – Gini Coefficients around the World

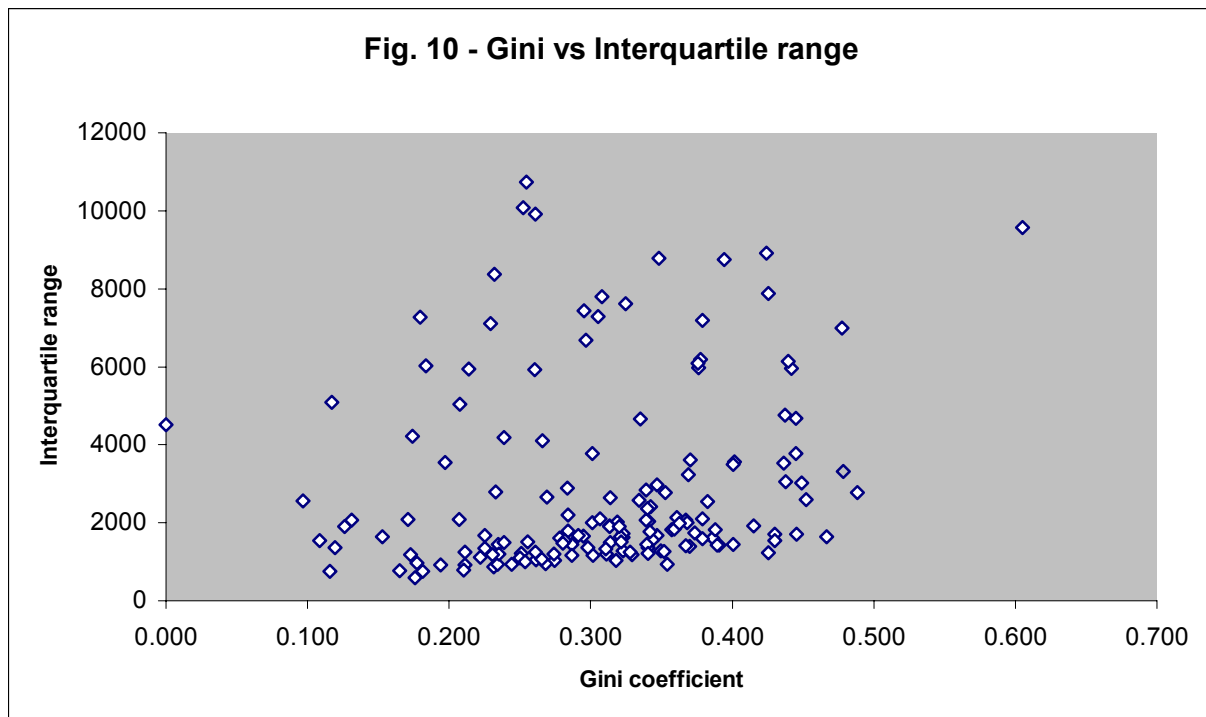
Country	Year	Coefficient
Algeria	1988	39
Australia	1990	42
Bangladesh	1992	28
Belgium	1992	27
Brazil	1989	60
Canada	1991	28
Chile	1994	56
China	1992	38
Egypt	1991	32
France	1984	35
Ghana	1992	34
India	1992	32
Japan	1990	35
Kenya	1992	55
Malawi	1993	62
Nigeria	1993	37
South Africa	1993	62
UK	1991	32
USA	1991	38
Zimbabwe	1990	57
Zambia	1996	56

The Lorentz curve and associated Gini coefficient were also calculated for the AMPS2001A dataset. The Gini coefficient is 53, in contrast with the 62 given above. The 1994 figure was 57.

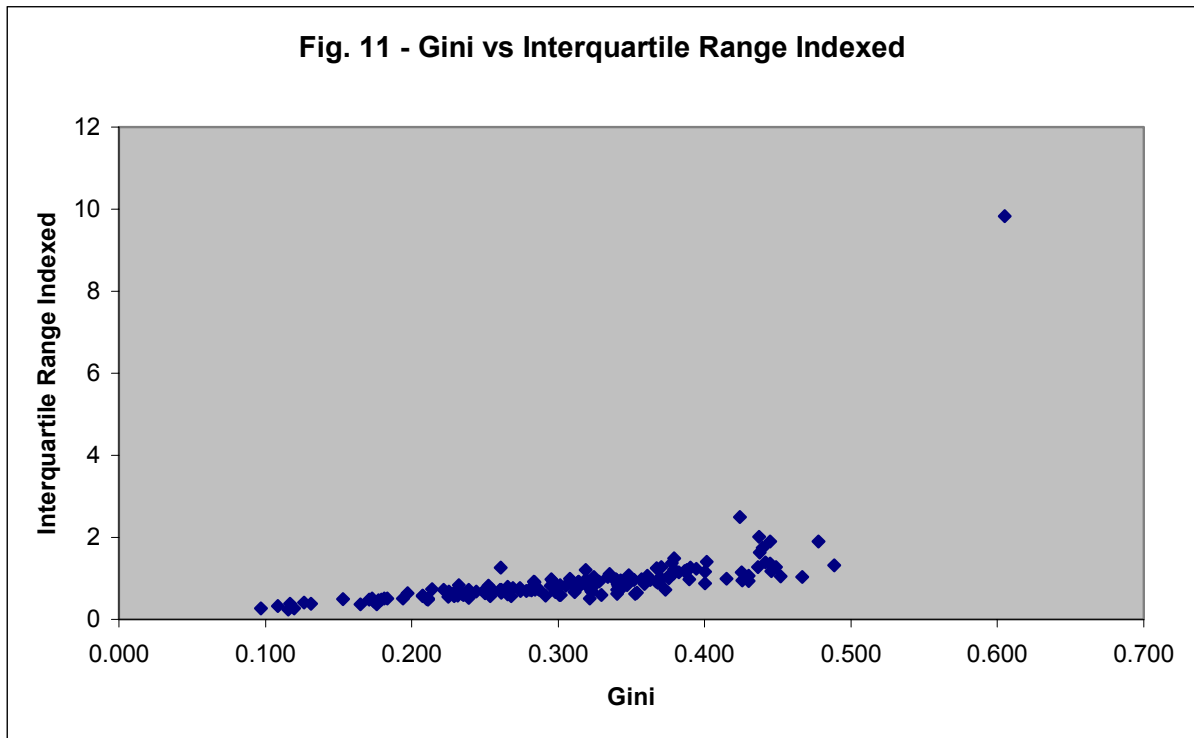
This suggests two conclusions: that South Africa is slowly improving its income inequality levels – but that AMPS may not cover the full range of incomes in South Africa.



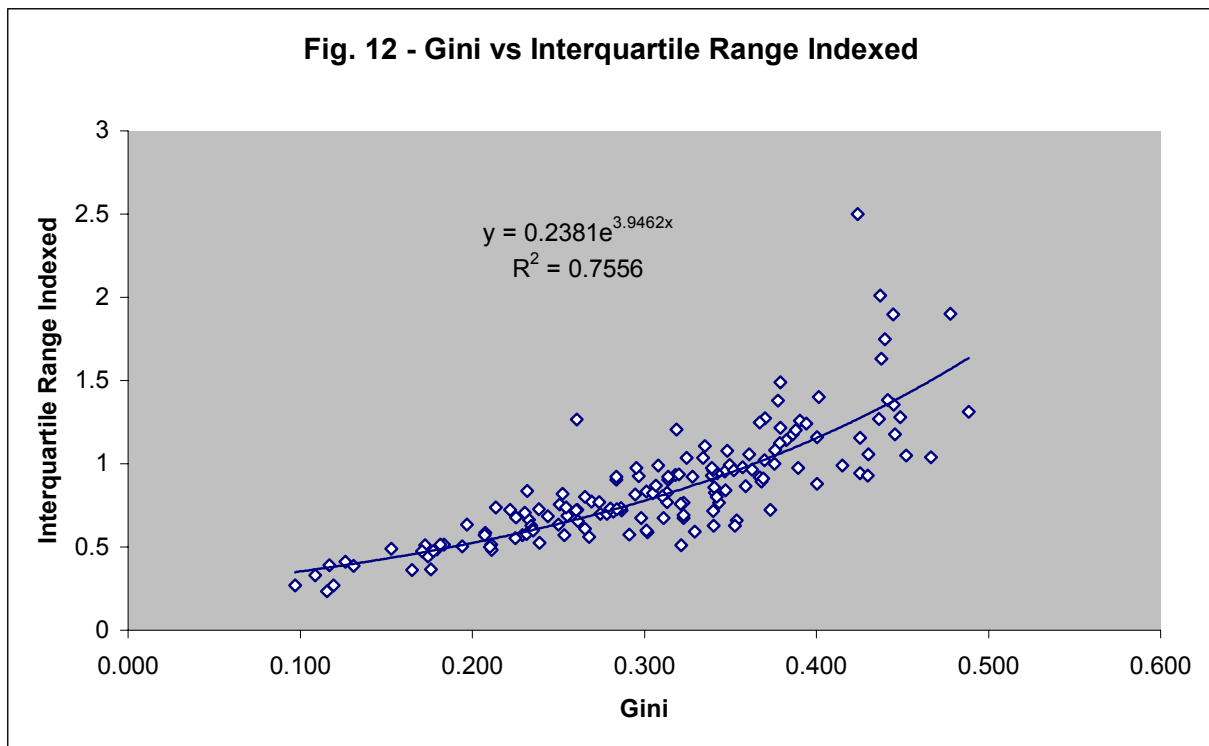
A comparison of interquartile ranges and Gini coefficients at suburb level is given below for the same dataset used above. The correlation between the two measures is poor, indicating that they are measuring fundamentally different things. It is important to understand the measure intimately before adopting it for any given purpose. Here, the Gini coefficient is a relative measure, independent of absolute income, whilst the interquartile range is affected by absolute income levels.



Therefore, it is more sensible to plot the Gini coefficient against the interquartile range indexed on the median for each area, so as to compare like with like. This yields the following picture.

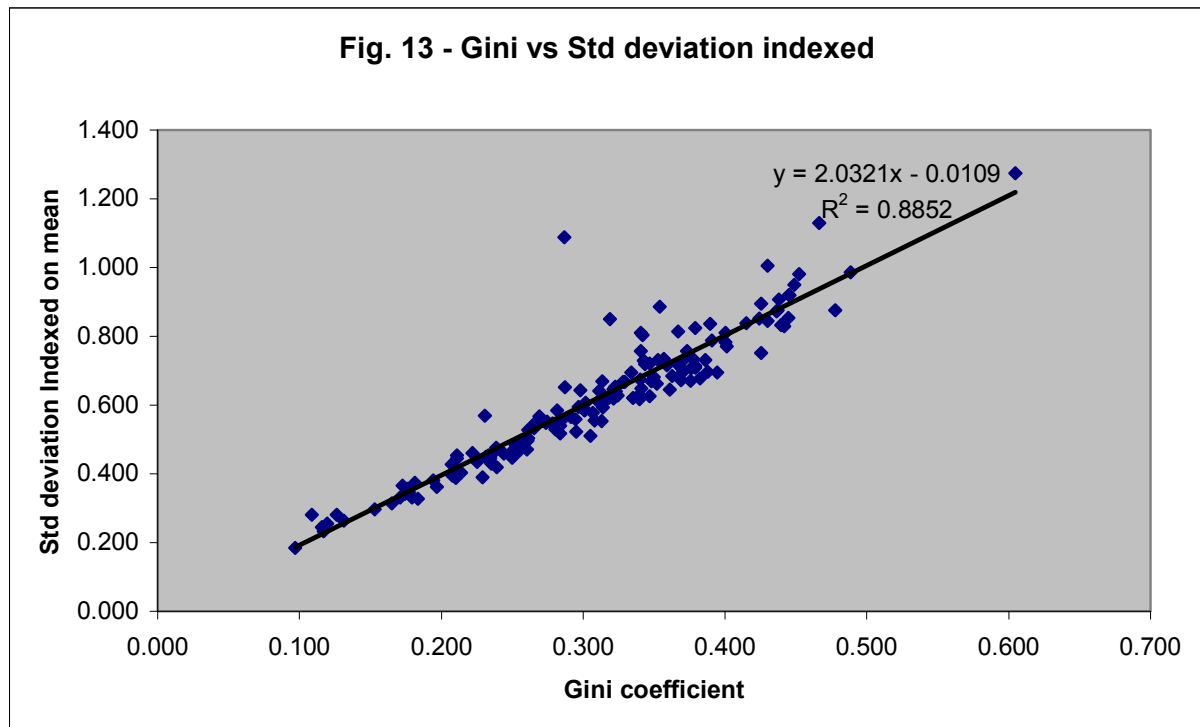


There is a fair correlation except for one outlier. This was an area where the income distribution was almost U-shaped, and clearly contained a very heterogeneous mix of people. If this outlier is removed, there is better correlation. However, there is a danger in simply removing this outlier as there is a considerable gap between the rest of the data points and this one. The relationship between Gini coefficient and interquartile range indexed may well be very non-linear: more data is clearly required to investigate this. Even with this point removed, a non-linear relationship gives the best fit (the linear r^2 was 66%):



The interquartile range is not the best measure to compare against the Gini coefficient, as it is less sensitive to extreme values, by definition – and it just these that are important in examining the issue of inequality. So, finally, plotting the Gini coefficient against the standard deviation indexed on the mean gives the plot below. Interestingly, the outlier is better handled and a simple linear relationship gives a good fit with an r^2 of 89%.

This reaffirms the need in this whole arena to give very careful consideration to what it is that a measure actually measures. The day of the eclectic mix of variables, chosen with little regard to any underlying model or theory or need, is certainly past.



The relationship above is a useful one. The Gini coefficient is straightforward to calculate but the procedure is rather cumbersome as one is working at an aggregated level. By using the above relationship, it is possible to calculate the Gini coefficient directly from the standard deviation and the mean. The robustness of this relationship needs to be verified by examining other datasets.

New Directions – Measuring Well-Being

Moving now to new directions, it has become clear that socio-economic status is simply the starting point. It is a measure than we initially saw as embedded in larger constructs such as measures of urbanisation and, ultimately, measures of well-being, although other conceptualizations begin to suggest themselves. For this to become clear, we need to consider the construct “well-being” more closely, and to consider moving beyond just the satisfaction of basic needs.

As we have seen, measures of well-being start with SES measures, both income and expenditure based, and levels of urbanisation (particularly looking at community facilities, quality of life from an infrastructural point of view, and security) but also add other arenas such as –

1. needs from a Maslow point of view (including higher order needs) as well as unsatisfied needs and concerns especially around housing, nutrition and health;
2. support systems that a person has, both formal and informal;
3. social aspects such as crime, debt, medical/health insurance/aid, household conflict, the need to care for sick household members (AIDS has an impact here), alcoholism, smoking and exposure to pollution (a key aspects in some urban areas with a high incidence of coal smoke);
4. nutrition levels and suitability of actual food intake, the taking of supplements, dieting, one's actual height and weight to measure obesity via the Body Mass Index (BMI – weight (kg) divided by height (m) squared $[w/h^2]$);
5. health from the point of view of illnesses and hospitalisation, exercise (type and frequency), leisure, emotional well-being, and stress; and
6. coping mechanisms – personality and satisfaction with one's life – and key psychographics.

In the US and Canada, well-being is regarded as a crucial concept for both government and business. Here in South Africa, the intention is to produce, for the first time, one comprehensive measure of well-being with a number of key sub-indices, bringing together many of these ideas. This will bring a new understanding to how people actually live their normal every-day lives, giving new insights into understanding why people do what they do, and what is really affecting them materially in their day-to-day living experiences. This will take us well beyond the simple and uni-dimensional concept of socio-economic status as the basis for understanding people into a much richer realm.

The "embedded model" outlined earlier is, in the light of the above, seen to be overly simplistic and linear in its nature. Socio-economic status is but one variable – albeit an important variable – affecting well-being. The diagram overleaf represents the next step in constructing such a more comprehensive model, but this too is still in its early stages.

Maximum well-being can be defined as a state of maximum physical and mental strength and stamina for a person's age, sex, lifestyle and state of health – getting the most out of life no matter one's age or state of health (Ewin, 2000).

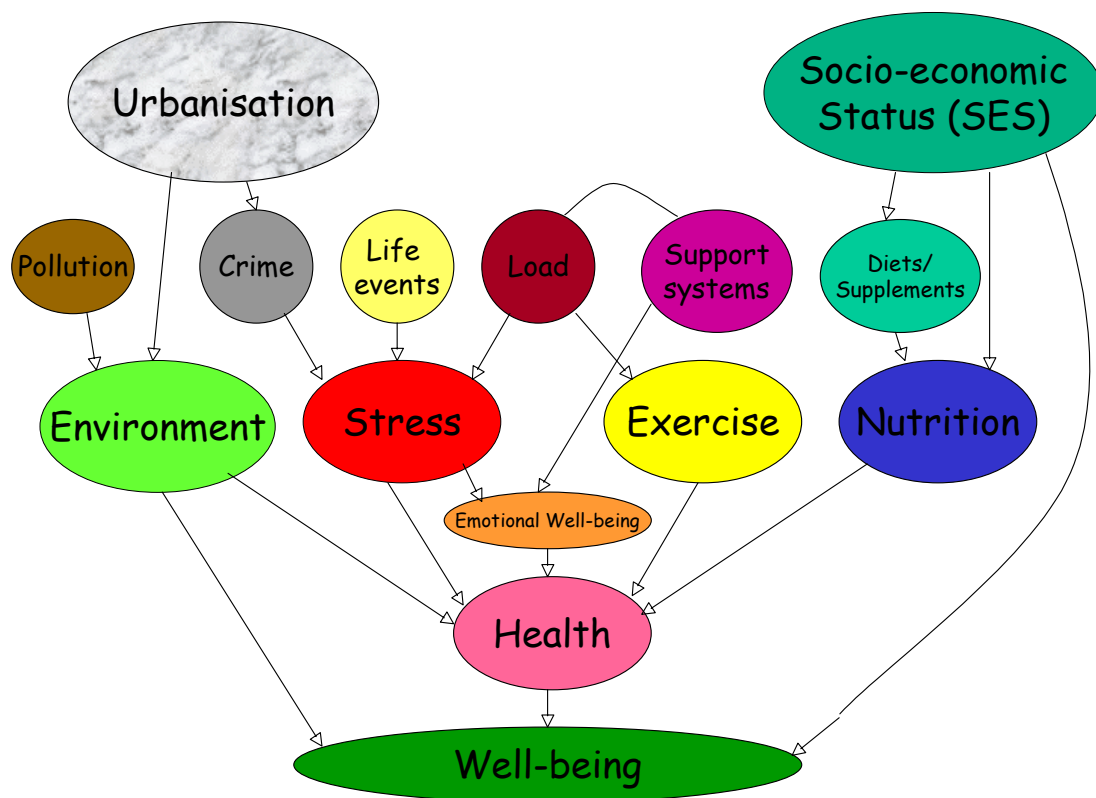
Is this not what marketers (should be perceived to) be delivering? Is this not what we should measure as routine in studies to see if they actually do achieve this? To say that this will be a crucial step in understanding how to understand, satisfy and then talk to people more intelligently is a gross understatement.

“Hardly surprising [that] the ordinary Joe is desperate to find some teeny little thing to trust...Modern consumers have been conned for far too long and far too often by empty promises and superficial concern for their well-being.

“Brand advertising is all about honesty, credibility, loyalty and creating a genuine warm feeling within the very soul of the consumer. It's about the best possible investment any company can make right now. And done with transparency, sincerity and substance, the returns will be quite munificent.”

Chris Moerdyk, p7, Sunday Times Business Times, Johannesburg, 4 August 2002

We suggest that we can no longer see socio-economic status as a surrogate for well-being. We need to understand the other intervening variables depicted in the initial model in Fig. 14.

Fig 14. An Early Structural Model of Well-being

Conclusions

The measurement of socio-economic status is beset by a true plethora of possibilities. This paper has traced the history of some of these and suggested a conceptual framework for the selection of suitable measures for different situations, using level of aggregation and income- vs expenditure-based variables as conscious choices one needs to make. In particular, it cautions against the use just of LSMs, no matter how good LSMs seem to be as a socio-economic status measure, and suggests that marketers and researchers need to begin looking for appropriate measures for appropriate situations. It is crucial to understand what the underlying construct is that is to be measured so that truly appropriate measures can be chosen for different situations.

We have also examined the idea that we need to ask questions about the variability of socio-economic status when we consider geo-demographic classifications relating to socio-economic status. In parenthesis, we saw that the power of money is greater at the poorer end of the wealth continuum than at the upper end. We also saw that lower-income areas are, at present, relatively more heterogeneous than their wealthier counterparts – this having implications for area-based target marketing.

We also need to spend more time comparing our regular and accepted industry measures with other datasets available to international researchers. The comparison of the AMPS Gini coefficient with that available on external datasets is a case in point.

An empirical relationship between the Gini coefficient and other measures of dispersion was discovered:

$$s/x = 2.0321g - 0.0109 \text{ with an } r^2 \text{ of } 89\%, \text{ where } s = \text{standard deviation of monthly income, } x = \text{mean monthly income and } g = \text{Gini coefficient.}$$

This relationship needs to be tested for robustness on other datasets, and its theoretical foundation also investigated: for example, one would expect, theoretically, that the intercept would take on a value of zero, rather than 0.0109.

It seems that a measure of well-being (with SES as a key starting component) will be a course worth pursuing. Maximum well-being can be defined as a state of maximum physical and mental strength and stamina for a person's age, sex, lifestyle and state of health (Ewin, 2000) – getting the most out of life no matter one's age or state of health. Is this not what marketers (should be perceived to) be delivering? Is this not what we should measure as routine in studies to see if they actually do achieve this?

There is much work to do in developing new measures and comparing them with existing measures, as well as deciding on when they should be used. This paper is a step in that direction.

References

- (1984). Harmonisation of demographics for easier international comparisons, European Research, p182, Vol 12 No 4, October 1984, ESOMAR, Amsterdam 1984
- Baker, K, Bermingham, J, and McDonald, C (1979) The Utility to market research of the classification of residential neighbourhoods, Journal of the Market Research Society, London, 1979, reprinted in Journal of the Market Research Society, Vol 39, No 1, January 1997, London, 1997
- Bollen, K.A, Glanville, J.I, Steckby, G (1999), Socioeconomic Status and Class in Studies of Fertility and Health in Developing Countries, Working Paper, Carolina Population Center, University of North Carolina at Chapel Hill, February 1999
- Bollen, K.A, Glanville, J.I, Steckby, G (1999), Economic Status Proxies in Studies of Fertility in Developing Countries: Does the Measure Matter? Working Paper, Carolina Population Center, University of North Carolina at Chapel Hill, May 2001
- Boniaszczuk, J. (1988). Urban/rural Segmentation for Today's Marketer, Proceedings of the 10th SAMRA Convention, South African Marketing Association, Johannesburg, 1988
- Cornish, P, and Denny, M (1989). Demographics are dead – long live demographics, Journal of the Market Research Society, Vol 31 No 3, July 1989, London, 1989
- Dimissie, K, Harley, J. A., Menzies, D, Joseph, L and Ernst, P (2000). Agreement in measuring socio-economic status: area-based vs individual measures, Chronic Diseases in Canada, Vol 21 No 1, Ministry of Health Canada, 2000
- Deonandon, R, Campbell, K, Ostbye, T, Tummon, I, Robertson, J (2000). A comparison of methods for measuring socio-economic status by occupation or postal area, Chronic Diseases in Canada, Vol 21 No 3, Ministry of Health Canada, 2000
- Ewin, J (2000). Nutritional Therapy. Element Books Limited, Shaftesbury, Dorset, 2000
- Higgs, N. T, (1983). Demographic and Social Group Measurement in Zimbabwe, Probe Market Research (Pvt) Ltd, Harare, 1983

Higgs, N. T. (1987). Measuring Change in a Multi-Ethnic Society – The Transition from Rural to Urban Living, Proceedings of the 9th SAMRA Convention, South African Marketing Association, Johannesburg, 1987

Higgs, N. T. (1991). Practical and Innovative Uses of Correspondence Analysis, The Statistician, 40 (2), Institute of Statisticians, London, 1991

Higgs, N. T. (1994). Little Boxes Made of Ticky Tacky, Proceedings of the 16th SAMRA Convention, South African Marketing Association, Johannesburg, 1994

Higgs, N. T. (2002). Measuring Socio-Economic Status: A Discussion and Comparison of Methods, Proceedings of the 22nd SAMRA Convention, South African Marketing Association, Johannesburg, 2002

Matisom, H (1991). It's Not What You Earn, It's the Way You Spend It, Proceedings of the 13th SAMRA Convention, South African Marketing Association, Johannesburg, 1991

Meier, E. and Moy, C. Social Classifications – a new beginning or more of the same? Journal of the Market Research Society, Vol 41, No 2, April 1999, The Market Research Society, London , 1999

Sleight, P (1997). Research's spotlight puts its focus nearer the individual, ResearchPlus, May 1997, The Market Research Society, London 1997

Winters, P (1997). Mapping the route to convergence, ResearchPlus, May 1997, The Market Research Society, London 1997

World Bank (2002). Deninger and Squires dataset (1996) downloaded from World Bank website www.worldbank.org, 2002